Molecular and geochemical insights into microbial life centimeters to kilometers below the seafloor

Thesis by

Elizabeth Trembath-Reichert

In Partial Fulfillment of the Requirements for
the degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2016
(Defended 04/26/2016)

ACKNOWLEDGEMENTS

While I have much to be grateful for in the past 30 years, as it specifically pertains to the attainment of a Doctorate of Philosophy, I would like to thank the following people.

I have a deep appreciation for the opportunities and respect afforded to me throughout my PhD for being a student of Victoria Orphan's. While graduate school certainly has many elements of fate, it was the international goodwill and regard she engenders that gave me the opportunity of a lifetime to work on IDOP Expedition 337. I also appreciated the independence and trust afforded me to tackle graduate school largely on my own terms. And while I may be biased as a microbial ecologist, Microbial Ecology was my favorite class at Caltech, and I hope to teach my own version some day.

I thank my collaborator and committee member, Woodward Fischer, who taught me how an idea (or two) can grow to become bigger than expected. Our projects also gave me the confidence to interact with a wider sphere of Geobiology than I would have thought possible. I also thank Alex Sessions, my academic advisor and committee chair, and committee member Jared Leadbetter for providing broader perspectives for my work.

I thank Yunbin Guan, Chi Ma, Nathan Delleska, and Fenfang Wu for analytical assistance and making the ions and isotopes run on time. I would also like to thank the GPS administrative staff, especially Jan Haskell, Marcia Hudson, Liz Boyd, and Dian Buchness. There was no problem they could not solve throughout my time at Caltech.

I thank the crew and science party of IDOP Expedition 337, specifically my collaborators Fumio Inagaki and Yuki Morono, whose continuous and extremely generous support made my research possible. Fumio shared with me a contagious passion for deep biosphere research and a demonstration of how to make the seemingly impossible, possible. One should only be so lucky as to work with a "Morono method," where every consideration has already been made, tested, and proved. I would also like to thank my collaborator,

Jonathan Wilson, through whom my love of plants grew (pun very much intended) to the next level.

From lab outings to paper discussions thank you to the past and present Orphan Lab members who all had a meaningful impact on my time at Caltech: Stephanie Connon, Anne Dekas, Abbie Green-Saxena, Katherine Dawson, Josh Steele, Shawn McGlynn, Roland Hatzenpichler, David Case, Jeffrey Marlow, Hiro Imachi, Kyle Metcalfe, Grayson Chadwick, Hank Yu, Silvan Scheller, Alexis Pasulka, Connor Skennerton, Sean Mullins, Lizzy Wilbanks, Patty Tavormina, Jennifer Glass, Min Sub Sim, and Paul Magyar. And more specifically, these folks for the following: Stephanie — her ceaseless care and concern made the lab, and most definitely my PCRs; Josh and Shawn — for allowing me to pester them with what was probably a near continuous stream of questions that got me through my qualifying exams; Anne and Abbie — for being who I want to be when I grow up and to this day providing fantastic support and advice; and Kat — for always being up for a geochemistry question, a run, or sometimes both at the same time. Thanks also to my Geobiology office crew Sean, Kyle, Lewis Ward and Jena Johnson for providing sounding boards, sound advice, and a lot of proof reading. I am grateful for the opportunity to mentor three young women in the Orphan Lab: Julia Brown, Heather Grotzinger, and Ana Gonzalez. Your questions made me a better scientist and your enthusiasm was always a great summer pick-me-up.

Thank you to the many great advisors and mentors I have had before graduate school: Brian Mailloux, Martin Stute, and Stephanie Pfirman at Barnard College; Ben Evans, Jeff Ferguson, Mike Brown at NOAA, and Del Bohnenstiel at NC State, especially Brian, who is still generous with his advice and his maxims stick with me to this day. I would also like to thank my teachers in the Durham Public School System, including Lisa Lord, Catherine Clark, and Steven Unruhe, for teaching me how to think, write, and question the world around me.

Thank you to Gina Marrone, Jane Khodarkovsky, Amy Burig, Katherine Roth, and Kacie Ross. You are amazing women who have reminded me there is more to life than work over

our decades of friendship, but I equally appreciate the time you have taken to get to know what I do, and why I do it.

And to my family. Thank you to my mother, Dawn Trembath, for a lifetime of support and intellectual encouragement. I also thank her, and the women in science before me, who pushed for a work environment where I can be a graduate student rather than a *female* graduate student. Thank you to my father, William Reichert, for providing the PI perspective, some combination of learned and genetic dogged determination to keep going when all signs suggest I should quit, and the confidence that I would always figure it out. Thank you to Stephen for being a supportive not-so-little-anymore brother, and I am so excited to see you off to graduate school next year in Physics.

ABSTRACT

At the broadest scale, this thesis is an investigation of how life modulates the movement of essential elements (carbon, sulfur, nitrogen, and silicon) on modern and geologic timescales. Chapters 1 and 2 explore carbon and sulfur cycling microbial communities found centimeters below the seafloor in hydrocarbon-rich methane seep ecosystems. At the Hydrate Ridge methane seep, we investigated how microbial partnerships direct the flow of methane and sulfide in these benthic oases by using identity-based physical separation methods developed in our lab (Magneto-FISH) in conjunction with community profiling and metagenomic sequencing. This method explores the middle ground between single cell and bulk sediment analysis by separating target microbes and their physically associated community for downstream sequencing applications. Magneto-FISH captures were done at a range of microbial taxonomic group specificities and sequenced with both clone library and next-gen iTag 16S rRNA gene methods. Chapter 1 provides a demonstration of how FISH probe taxonomic specificity correlates to resultant Archaeal taxonomic diversity in Magneto-FISHed seep sediments, with specific attention to preparation of Archaea-enriched samples for downstream metagenomic sequencing. In Chapter 2, a Bacteria-focused parallel environmental isolation and sequencing effort was subjected to co-occurrence analyses which suggested there may be far more microbial associations in methane seep systems than are currently appreciated, including partnerships that do not involve the canonical anaerobic methane oxidizing archaea and sulfate reducing bacteria. With samples from IODP Expedition 337 Shimokita coalbed biosphere, Chapter 3 provides evidence for an active microbial assemblage kilometers below the sea floor in the deepest samples ever collected by marine scientific ocean drilling. Using in situ temperature Stable Isotope Probing (SIP) incubations and NanoSIMS, we investigated whole community activity (with the passive tracer D_2O) and substrate specific activity with C1-carbon compounds methylamine and methanol. We found deuterium-based turnover times to be faster (years) than previous deep biosphere estimates (hundreds to thousands of years), but methylotrophy rates to be slower than previous carbon metabolic rates.

PUBLISHED CONTENT AND CONTRIBUTIONS

E. Trembath-Reichert et al. (2016). "Characterization of microbial associations with methanotrophic archaea and sulfate-reducing bacteria through statistical comparison of nested Magneto-FISH enrichments." In: *PeerJ*. e1913. doi: 10.7717/peerj.1913.

E.T.R participated in conception of project, designed study, prepared data and analysis, and wrote the manuscript.

E. Trembath-Reichert et al. (2016). "Gene Sequencing-Based Analysis of Microbial Mat Morphotypes, Caicos Platform, British West Indies." In: *Journal of Sedimentary Research*. (in press June 2016)

E.T.R participated in conception of project, designed study, prepared data and analysis, and participated in writing of the manuscript.

E. Trembath-Reichert et al. (2015) ."Four hundred million years of silica biomineralization in land plants." In: *PNAS*. 112:5449-5454. doi:10.1073/pnas.1500289112

E.T.R participated in conception of project, study design, data preparation and analysis, and writing of the manuscript

E. Trembath-Reichert et al. (2013). "Whole Cell Immunomagnetic Enrichment of Environmental Microbial Consortia Using rRNA-Targeted Magneto-FISH." In: Edward FD, ed. *Methods in Enzymology*: Academic Press, 21-44. doi:10.1016/B978-0-12-407863-5.00002-2

E.T.R participated in conception of project, designed study, prepared data and analysis, and wrote the manuscript.

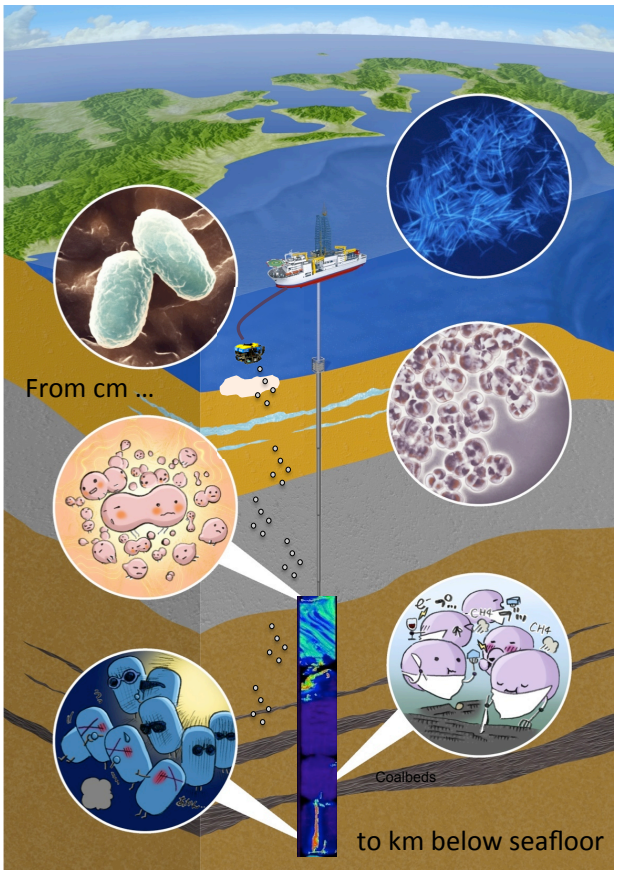
TABLE OF CONTENTS

LIST OF ILLUSTRATIONS AND TABLES

Figure *Page*

Table *Page*

Introduction



While not unified by a single method or study location, this thesis provides four examples of how targeted methods are uniquely able to resolve the character of biologically-mediated carbon, sulfur, silicon, and nitrogen cycling. Discerning the biological component of the systems explored herein is fraught with difficulty stemming from their complexity (Chapters 1 and 2), age (Appendix A and B), or metabolic reticence (Chapter 3).

Chapters 1 and 2 provide a method (Magneto-FISH) and an application (modern methane seep sediments) for dealing with complex microbial communities where multiple species may have, at least superficially, similar roles, such as sulfate reducing bacteria and anaerobic methane-oxidizing archaea, but yet certain partnerships appear preferred over others. By a phylogenetically-selective mechanism, we are able to enrich for target microbes and their physically associated microbial partners to explore spatial arrangement and sequence space in tandem. It is an attractive method for any environment with physically associated microorganisms that can bridge work done at the single-cell and bulk microbial community levels to provide a more holistic framework for microbial interactions.

The Appendices address more historical geobiological questions of how the evolution of land plants may have affected global silicon and carbon cycling (Appendix A) and if microorganisms may be responsible for the structures preserved in microbial mats from the rock record (Appendix B). In both of these systems the original biomaterial is a palimpsest – no longer present or too altered to

directly address our research questions. To overcome the effects of time, we utilized comparative biology methods to determine how extant plants (early evolving land plant lineages) and mats (modern carbonate platform analogs) create the biominerals and biostructures, respectively, that we may see preserved. The applications herein provide application and integration of modern biology to Earth history questions that exemplifies the toolbox of geobiology.

The final chapter (3) interrogates a unique deep biosphere sedimentary environment where terrestrial organic matter from a paleo-swamp has been buried for millions of years under what has now transitioned to an open marine environment. Initial genetic and geochemical results from IDOP Expedition 337 indicate an active assemblage of microbes similar to a modern swamp community (Inagaki et al. 2015), but cell abundances lower than retrieved from any other IODP cruise (1-100 cells/cm^3), despite the extremely high cell abundances at the sediment surface (10^9 cells/cm^3). This extremely low biomass provides a technical challenge to both measuring activity and ensuring the measured activity reflects that of the *in situ* community, rather than any of the myriad contamination sources from drilling to sequencing or an overprinting abiotic process. One could even argue that deep sea drilling is even harder than detecting life in Martian samples, as the contamination on the Earth's surface is so much higher.

As we abut the limit of cell detection, we can no longer hope that the in situ cell concentration will be above the background contamination signal. One of the biggest sources of contamination, drilling mud, is also required for the riser drilling technology that allows deep core recovery. Stringent contamination control, such as identification of samples with high porosity and fracture planes via onboard tomography (CAT scan), can aid in determining the most pristine samples in real-time to avoid using them for stable isotope probing (SIP) incubations. However, it was not possible to remove all sources of contamination from all samples. Therefore, tracking contamination is a more viable pursuit than attempting to remove it completely. This can be done onboard by adding chemical tracers like perfluorocarbon (PFC) to drilling mud and monitoring its concentration, or performing sequencing assays for known microbial contaminants such as water column marine organisms for all downstream biological samples. In addition to these microstructural, chemical, and genetic contamination identification methods, hydrogenase enzymatic and SIP-NanoSIMS activity-based controls showed that when putative contaminant cells did come into contact with samples, they were "dead on arrival," making our activity based measures robust even to contaminant cells for determining viability of in situ populations. While it

cannot be ignored that contaminant cells may provide a potential organic carbon source, we did not determine that any contaminant cells were present (based on expected size for deep biosphere cells) in the incubations discussed in this thesis.

In addition to tracking, technological advancements in sample collection were also used to reduce contamination. Cruises rely heavily on porewater data to determine potential metabolisms, activity profiles, effects of transition from in situ to incubation, or even simply concentrations to use for incubation conditions, but we were either unable to recover any porewater, or what was recovered was too contaminated with drilling mud, through traditional onboard squeezing methods from the 2 km below seafloor coalbeds. To overcome porewater exposure to drilling mud, Exp. 337 was able to use a specialized formation water-sampling device, Schlumberger's Quicksilver probe, for a few select horizons. This allowed us to recover more pristine interstitial water at formation pressures, which is extremely important for gaseous substrate concentration measurements. These Qucksilver probe samples allowed us to confirm the high (mM) levels of ammonium measured in our SIP incubations, highlighting a conundrum of the deep biosphere that has also been found in other studies: Why do deep biosphere cells show a clear preference for nitrogen incorporation over carbon, if they live an ammonium replete environment? The cause of this phenomenon should be a high priority for future deep biosphere research and emphasizes how much remains to be discovered about deep biosphere physiology.

Another approach to understanding deep life physiology has been through attempting to constrain metabolic rates and relating them to turnover of elements deep essential for life (i.e. hydrogen, nitrogen, and carbon). The deep biosphere literature has gravitated toward using a discussion of turnover time, as opposed to growth rate or doubling time, since production of new biomass cannot be predominantly attributed to production of new cells (doubling) over maintenance in energy limiting environments. There is also evidence that deep life is good at recycling biomaterials as a potential energy conservation technique (Takano et al. 2010), which provides an additional caveat to turnover calculations that requires further exploration. However, Morono et al. 2011 showed that most deep biosphere cells were viable, if not actively replicating. At its base level, turnover is the reciprocal rate of some process, be it sulfate reduction coupled to acetate oxidation or amino acid degradation. Converting these rates to turnover times enables comparison to geologic processes, such as sedimentation rate or thermal degradation of organic matter. Previously published deep biosphere biomass turnover times have all been upwards of thousands of years before SIP-

NanoSIMS based times, which appear to be months to years based on our findings that those of Morono et al. SIP-NanoSIMS provides powerful, single-cell resolution of minimal microbial activity that is not possible with bulk geochemical or SIP-RNA/DNA methods. Our study was also able to show that microbial assemblages appear to have different modes of activity within these viable populations and different biosynthesis rates between hydrogen, nitrogen, and carbon, as discussed in Chapter 3. The 2 kmbsf biosynthesis-based turnover times are slower than times for shallower samples that were provided more substrates (Morono et al. 2011), but we do not know if this is an effect of substrates provided and/or differences in the microbial potential for activity from 200 m to 2000 m below seafloor. As we continue to use SIP-NanoSIMS to study the deep biosphere, these distinctions may become more resolved. The continued application of deuterated water as a passive tracer can also provide a baseline metric for unamended, or minimally amended, activity conditions in each new system to connect all future SIP-NanoSIMS experiments, and better determine what is unique to a new environment and what may be universal to deep life.

Finally, the results from Exp. 337 have opened new avenues for conceptualizing the residence time of carbon in coals that have never reached sterilization conditions. With global lignite reserves estimated at 839 Gt (Killops & Killops 2013), understanding what portion of this carbon, assumed stabilized in the lithosphere, may be biologically mobilized and potentially returned to the surface biosphere is important for understanding both deep life and global carbon cycle regulation. Initial investigations into carbon isotopic composition of methoxy-groups in Exp. 337 coal samples are order 50 per mil enriched over the bulk coal carbon values, which provides the tantalizing prospect of a signal for microbial distillation over millions of years. While at the same time, other work suggests that high-pressure environments cause a significant (–20 ‰) depletion in biomass carbon from their carbon source (Fang et al. 2006).

Even with a high-resolution, single-cell activity technique, we must know more about microbial physiology under high pressure, high temperature, and slow growth conditions to be able to contextualize in situ deep biosphere measurements and tease apart what is physiology versus environmental in future deep biosphere SIP-NanoSIMS incubations. Target questions include: Why do microbial populations appear to show different physiological modes, even when in theoretically uniform conditions (Kopf et al. 2015)? How does high pressure affect both natural and labeled isotopic enrichments? What are reasonable water assimilation constants for slow growth conditions, and archaea in general? Can we overcome limitations of deep biosphere biomass and

develop methods to discern biosynthesis of new biomass from microbial maintenance and repair or recycling of necromass? These constraints on biomass turnover, along with technological advancements in three-dimensional imaging of deep biosphere spatial relationships, will then lay the groundwork for myriad other deep biosphere constraints, such as genetic exchange and evolution rates.

The more we know about how life thrives and survives in the present, the more we can plan for our future and interpret our past. The deep biosphere provides a unique environment that blends active biological processes operating in geologic time capsules that are isolated from solar primary production. Through the combination of additional microbially–motivated IODP cruises to more environments and carefully cultivated laboratory experiments, we have much to explore in the years to come that will advance our understanding of life's most extraordinary forms on our planet and maybe others.

Fang J, Uhle M, Billmark K, Bartlett DH, and Kato C. 2006. Fractionation of carbon isotopes in biosynthesis of fatty acids by a piezophilic bacterium Moritella japonica strain DSK1. *Geochimica et Cosmochimica Acta* 70:1753-1760. http://dx.doi.org/10.1016/j.gca.2005.12.011

Inagaki F, Hinrichs K-U, Kubo Y, Bowles MW, Heuer VB, Hong W-L, Hoshino T, Ijiri A, Imachi H, Ito M, Kaneko M, Lever MA, Lin Y-S, Methé BA, Morita S, Morono Y, Tanikawa W, Bihan M, Bowden SA, Elvert M, Glombitza C, Gross D, Harrington GJ, Hori T, Li K, Limmer D, Liu C-H, Murayama M, Ohkouchi N, Ono S, Park Y-S, Phillips SC, Prieto-Mollar X, Purkey M, Riedinger N, Sanada Y, Sauvage J, Snyder G, Susilawati R, Takano Y, Tasumi E, Terada T, Tomaru H, Trembath-Reichert E, Wang DT, and Yamada Y. 2015. Exploring deep microbial life in coal-bearing sediment down to ~2.5 km below the ocean floor. *Science* 349:420-424. 10.1126/science.aaa6882

Killops SD, and Killops VJ. 2013. *Introduction to organic geochemistry*: John Wiley & Sons.

Takano Y, Chikaraishi Y, Ogawa NO, Nomaki H, Morono Y, Inagaki F, Kitazato H, Hinrichs K-U, and Ohkouchi N. 2010. Sedimentary membrane lipids recycled by deep-sea benthic archaea. *Nature Geosci* 3:858-861.

Chapter 1

WHOLE CELL IMMUNOMAGNETIC ENRICHMENT OF ENVIRONMENTAL MICROBIAL CONSORTIA USING RRNA-TARGETED MAGNETO-FISH

Abstract

Magneto-FISH, in combination with metagenomic techniques, explores the middle ground between single cell analysis and complex community characterization in bulk samples to better understand microbial partnerships and their roles in ecosystems. The Magneto-FISH method combines the selectivity of Catalyzed Reporter Deposition -Fluorescence In-Situ Hybridization (CARD-FISH) with immunomagnetic capture to provide targeted molecular and metagenomic analysis of co-associated microorganisms in the environment. This method was originally developed by Pernthaler et al. and Pernthaler & Orphan (2008; 2010). It led to the discovery of new bacterial groups associated with anaerobic methane-oxidizing (ANME-2) archaea in methane seeps as well as provided insight into their physiological potential using metagenomics. Here, we demonstrate the utility of this method for capturing aggregated methanotrophic consortia using a series of nested oligonucleotide probes of differing specificity designed to target either the ANME archaea or their *Deltaproteobacteria* partner, combined with 16S rRNA and *mcrA* analysis. This chapter outlines a modified Magneto-FISH protocol for large and small volume samples and evaluates the strengths and limitations of this method predominantly focusing on 1) the relationship between FISH probe specificity and sample selectivity, 2) means of improving DNA yield from paraformaldehyde-fixed samples, and 3) suggestions for adapting the Magneto-FISH method for other microbial systems, including potential for single cell recovery.

Introduction

As advancements in high-throughput sequencing technology allow deeper and more cost effective means of sequencing complex microbial assemblages, we are left with more data, but not necessarily more means to understand it. The development of microbiological techniques to isolate and visualize environmental microorganisms *a priori* can be used to meaningfully parse environmental samples before metagenomic processing, and thereby provide additional context for downstream bioinformatic data interpretation. There is also increasing awareness that microbe-environment and microbe-microbe interactions are important factors in assessing microbial systems, their metabolic potential, and how these relationships affect larger scale processes such as ecosystem nutrient cycling.

A range of *in situ* techniques are currently available for physical separation of microorganisms of interest from environmental samples. Methods involving selection from a complex microbial sample often involve a stage of phylogenetic identification, such as 16S rRNA-based fluorescence *in situ* hybridization (FISH), coupled to a means of physical separation such as flow sorting (Amann et al. 1990; Yilmaz et al. 2010) (also see chapters in this volume by Zehr and Haroon), optical trapping (Ashkin 1997), microfluidics (Melin & Quake 2007), or immunomagnetic beads (Šafařík & Šafaříková 1999). This is in contrast to separation methods where selection is based on a property other than identity, such as metabolic activity (Kalyuzhnaya et al. 2008), followed by downstream identification of the population exhibiting the property of interest. The majority of these methods have focused on single cell analysis, rather than examining intact multi-species microbial associations, with the exception of intracellular microbial interactions (Yoon et al. 2011).

The Magneto-FISH method was originally developed by Pernthaler et al. (2008) to enrich for and characterize microbial associations in the environment. This technique was specifically developed for studying inter-species partnerships between anaerobic methane-oxidizing (ANME) archaea and sulfate-reducing *Deltaproteobacteria* (SRB) in anoxic marine sediments (Boetius et al. 2000; Orphan et al. 2002). This method is based on 16S rRNA Catalyzed Reporter Deposition (CARD)-FISH (identity) (Pernthaler et al. 2002) and immunomagnetic bead capture (separation) (Pernthaler et al. 2008; Pernthaler & Orphan 2010). The Magneto-FISH method was shown to successfully

concentrate the population of interest and aid in microbial association hypothesis development that could be further supported with metagenomics, microscopy, and isotope-labeling techniques. This provides a means to study metabolic potential at a level that is not defined in separate units of species identity, but operational groups of organisms that have evolved to serve a function, such as the symbiotic consortia mediating methane oxidation coupled to sulfate reduction. Magneto-FISH is also compatible with the physical challenges of sediment associated ANME-SRB aggregates, namely their heterogeneous morphology, wide size range (~3-100 μm diameter), and frequent association with mineral and sediment particles.

In evaluating the application of Magneto-FISH to other environmental populations, it is important to consider sample input constraints such as microbe size and morphology, sample output requirements such as yield and purity, and of course time and expense. Autofluorescent sediment particles and diverse ANME/SRB consortia size complicated the successful application of flow sorting approaches to the AOM system. In other environments, FAC sorting has been shown to be an effective means of cell separation, but often requires DNA amplification (Rodrigue et al. 2009; Woyke et al. 2011). Yield and purity are also often opposing constraints. For example, FAC sorting can provide high sample purity, but may require significant instrument time for collecting sufficient material without including a post-amplification step (Woyke et al. 2011). Sample yield remains an issue with Magneto-FISH, as well. Initial application of Magneto-FISH required Multiple Displacement Amplification (MDA) before construction of metagenomic libraries for 454 pyrosequencing (Pernthaler et al. 2008). However, advances in library preparation (e.g. Nextera XT) have significantly lowered the minimum DNA concentrations required. Magneto-FISH can be completed in a day and does not require the use of any specialized equipment beyond an epifluorescent microscope. The main expense is reagents, which scales with amount of sample processed and diversity of FISH probes needed. Another advantage is the versatility of this method. It is compatible with a broad range of oligonucleotide probes incorporated into the same basic protocol; no instrument adjustment or recalibration is required between runs or with different microbial targets.

This chapter introduces three modifications to the Magneto-FISH protocol of Pernthaler et al. (2008) to improve DNA recovery and labor efficiency: 1) immuno-based attachment of magnetic beads for single cell capture, 2) magnetic separation in a standard magnetic holder, and 3) DNA

cross-link reversal incubation during extraction. Using this modified method, we evaluate the DNA recovery and microbial target specificity using a nested set of oligonucleotide probes and discuss 1) increasing target DNA yield for current template requirements amplification, 2) the relationship between sample purity and FISH probe specificity, 3) controls for association selectivity, and 4) DNA quality for metagenomic techniques.

Methods

Samples and controls used in Magneto-FISH capture experiments

Sediment samples were collected in September 2011 from methane seeps within the S. Hydrate Ridge area off the coast of Oregon at a depth of 775 m using the R.O.V *JASON* and the R/V *Atlantis*. Marine sediment was collected in a push core (PC-47) associated with a sulfide-oxidizing microbial mat adjacent to an actively bubbling methane vent. A sediment slurry from the upper 0-15 cm depth horizon was prepared with one volume N_2 sparged artificial seawater to one volume sediment, over-pressured with methane (3 bar) and incubated at 8ºC in a 1L Pyrex bottle sealed with a butyl rubber stopper. A 4 ml sample from the incubation was collected on November 19, 2012. Samples were immediately fixed in 0.5 ml sediment aliquots in 1.5% paraformaldehyde (PFA) for 1 hr at room temperature (fixation can alternatively be performed at 4^{o}C overnight). Samples were washed in 50% 1x PBS: 50% EtOH, then 75% EtOH: 25% DI water, and resuspended in 2 volumes (1 ml) 100% ethanol. Samples were centrifuged at 1000 x*g* for 1 min between wash steps.

As a control to test association specificity, 0.5 ml of sediment slurry was spiked with 10 µl of turbid *Paracoccus denitrificans*, strain ATCC 19367. After addition the sample was quickly vortexed, fixed, and washed as described above. 16S rRNA diversity surveys of the original sediment incubation sample supported the absence of *P. denitrificans* in the bulk sediment.

Magneto-FISH

A detailed protocol is provided in Table 1 and additional information and explanation of the major steps are provided below. When using Magneto-FISH with marine sediment samples, 100 µl of fixed sediment slurry (resuspended in 100% ethanol) is the recommended starting volume for the

recovery of PCR-amplifiable DNA. The method has been tested with sediment volumes ranging from 75-3000 µl. Smaller sample sizes have higher target purity, but lower DNA yield. For the purposes of this chapter, all reagent amounts are given for the 100 µl starting sample size (small scale prep), but can be scaled up as indicated for larger samples. There are two means to scale up these reactions: 1) using more of the starting sample with the same oligonucleotide probe, 2) or using more of starting sample, but with different probes. With option 1, all sample aliquots can be combined during wash steps as indicated. For option 2, sample aliquots can be combined during the initial permeabilization stages, but can no longer be combined after probes have been applied. All reagents should be sterilized by filtration (0.22µm) prior to use, and sterile sample containers should be used in subsequent steps. Additionally, after fluor addition samples should treated as light-sensitive.

Permeabilization and inhibition of endogenous peroxidases

The TE pH 9 heating step serves to permeabilize cells and loosen sediment particles. The hydrogen peroxide addition inhibits endogenous peroxidases prior to the CARD reaction. To remove ethanol, spin sediment-ethanol slurry at 16,000 x*g* for 1 min, remove supernatant, and resuspend in TE (pH 9). When performing multiple reactions with the same sediment, they can be combined during these steps (i.e. for 6 captures, add 600 µl (original volume) of sediment slurry to 100ml Tris-EDTA (TE), pH 9) after removing ethanol.

Liquid CARD-FISH

All oligonucleotide probes and corresponding formamide concentrations used are summarized in Table 2. When using a histological microwave for hybridization, formamide concentrations were lowered by 10% below the concentrations optimized for a conventional hybridization oven (Fike et al. 2008). *Note*: A hybridization oven can also be used for liquid CARD-FISH, but incubation time should be increased to at least 2 hrs or more.

If doing multiple reactions, evenly divide sediment pellet among all samples in each CARD hybridization buffer with the appropriate formamide concentration. For histological microwave use, orient the beaker and samples such that only water, and no samples, is in the path of the temperature probe. Inverting or vortexing samples a few times during this incubation can improve

mixing of probe and sample, since sediment will tend to settle out of suspension during the incubation. All samples with the same probe can be combined during wash steps, but different probe samples must be kept separate.

For the amplification reaction, samples must be evenly divided into their initial starting proportions (if started with 600 µl of slurry, then separate into 6 aliquots) for proper target to probe ratios, but like-samples can be recombined during subsequent wash steps. For a larger combined wash, samples can remain in a 50 ml tube with the appropriate amount of PBS and PPi after blocking reagent and washing steps. Hybridized samples can also be stored overnight at 4°C before proceeding with magnetic capture.

Table 1: Step-by-step detailed instructions for Magneto-FISH protocol. Additional information and suggestions are included in the text for each section. Recommended equipment list: hybridization microwave [BP-111-RS-IR, Microwave Research & Applications], centrifuge (microtubes and 50 ml tubes), sonicator with tapered microtip probe [Branson Sonifier W-150 ultrasonic cell disruptor], rotating or shaking incubator/hybridization oven, magnetic holder [Dynal MPC-1.5ml], waterbath, bead beating tubes with garnet sand [PowerSoil DNA Kit PowerBead Tubes, MO BIO] and bead beater [FastPrep FP120, Thermo Electron Corp.], cellulose spin columns [Microcon, Millipore], vortex [Vortex-Genie 2, MO BIO], 1.5 ml maximum recovery centrifuge tubes [Flex-Tubes 1.5ml, Eppendorf]. Special reagents: Linear Acrylamide, Dextran Sulfate, Blocking Reagent, HRP-probes, fluor-labeled tyramide(s), biotin tyramide, anti-fluor mouse monoclonal IgG antibody [Life Technologies], Dynabeads Pan Mouse IgG [Life Technologies].

Magneto-FISH

1. Permeabilization and inhibition of endogenous peroxidase
    a. Add 100µl sediment slurry to 100 ml TE pH 9 in a sterile 250 ml glass beaker (or other flat-bottomed vessel to maximize surface area).
    b. Microwave 2 min at 65°C in a hybridization microwave (100% power) [BP-111-RS-IR, Microwave Research & Applications].
    c. Transfer to two 50 ml Falcon tubes and spin at 5000 x*g* for 5 min at 4°C (all spin steps should be performed in this manner unless otherwise indicated).
    d. Decant supernatant taking care to retain the sediment pellet by pouring slowly and all in one motion.
    e. Resuspend in 50 ml 1x Phosphate Buffered Saline (PBS), 0.01M Sodium Pyrophosphate (PPi), 0.1% H_2O_2 and incubate at room temperature for 10 min, inverting tubes occasionally to keep sediment in suspension.
    f. Sonicate for three 10 s pulses on setting 3 (~6V(rms) output power) [Branson Sonifier W-150 ultrasonic cell disruptor] at room temperature with sterile remote tapered microtip probe [Branson] inserted into the liquid.
    g. Spin and decant.
2. Liquid CARD-FISH
    a. Resuspend sediment in 2 ml CARD buffer [0.9M NaCl, 20mM Tris-HCl

pH 7.5, 10% w/v Dextran Sulfate, 1% Blocking Reagent (in pH 7.5 maleic acid buffer), 0.02% w/v SDS] and transfer to a 2 ml Eppendorf tube.

b. Add 20 µl of 50 ng/µl CARD probe and vortex [Vortex-Genie 2, MO BIO] briefly to mix.
c. Wrap tubes in parafilm and tape to the sides of a beaker filled with DI water, such that tubes float in an approximately horizontal orientation.
d. Microwave for 30 min at 46^{o}C, power setting of 100%.
e. Remove samples from the water bath and remove parafilm.
f. Spin tubes at 10,000 x*g* for 2 min.
g. Decant supernatant into formamide waste and resuspend hybridized sediment in 50 ml 1x PBS.
h. Incubate at room temperature for 10 min, shaking occasionally.
i. Centrifuge, decant supernatant, resuspend in fresh 1x PBS, centrifuge and decant again, leaving pellet.
j. Resuspend in 2 ml amplification buffer [1x PBS, 1% Blocking Reagent, 10% w/v Dextran Sulfate, 2M NaCl] in 2 ml Eppendorf tube.
k. Add 2 µl fluor-labeled tyramide (0.5 µg/ml), 2 µl biotin tyramide (0.5 µg/ml), and 5 µl 0.0015% H_2O_2.
l. Wrap tube(s) in foil to protect from light and incubate with gentle shaking or rotating at 37^{o}C for 1.5 hrs.
m. Spin at 10,000 x*g* for 2 min.
n. Decant supernatant and resuspend in 50 ml 1x PBS in 50 ml centrifuge tube.
o. Incubate for 10 min at room temperature in the dark, shaking occasionally.
p. Spin, resuspend in 50 ml 1x PBS, and spin again.
q. Resuspend in 49.5 ml 1x PBS and 0.5 ml 10% blocking reagent in a 50 ml falcon tube.
r. Microwave [BP-111-RS-IR, Microwave Research & Applications] in a vessel large enough to submerge 50 ml tubes for 20 min at 40^{o}C in DI water.
s. Centrifuge, decant, and resuspend in 50 ml 1x PBS, then centrifuge and decant again.
t. Resuspend each sample in 1 ml 1x PBS, 0.01M PPi in a 1.5 ml Eppendorf tube
u. Counterstain a sample aliquot with DAPI and verify hybridization by microscopy.

3. Magnetic Bead preparation and Magnetic Cell Capture
    a. Sonicate sample in 1.5 ml tube for 5 sec, setting 3 at room temperature to resuspend cells.
    b. Add 5 µl anti-fluor mouse monoclonal IgG antibody [Life Technologies] per 1 ml reaction volume and incubate at 4^{o}C for 20 min rotating to keep sediment in suspension [Hybridization Oven, VWR].
    c. While the sample is incubating, prepare beads:
        i. Add 25 µl of Dynabeads Pan Mouse IgG [Life Technologies] per reaction to 1 ml of Buffer1 [1x PBS, 0.1% BSA] and place in magnetic holder [Dynal MPC-1.5ml].

ii. Invert holder and tube(s) multiple times to wash all beads down to magnet. Remove liquid with pipet and treat as azide waste. Remove tube from holder and resuspend washed beads in 30 µl of Buffer1.

d. After 20 min incubation, spin sample at 300 x*g* for 8 min at 4°C.
e. Decant supernatant, resuspend sediment pellet in Buffer1, and spin again as in step 3d. Decant supernatant.
f. Add 30 µl of washed beads and 1 ml Buffer1 per sample volume.
g. Incubate 1.5 hrs at 4°C in dark while rotating to keep sediment in suspension.
h. Place sample(s) into magnetic holder slots. Invert multiple times and let sit 1 min until sediment has settled to the bottom of the tube. Remove liquid including all sediment while trying not to disturb magnetic beads.
i. To wash beads and target cells, remove tube from magnetic holder and add 1 ml Buffer1 while aiming pipet tip at magnetic beads to resuspend them. If all beads are not resuspend when adding 1 ml, pipet up and down slowly to resuspend remaining beads from side of the tube. After a few washes, counterstain a sample aliquot with DAPI and verify bead attachment by microscopy. Repeat wash step at least 9 more times (10 total).
j. Save any sample necessary for further microscopy before proceeding to DNA extraction.
k. After final wash, resuspend washed beads and cells in 400 µl of TE buffer (pH 8).

DNA Processing

1. Cell lysis and reversing crosslinks in DNA
    a. Add lysis reagents (10 µl 5M NaCl and 25 µl 20% SDS) to 400 µl TE with beads from step 3k.
    b. Remove liquid from screw cap 2 ml bead beating tube with garnet sand (PowerSoil DNA Kit PowerBead Tubes, MO BIO].
    c. Add total volume of sample and lysis reagents (435 µl) to bead beating tube.
    d. Bead beat at setting 5.5 for 45 s [FastPrep FP120, Thermo Electron Corp.].
    e. 3 rounds of alternating Freeze/Thaw (-80°C and 65°C were used in this study).
    f. Incubate samples for at least 2 hrs, up to 48 hrs, in a 65°C water bath.
2. DNA Extraction
    a. Add 0.5 ml phenol [pH 8, 0.1% hydroxyquinoline] to bead beating tube.
    b. Vortex to mix, and spin for 2 min at 16,000 x*g*.
    c. Transfer supernatant to a new tube while avoiding particulates at TE/phenol interface.
    d. Add 250 µl phenol and 250 µl Chloroform:IAA (24:1).
    e. Vortex to mix, spin 1 min at 16,000 x*g*, and transfer supernatant to new tube.
    f. Add 500 µl Chloroform:IAA, vortex briefly, spin 2 min at 16,000 x*g*.

g. Add 200 µl TE to cellulose spin column [Microcon, Millipore], then add DNA supernatant.
h. Spin 8 min, 14,000 x*g*. Wash DNA on spin column 3x with 500 µl TE.
i. Elute into new tube at 1,000 x*g* for 3 min, as per manufacturer directions.

3. Concentration
   a. Transfer DNA from elution tube to 1.5 ml maximum recovery centrifuge tube [Flex-Tubes 1.5 ml, Eppendorf] and bring volume up to 37.5 µl with TE.
   b. Add 12.5 µl 10M Ammonium Acetate (2.5 M final concentration), 0.2 µl Linear Acrylamide, and 125 µl cold EtOH (2.5 volumes).
   c. Precipitate DNA overnight in wet ice (0^{o}C).
   d. Spin 18,000 x*g* in a microfuge for 30 min at 4^{o}C to pellet DNA.
   e. Decant supernatant, careful to retain pellet.
   f. Lay tube on its side with cap open on a heat block at 65^{o}C to evaporate remaining liquid. Resuspend in 10 µl Tris-HCl (pH 8).

Magnetic Capture

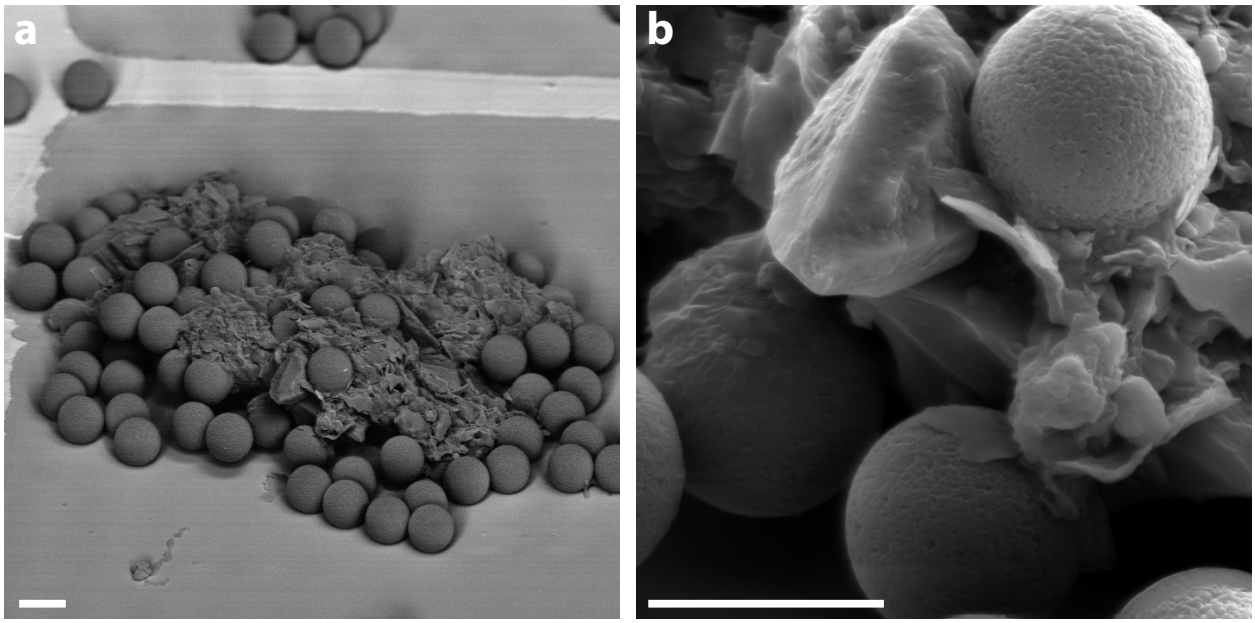


Figure 1. a. Electron backscatter image and b. close-up secondary electron image of Magneto-FISH aggregate. Bar is 5 µm.

The magnetic capture consists of three main steps: 1) Incubation of anti-fluor antibodies with fluor labeled cells (Figure 1), followed by two centrifugation wash steps to remove any unassociated anti-fluor. Increasing the centrifugation speed/force does not appear detrimental and could be

optimal for other systems in order to retain more material (for example, non-sediment associated microbes, single cells, and smaller aggregates). We recommended saving supernatants from the washing steps until satisfied with magneto capture, in the event that steps need to be repeated or re-optimized during bead attachment. 2) Incubation of magnetic beads with anti-fluor attached cells. 3) Removing remaining sediment and cells that did not attach to beads using a magnetic tube holder. It is also recommended to retain the first two sediment washes until satisfied with magnetic capture and to evaluate efficiency (number of captured cells/cells remaining in wash). Bead resuspension between washes should be done as gently as possible to reduce the strain on bead-cell association. When performing larger reactions, multiply number of reactions by 1 ml Buffer1 to calculate volume of wash to use. Larger magnetic holders for 15 ml or 50 ml tubes may also be necessary. To reduce larger volumes down to 400 µl for extraction, adding additional washes in increasingly smaller volumes before final suspension in TE may be helpful. After the final resuspension, it is easier to work with low retention tips as beads can stick to tips and tubes when in TE.

In Pernthaler et al. (2008) the magnetic beads and anti-fluor antibodies were incubated together before application to the sediment. Here, anti-fluor and magnetic beads are added in separate, successive reactions. We have found that addition of the anti-fluor antibodies independently, followed by subsequent addition of magnetic beads, resulted in higher recoveries, likely a result of improved antibody-cell hybridization, which may avoid steric hindrance caused by bulky magnetic beads during the attachment stage (*R.S. Poretsky and V.J. Orphan*, unpublished). Pernthaler et al. (2008) also developed a separatory funnel apparatus outfitted with a neodymium ring magnet to allow large volumes of buffer to continually wash the magnetic beads and attached cells (Pernthaler & Orphan 2010). To simplify this procedure, and increase the recovery of cells after magnetic capture, a conventional magnetic tube holder for 1.5 ml and 50 ml falcon tubes (Dynal) was used in combination with multiple washes to remove residual sediment particles and collect the bead-attached cells. We found that these modifications achieved a similar level of target cell enrichment with small samples.

DNA Processing

Lysis and reversing paraformaldehyde crosslinks

Higher DNA yields have been reported after 48 hrs cross-link reversal incubation with no degradation of sample (Gilbert et al. 2007), but may not be necessary if fixation duration and time since fixation are short, or a different fixative is used. Gilbert et al. (2007) also provide a review of other published amendments to DNA extraction methods for PFA fixed DNA that may provide further insight for optimizing this method for different sample types or downstream goals.

Extraction and Concentration

DNA extraction and concentration methods are based on Sambrook and Russell (2001) and Crouse and Amorese (1987). Bead beating can be replaced by vortexing at maximum speed for 10 min. Freeze/thaw cycles can be performed at a range of freezing and thawing temperatures. -80^{o}C and 65^{o}C were chosen based on equipment available and for rapid cycling between states.

Quantification

The extremely low DNA concentrations from magneto-FISH samples requires the highest possible sensitivity for detection, reduction of sample loss during quantification, and minimization of contamination during processing or from reagents (Woyke et al. 2011). For DNA quantification prior to metagenomic library construction, the use of a Qubit fluorometer and HS dsDNA Assay kit (Life Technologies) is recommended, though it may require as much as half of the final DNA extract for the small-scale preparation (5 µl) to obtain a reading above detection.

PCR and Cloning

Archaeal *16S rRNA* Primers, annealing 54^{o}C:

- Arc23F (DeLong 1992; Waldron et al. 2007) – TCC GGT TGA TCC YGC C
- U1492R (Lane 1991) – GGY TAC CTT GTT ACG ACT T

mcrA Primers, annealing 52^{o}C:

- ME1 (Hales et al. 1996) – GCM ATG CAR ATH GGW ATG TC
- ME2 (Hales et al. 1996) – TCA TKG CRT AGT TDG GRT AGT

Paracoccus denitrificans, annealing 50^{o}C:

- Bac27F (Lane 1991) – AGA GTT TGA TYM TGG CTC
- PAR1244R (Neef et al. 1996) – GGA TTA ACC CAC TGT CAC

Hot start Taq DNA polymerases, such as HotMaster (5 PRIME), are recommended for PFA fixed samples, especially when trying to amplify larger (>1000 bp) fragments such as full length 16S

rRNA (Imyanitov et al. 2006). All Magneto-FISH PCR reactions were 12.5 µl total volume containing 1 µl DNA template. The following thermocycler conditions were used: 95°C initial denaturation of 2 min, followed by 40 cycles of 94°C for 20 s, annealing for 20 s at temperatures listed above for primers, 1-1.5 min extension at 72°C, and a final extension of 10 min at 72°C. PCR reagents were used at the following concentrations: 1x HotMaster buffer with 25 mM Mg^{2+}, 0.22 mM dNTPs, 0.2 µM forward and reverse primer, 0.2 U HotMaster Taq per µl reaction. Prior to cloning, an additional reconditioning PCR step of 5 to 8 cycles was performed in 25 µl, using 5 µl of template from the original PCR reaction (Thompson et al. 2002). Reconditioned PCRs were quantified by gel electrophoresis (1% gel, SYBR safe stain), filtered (MultiScreen PCR Filter Plate #MSNU03010, Millipore) to remove primers, and concentrated in 10µl Tris-HCl (pH 8). Approximately 4 µl of PCR product was used per reaction according to guidelines for TOPO TA Cloning Kit for Sequencing with pCR4-TOPO Vector and One Shot Top 10 chemically competent *E. coli* (Life Technologies). An ABI Prism 3730 DNA sequencer was used for all sequencing.

Phylogenetic analysis of 16S rRNA and metabolic genes (mcrA)

Translated methyl-coenzyme reductase alpha subunit (*mcrA*) nucleotide sequences were added to an *mcrA* database and aligned in ARB utilizing the ARB alignment features (Ludwig et al. 2004). 16S rRNA sequences were aligned using Silva online aligner (Quast et al. 2013) and then imported into ARB to verify alignment. Representative sequences were selected from the alignments and cropped to a common region containing no primers: 451 nucleotide containing positions for *mcrA* and 901 nucleotide containing positions for 16S rRNA. Sequences were then exported from ARB and phylogenies were computed using MrBayes (Ronquist et al. 2012). Convergence was determined by an average standard deviation of split frequencies <0.01. Both phylogenies were computed by nucleotide. Inverse gamma rates and default recommendations from Hall (2004) were used for all other MrBayes parameters.

Results and Discussion

Evaluating the quantification and specificity of captured targets using general and species-specific FISH probes

In the initial Magneto-FISH publication by Pernthaler et al. (2008), a clade specific probe targeting the archaeal subgroup ANME-2c (Knittel et al. 2005) was used to successfully enrich this group and physically associated bacteria from Eel River Basin methane seep sediments, increasing the percentage of recovered ANME-2c from 26% in the original sediments to 92% of the Magneto-FISH captured archaeal diversity. Here we expand upon this work, specifically evaluating how FISH probe selectivity affects Magneto-FISH microbial target selectivity. Five different CARD-FISH probes, including Domain-level and group-specific probes targeting major methane seep archaeal and sulfate-reducing bacterial groups were evaluated (Figure 2 and Table 2). The three archaeal probes used were ANME-2c_760, Eel-MSMX_932 (general ANME; Boetius et al. 2000) and Arc_915 (general archaea; Stahl & Amann 1991). Two bacterial probes, Seep-1a_1441 (Schreiber et al. 2010) and Delta_495a (Loy et al. 2002), were also used to target *Deltaproteobacteria* that commonly associate with ANME archaea. Seep-1a_1441 is a probe designed to hit a specific subgroup of the *Desulfococcus/Desulfosarcina* (DSS), shown to be a dominant partner of ANME-2c archaea in methane seeps (Schreiber et al. 2010). However, greater diversity of SRB and other bacteria exist in association with ANME in seeps (Holler et al. 2011; Knittel et al. 2003; Löesekann et al. 2007; Niemann et al. 2006; Orphan et al. 2002). Delta_495a targets a broader range of SRB, and is expected to recover additional diversity if present in the sample. This allows investigation of both the effectiveness of target species enrichment, as well as providing information on the breadth of associated ANME partners.

Total DNA recoveries from each Magneto-FISH capture ranged from below detection to 1.2 ng, depending on the specificity of the FISH capture probe (Table 2). The total DNA extracted for each sample was consistent with the predicted yield based on oligonucleotide probe specificity, where clade specific probes (ANME-2c_760 and Seep-1a_1441) yielded lower DNA recoveries relative to Magneto-FISH captures with more general probes (Eel-MSMX_932, Delta_495a, and Arc_915). The DNA recovered from the group-specific Seep-1a_1441 probe is reported as not detected in the table, however, only 1µl (10%) of the total DNA extract was quantified to preserve sample material. Typically 5 µl (50%) was necessary for detection of other Magneto-FISH captures. Based on PCR amplification, the Seep-1a_1441 Magneto-FISH capture most likely recovered a DNA concentration similar to that observed with ANME-2c_760.

In an attempt to quantify the level of confidence in Magneto-FISH microbial associations, we spiked a bulk sediment sample with a known volume of an alien cultured organism, *Paracoccus denitrificans*. This pure culture has a diagnostic morphology and was not detected in any of our bulk sediment analyses. After confirming with FISH and microscopy that the introduced *P. denitrificans* cells were present in the sediment sample after fixation and at an environmentally relevant concentration (visible in each field of view, but not a dominant species), this spiked sample was used for Magneto-FISH with the Eel-MSMX_932 probe. Using a primer specific to *P. denitrificans* (Neef et al. 1996), DNA recovered from the capture did not reveal *P. denitrificans* contamination after 40 cycles of PCR. There was also positive PCR amplification of *P. denitrificans* from the spiked bulk sediment DNA extraction. Universal bacterial 16S rRNA primers were used to confirm that the Eel-MSMX_932+*P. denitrificans* sample did not have amplification inhibition. This suggests that microorganisms associated with target Magneto-FISH samples are unlikely present due to non-specific attachment during the Magneto-FISH or sample preservation protocol.

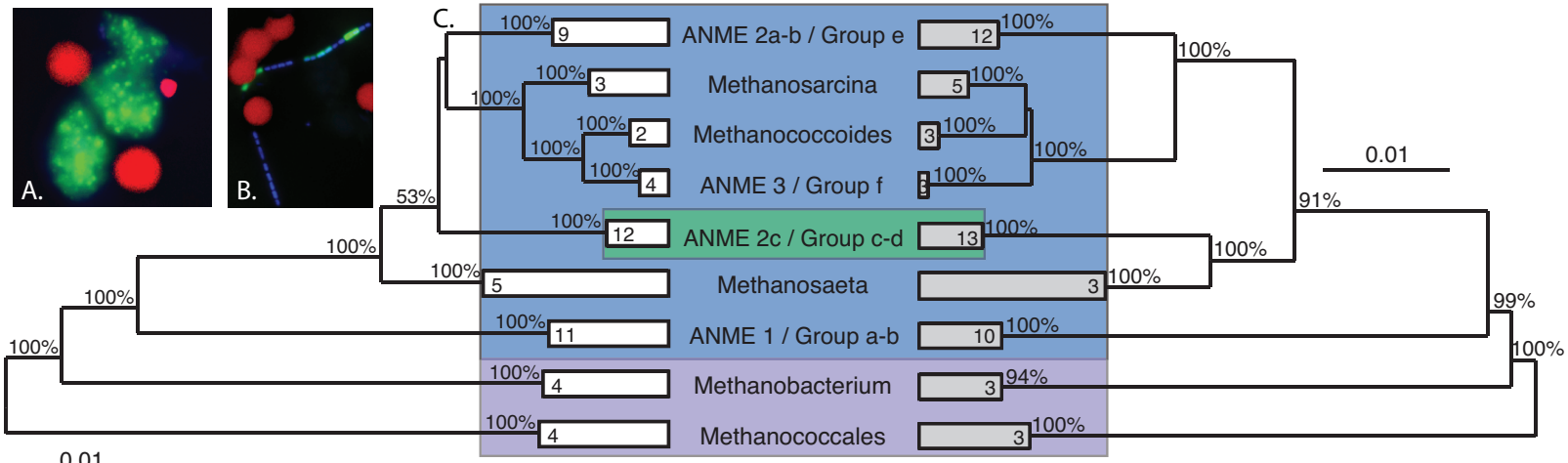
Table 2: CARD-FISH probes, microbial target organisms (Ar, archaea; Ba, bacteria) and associated formamide concentrations (FA %) used in this study with corresponding total DNA yield in nanograms quantified on a Qubit fluorometer from small scale (100 µl) Magneto-FISH captures. The percent bulk yield is calculated by dividing the total DNA recovered for each Magneto-FISH capture by the total bulk DNA recovered from a same volume of paraformaldehyde fixed sediment. Both loss during processing and selectivity of FISH probes used in Magneto-FISH contribute to the estimated percent bulk yield. Seep-1a_1441 DNA concentration was below detection (BD), but only 10% of sample was analyzed due to sample volume constraints. Probe references: Seep-1a_1441 (Schreiber et al. 2010); ANME-2c_760 (Knittel et al. 2005); Eel-MSMX_932 (Boetius et al. 2000); Arc_915 (Stahl & Amann 1991); Delta_495a (Loy et al. 2002).

Sample	*Target Organism(s)*	*FA %*	*Total DNA (ng)*	*Bulk Yield (%)*
ANME-2c_760	Ar, ANME subgroup 2c	50	0.4	3
Seep-1a_1441	Ba, Desulfobacteraceae subgroup Seep-SRB1a	40	BD	-
Eel-MSMX_932	Ar, General ANME	35	0.9	8
Delta_495a	Ba, General Deltaproteobacteria	25	0.8	7
Arc_915	Ar, General Domain-level Archaea	25	1.2	11
Bulk sediment	-	-	11.0	-

To evaluate Magneto-FISH enrichment of target species, clone libraries for both archaeal 16S rRNA and methyl-coenzyme M reductase alpha subunit (*mcrA*) were constructed (Figure 2). *mcrA*

encodes for an enzymatic step common to methanogenic and methanotrophic archaea (Hallam et al. 2003; Luton et al. 2002). Conserved regions can be used as a measure of archaeal diversity in methane seeps, with similar tree topology to archaeal 16S rRNA (Hallam et al. 2003; Hallam et al. 2011; Luton et al. 2002). Parallel analysis of 16S rRNA and metabolic gene diversity in Magneto-FISH capture experiments using complementary (or nested; Amann et al. 1995) suites of FISH probes with differing specificities can assist in evaluating the affiliation of specific metabolic genes with a 16S rRNA phylotype. Results from five independent Magneto-FISH capture experiments, using different probes on the same starting material, recovered the predicted level of archaeal diversity, based on the specificity of the capture probe. For example, archaeal 16S rRNA diversity recovered from ANME-2c_760 and Seep-1a_1441 Magneto-FISH experiments were 100% affiliated with the ANME-2c group, with parallel *mcrA* analysis recovering 95% and 100% of *mcrA* groups c/d, respectively. The *DSS*-affiliated Seep-SRB1a group has been shown in environmental FISH surveys to predominately pair with ANME-2c (Schreiber et al. 2010). The abundance of ANME-2c in both the archaeal 16S rRNA and *mcrA* gene surveys from the SRB targeted Seep-1a_1441 capture is consistent with these findings. These experiments support the results from Pernthaler et al. (2008), demonstrating that high specificity can be achieved with Magneto-FISH. These data also demonstrate the ability to corroborate a microbial association hypothesis, such as ANME-2c/ Seep-SRB1a (Schreiber et al. 2010), with complementary Magneto-FISH experiments.

Figure 2: A. CARD-FISH epifluorescent image of an ANME-2c (FITC) aggregate counterstained with DAPI. B. CARD-FISH epifluorescent image of an ANME-1 (FITC) rod chain counterstained with DAPI. In both images cy3 was over-exposed to show beads (beads are 5 µm for scale). C. Consensus trees of Archaeal 16S rRNA (white boxes) and *mcrA* (gray boxes) genes with ANME clade (16S rRNA) and Group (*mcrA*) names separated by a slash. All other group names apply to both trees. The target range of CARD-FISH probes ANME-2c_760 and Seep1a_1441 (green), Eel_932 and Delta_495 (blue), and Arc_915 (purple) are indicated in the trees and table. D. Table includes percent of total archaeal clones from each Magneto-FISH capture for each archaeal group. *Thermoplasmata* was not included in the trees but is shown in the table to demonstrate the full diversity recovered. No *mcrA* group a-b or f were recovered from Magneto-FISH or Bulk samples, and are not included in the table. *mcrA* clone libraries were not constructed for the Arc_915 Magneto-FISH capture.



D.

	Archaeal 16S rRNA								*mcrA*		
	Methanomicrobia					**Thermoplasmata**			**Methanomicrobia**		
Sample	*2c*	*2b*	*2a*	*1*	*3*	*MBG-D*	*Other*	*Total Clones*	*c-d*	*e*	*Total Clones*
ANME 2c 760	100	0	0	0	0	0	0	48	95	5	20
Eel 932	81	0	19	0	0	0	0	48	100	0	24
Arc915	44	0	4	38	0	4	10	48	-	-	-
Seep1a-1441	100	0	0	0	0	0	0	48	100	0	23
Delta 495	42	42	0	17	0	0	0	48	63	37	19
Bulk	34	10	10	41	2	2	0	41	75	25	24

The interesting pattern of ANME-2 diversity in the more general Eel-MSMX_932 and Delta_495a Magneto-FISH samples is another example of the more nuanced information that can be recovered by this technique. Note that the simplified trees in Figure 2 do not resolve the distinction between ANME-2a and 2b; ANME-2b sequences form a coherent clade related to the ANME-2a group (see Figure 2 in Orphan et al. 2001). While ANME-2a and ANME-2b were equally represented in the bulk sediment diversity survey, these closely related archaeal groups showed differential distribution in the Eel-MSMX_932 and Delta_495a Magneto-FISH captures. While these two probes are expected to have similar levels of target group specificity in this system, ANME-2a was not detected in Eel-MSMX_932 samples and conversely ANME-2b was absent in the Delta_495a capture. The Eel-MSMX_932 probe was designed to target all Eel River Basin clones affiliated with the order *Methanosarcinales* (Boetius et al. 2000), but archaeal 16S rRNA ANME-2 diversity contained ANME-2c (81%) and ANME-2a (19%) sequences and no 2b. The Delta_495a Magneto-FISH capture, selecting for general *Deltaproteobacteria*, recovered an equal number of ANME-2c and ANME-2b clones (42%), as well as 17% affiliated with ANME-1, but no 2a. This would suggest that, in this sample, ANME-2c and ANME-2b might be more likely to form associations with *Deltaproteobacteria* than ANME-2a. These hypotheses can be tested with independent FISH hybridization experiments with the original sediment sample (see Pernthaler et al. 2008).

Magneto-FISH can also aid in correlating diagnostic metabolic genes (e.g. *mcrA, dsrAB, aprA, nifH,* etc.) to 16S rRNA identity. Since 42% of the clones in the Delta_495a capture were ANME-2b, *mcrA* sequences that are distinct from the previously described ANME-1 group a-b or ANME-2c affiliated group c-d, may be associated with ANME-2b, a currently undefined *mcrA* group designated here as e'. The bulk sediment distribution within ANME-2 archaeal 16S rRNA sequences alone is 64% - 2c, 18% - 2b, 18% - 2a. The Delta_495a ANME-2 archaeal 16S rRNA distribution is 50% - 2c and 50% - 2b. The bulk sediment distribution of ANME-2 *mcrA* sequences is 75% c-d (2c), 4% e' (2b), 21% e (2a). The Delta_495a distribution of ANME-2 *mcrA* sequences is 63% c-d (2c), 37% e' (2b), 0% e (2a). Since all three ANME-2 groups are found in both archaeal 16S rRNA and *mcrA* clone libraries for bulk sediment, but only 2c (c-d) and 2b (e') are found in Delta_495a there are multiple lines of evidence to support the hypothesis of group e' *mcrA*. It should also be noted that the *mcrA* primers are not complementary to the majority of

ANME-1 sequences, so investigation of ANME-1 correlations between archaeal 16S rRNA and *mcrA* was not possible.

Optimization for Metagenomics

Advances in library preparation and high throughput sequencing protocols have significantly lowered the required amount of DNA for metagenomics (as low as 1 ng DNA with the Nextera XT). However, our small-scale Magneto-FISH captures yield DNA in amounts that are still below current thresholds without including a post DNA amplification (e.g. Multiple displacement amplification, MDA), similar to that used in single cell genomics ((Woyke et al. 2011) and other chapters in this volume) and used in the Magneto-FISH ANME-2c metagenome (Pernthaler et al. 2008).

To determine where the protocol could be optimized to increase recovery and DNA yield, we evaluated the losses associated with the different steps of the Magneto-FISH protocol. The Magneto-FISH cell retention efficiency was assessed by extracting DNA from wash step supernatants during a large-scale Magneto-FISH ANME-2c_760 capture (Table 3). The DNA concentration of the supernatants was then compared to the amount of DNA extracted from PFA fixed bulk sediment of the same initial volume (3 ml slurry). We estimate that ~6% total DNA is lost during the initial liquid CARD-FISH hybridization. An additional 28% is lost after the antibody (IgG) incubation, which can be improved by increasing the speed during centrifugation (discussed in methods). The sample remaining in the post-capture wash is due to both intended (selectivity from magnetic capture) and unintended (poor hybridization and/or unsuccessful magnetic capture) losses. The DNA yield from Magneto-FISH before the magnetic capture step can be estimated by adding the DNA recovered from the post Magneto-FISH wash (430 ng) to the yield from Magneto-FISH sample (145 ng) for a total of 575 ng. Using the specific ANME-2c_760 capture probe, the DNA yield from ANME2c-760 Magneto-FISH is 25% of this estimated total yield. As 33% of the recovered bulk sediment clones are ANME-2c, this is close to the expected level of selectivity. Assuming ANME archaea are the dominant archaea and about 1/3 of the total microbial assemblage based on the ANME:SRB ratio of 1:3 from other Hydrate Ridge studies (Nauhaus et al. 2007; Orphan et al. 2009), and 1/3 of those archaea are ANME-2c (bulk clone library results, Figure 2D), then 1/6 of the bulk sediment extracted DNA would result in a

theoretical yield of 166 ng. The experimental ANME-2c DNA yield (144.6 ng) is 87% of this theoretical yield.

Table 3: DNA recovered from different stages of a large-scale ANME 2c-760 Magneto-FISH sample to examine losses and selectivity. Initial sample was from 3 ml of PFA fixed slurry in EtOH. The percent bulk yield is calculated by dividing the total DNA recovered at each Magneto-FISH step (accounting wash volume differences) by the total bulk DNA recovered from a same volume of paraformaldehyde fixed sediment. *ANME2c-760, post capture sample* is target cells attached to beads at the end of the protocol. *ANME2c-760, post liquid CARD-FISH* is a 50 ml 1x PBS wash supernatant. *ANME-2c_760, post IgG* is the supernatant after 300 x*g* spin to remove remaining anti-body. *ANME-2c_760, post capture wash* is the sediment and Buffer1 removed after the first wash when the sample is in the magnetic holder (remaining non-target cells).

Sample	Total DNA (ng)	Bulk Yield (%)
ANME-2c_760, post capture sample	144.6	14
ANME-2c_760, post liquid CARD-FISH	63.4	6
ANME-2c_760, post IgG	276.3	28
ANME-2c_760, post capture wash	430.5	43
Bulk	1000.0	-

We also examined DNA extraction efficiency by testing a range of methods to improve cell lysis, removal of formalin crosslinks, and losses during DNA precipitation. As discussed in the methods, implementation of an extended heating step was found to reduce PFA crosslinking issues and yielded the greatest improvement to DNA extraction efficiency. The use of conventional organic extraction with phenol:choloroform resulted in higher yields than tested kit protocols (PowerSoil DNA Isolation kit, MO BIO). Recovery of DNA after ethanol precipitation was enhanced by the use of ammonium acetate and linear acrylamide at 0^{o}C (Crouse & Amorese 1987). The theoretical yield of bulk sediment DNA per ml slurry is 10^{-6} g per ml. This is based on 10^{7} aggregates per ml sediment slurry (calculated for this study) and estimates of 10^{2} cells per aggregate (Nauhaus et al. 2007) and 10^{-15} DNA per cell (Button & Robertson 2001; Simon & Azam 1989). Although this calculation does not account for single cells that also contribute to the bulk DNA, single cells are estimated to be 10% or less of the total biomass at Hydrate Ridge (Nauhaus et al. 2007). This theoretical yield is the same order of magnitude as the bulk sediment DNA (experimental) yield of 1000 ng per ml sediment slurry, indicating efficient DNA extraction.

The age of the fixed sample (time since fixation) can also impact the success of the Magneto-FISH capture and DNA recovery. Freshly fixed samples are recommended, when possible. We also evaluated ethanol as an alternative fixative to reduce cross-linking issues during DNA recovery. While CARD-FISH signals were not as bright, bead association was successful and expected clone diversity was recovered. Fixative choice and strength are recommended optimization areas for application of Magneto-FISH to other systems.

We also evaluated the ability of the Magneto-FISH procedure to meet metagenomic library preparation DNA concentration requirements without MDA amplification, by scaling up starting sample volume (large scale Magneto-FISH prep). This large-scale prep is similar to the procedure originally reported in Pernthaler et al. (2008) and outlined in Schattenhofer and Wendeberg (Schattenhofer & Wendeberg 2011) with a few modifications to the magnetic capture and washing steps (described in methods). In the large-scale Magneto-FISH prep, 3 ml of sediment slurry was used instead of 0.1 ml. From this volume of slurry, 48.2 ng DNA per ml slurry was obtained using the ANME-2c_760 specific probe. This is almost 14 times more DNA than a small scale ANME-2c_760 capture, and enough DNA for library preparation using the Nextera XT kit (minimum 1 ng) for Illumina miseq or highseq sequencing. However, the gain in total DNA yield also corresponded with a decrease in specificity. Only 53% of the 16S rRNA phylotypes associated with ANME-2c, compared with 100% in small scale Magneto-FISH captures. The scaled up protocol is still useful for enrichment of the target population, with 33% of the archaeal diversity associated with the ANME-2c target relative to 20% in the bulk sediment. For the larger volume Magneto-FISH protocols, the incorporation of more extensive washing procedures using a separatory funnel apparatus may aid in the removal of contaminating particles and enhance enrichment of the microbial target, as described in Pernthaler et al. (2008) and Schattenhofer & Wendeberg (2011).

Sample specificity and DNA yield should therefore be optimized for downstream needs; if high specificity is required then pooling many small-scale reactions is recommended, otherwise one large-scale reaction may be sufficient. It is also recommended that any samples that need be compared are run together with the same conditions and reagents to reduce any methodological variation.

Optimization for Other Environmental Systems

This Magneto-FISH protocol was developed and optimized for sediment-associated aggregated microorganisms, so optimal application to other systems likely requires adjustments to the liquid CARD-FISH protocol and washing steps for optimal cell recovery. Schattenhofer & Wendeberg (2011) reported enrichment of single SRB cells from hydrocarbon contaminated sediment using a Magneto-FISH protocol similar to Pernthaler et al. (2008). Schattenhofer & Wendeberg (2011) incubated cells with magnetic beads already labeled with antibodies, which may reduce single cell loss during antibody wash steps in the method described here.

To evaluate the method presented here for single cell Magneto-FISH, we focused on ANME-1. At Hydrate Ridge, ANME-1 are found predominately as single cells or chains of single cells rather than in association with SRB and have a distinctive rod-shaped morphology (Knittel et al. 2005). When using the general Arc_915 probe to target all archaea in small-scale Magneto-FISH experiments, we were able to recover ANME-1 phylotypes at bulk sediment clone abundance. We also observed single cells and chains attached to beads indicating the potential to enrich for non-aggregated cell types using this Magneto-FISH method. We then tried Magneto-FISH with an ANME-1 specific probe to select for a single cell population. We used ANME-1_350 (Boetius et al. 2000) with 30% formamide. We confirmed single cells and chains attached to beads by microscopy (Figure 2B). However, we did not recover quantifiable amounts of DNA and clone abundances were below bulk sediment ratios. Since ANME-1 represented 44% of the recovered archaeal bulk sediment diversity, this should not be due to issues with targeting too small a population.

A possible explanation is that more specific probes are more successful if they work at a higher stringency. When testing Magneto-FISH with without adding probe or adding non-sense probes at 5-10% formamide, it is possible to collect non-specifically bound aggregates. Non-specific capture was confirmed by microscopy (beads attached to aggregates without any CARD signal) and DNA extraction yields. DNA yields from these samples was below or near the limit of detection, but similar to the DNA concentration of ANME-2c_760 and Seep-1a_1441 samples. However, ANME-2c and Seep-1a captures return only the expected single species and do not show signs of non-specific binding. ANME-2c_760 (50%) and Seep-1a_1441 (40%) probes had higher

formamide concentrations than ANME-1_350 (30%). Only less specific probes such as Arc_915 and Delta_495a (25% formamide) returned the expected population at lower formamide concentrations. Optimization of Magneto-FISH for other systems and/or non-aggregate forming populations may be more successful when utilizing probes with targeted, high specificities.

Summary

Magneto-FISH provides a method to target microbial associations from environmental samples for metagenomic and other molecular analyses with high specificity. It is adaptable to a range of target populations within a system, working from the vast array of already vetted FISH probes or developing new ones. It is also an affordable technique since it does not require any special training or equipment beyond the contents of a normal microbiology laboratory. While the method was designed for ANME-2 aggregates and associated bacteria, it can be applied to and optimized for a range of microbial systems utilizing the recommendations described herein. By enriching for associations prior to metagenomic analysis, the genetic information obtained is for a working partnership that may otherwise be lost in a bulk environmental analysis. This middle ground will be invaluable in the effort to better understand all levels at which microbes function in an environment, and in particular in understanding how microbial associations on small scales reflect larger scale chemical and nutrient cycling.

Acknowledgements

Thanks to Annelie (Pernthaler) Wendeberg, Rachel Poretsky, Joshua Steele, Stephanie Cannon, Jen Glass, Kat Dawson, Hiroyuki Imachi, and the Caltech Genomics Center. Funding for this work was provided by the Gordon and Betty Moore foundation and a DOE early career grant (to VJO) and NIH/NRSA training grant 5 T32 GM07616 (to ETR).

References

Amann RI, Binder BJ, Olson RJ, Chisholm SW, Devereux R, and Stahl DA. 1990. Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Applied and Environmental Microbiology* 56:1919-1925.

Amann RI, Ludwig W, and Schleifer KH. 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiological Reviews* 59:143-169.

Ashkin A. 1997. Optical trapping and manipulation of neutral particles using lasers. *Proceedings of the National Academy of Sciences* 94:4853-4860.

Boetius A, Ravenschlag K, Schubert CJ, Rickert D, Widdel F, Gieseke A, Amann R, Jorgensen BB, Witte U, and Pfannkuche O. 2000. A marine microbial consortium apparently mediating anaerobic oxidation of methane. *Nature* 407:623-626.

Button D, and Robertson BR. 2001. Determination of DNA content of aquatic bacteria by flow cytometry. *Applied and Environmental Microbiology* 67:1636-1645.

Crouse J, and Amorese D. 1987. Ethanol precipitation: ammonium acetate as an alternative to sodium acetate. *Focus* 9:3-5.

DeLong EF. 1992. Archaea in coastal marine environments. *Proc Natl Acad Sci U S A* 89:5685-5689.

Fike DA, Gammon CL, Ziebis W, and Orphan VJ. 2008. Micron-scale mapping of sulfur cycling across the oxycline of a cyanobacterial mat: a paired nanoSIMS and CARD-FISH approach. *ISME J* 2:749-759.

Gilbert MTP, Haselkorn T, Bunce M, Sanchez JJ, Lucas SB, Jewell LD, Marck EV, and Worobey M. 2007. The Isolation of Nucleic Acids from Fixed, Paraffin-Embedded Tissues-Which Methods Are Useful When? *PLoS ONE* 2:e537. 10.1371/journal.pone.0000537

Hales BA, Edwards C, Ritchie DA, Hall G, Pickup RW, and Saunders JR. 1996. Isolation and identification of methanogen-specific DNA from blanket bog peat by PCR amplification and sequence analysis. *Applied and Environmental Microbiology* 62:668-675.

Hall BG. 2004. *Phylogenetic trees made easy: a how-to manual*: Sinauer Associates Sunderland.

Hallam SJ, Girguis PR, Preston CM, Richardson PM, and DeLong EF. 2003. Identification of methyl coenzyme M reductase A (mcrA) genes associated with methane-oxidizing archaea. *Applied and Environmental Microbiology* 69:5483-5491.

Hallam SJ, Page AP, Constan L, Song YC, Norbeck AD, Brewer H, and Pasa-Tolic L. 2011. 4 Molecular Tools for Investigating ANME Community Structure and Function. *Methods in Enzymology* 494:75.

Holler T, Widdel F, Knittel K, Amann R, Kellermann MY, Hinrichs KU, Teske A, Boetius A, and Wegener G. 2011. Thermophilic anaerobic oxidation of methane by marine microbial consortia. *ISME J* 5:1946-1956. 10.1038/ismej.2011.77

Imyanitov EN, Suspitsin EN, Buslov KG, Kuligina ESh BE, Togo A, and Hanson K. 2006. Isolation of nucleic acids from paraffin-embedded archival tissues and other difficult sources. *The DNA Book: Protocols and Procedures for the Modern Molecular Biology Laboratory Sudbury, MA, Jones and Bartlett Publishers*:85-97.

Kalyuzhnaya M, Lidstrom M, and Chistoserdova L. 2008. Real-time detection of actively metabolizing microbes by redox sensing as applied to methylotroph populations in Lake Washington. *The ISME Journal* 2:696-706.

Knittel K, Boetius A, Lemke A, Eilers H, Lochte K, Pfannkuche O, Linke P, and Amann R. 2003. Activity, distribution, and diversity of sulfate reducers and other bacteria in sediments above gas hydrate (Cascadia Margin, Oregon). *Geomicrobiology Journal* 20:269-294.

Knittel K, Lösekann T, Boetius A, Kort R, and Amann R. 2005. Diversity and distribution of methanotrophic archaea at cold seeps. *Applied and Environmental Microbiology* 71:467-479.

Lane DJ. 1991. 16S/23S rRNA sequencing. In: Stackebrandt EaG, M., ed. *Nucleic acid techniques in bacterial systematics*. Chichester, England: John Wiley & Sons, 115-175.

Löesekann T, Knittel K, Nadalig T, Fuchs B, Niemann H, Boetius A, and Amann R. 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.

Loy A, Lehner A, Lee N, Adamczyk J, Meier H, Ernst J, Schleifer K-H, and Wagner M. 2002. Oligonucleotide Microarray for 16S rRNA Gene-Based Detection of All Recognized Lineages of Sulfate-Reducing Prokaryotes in the Environment. *Applied and Environmental Microbiology* 68:5064-5081. 10.1128/aem.68.10.5064-5081.2002

Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar, Buchner A, Lai T, Steppi S, Jobb G, Förster W, Brettske I, Gerber S, Ginhart AW, Gross O, Grumann S, Hermann S, Jost R, König A, Liss T, Lüßmann R, May M, Nonhoff B, Reichel B, Strehlow R, Stamatakis A, Stuckmann N, Vilbig A, Lenke M, Ludwig T, Bode A, and Schleifer K-H. 2004. ARB: a software environment for sequence data. *Nucleic Acids Research* 32:1363-1371. 10.1093/nar/gkh293

Luton PE, Wayne JM, Sharp RJ, and Riley PW. 2002. The mcrA gene as an alternative to 16S rRNA in the phylogenetic analysis of methanogen populations in landfill. *Microbiology* 148:3521-3530.

Melin J, and Quake SR. 2007. Microfluidic large-scale integration: the evolution of design rules for biological automation. *Annu Rev Biophys Biomol Struct* 36:213-231.

Nauhaus K, Albrecht M, Elvert M, Boetius A, and Widdel F. 2007. In vitro cell growth of marine archaeal,Äêbacterial consortia during anaerobic oxidation of methane with sulfate. *Environmental Microbiology* 9:187-196.

Neef A, Zaglauer A, Meier H, Amann R, Lemmer H, and Schleifer KH. 1996. Population analysis in a denitrifying sand filter: conventional and in situ identification of Paracoccus spp. in methanol-fed biofilms. *Appl Environ Microbiol* 62:4329-4339.

Niemann H, Loesekann T, de Beer D, Elvert M, Nadalig T, Knittel K, Amann R, Sauter EJ, Schluter M, and Klages M. 2006. Novel microbial communities of the Haakon Mosby mud volcano and their role as a methane sink. *Nature* 443:854-858.

Orphan VJ, Hinrichs KU, Ussler W, Paull CK, Taylor LT, Sylva SP, Hayes JM, and Delong EF. 2001. Comparative analysis of methane-oxidizing archaea and sulfate-reducing bacteria in anoxic marine sediments. *Applied and Environmental Microbiology* 67:1922-1934.

Orphan VJ, House CH, Hinrichs K-U, McKeegan KD, and DeLong EF. 2002. Multiple archaeal groups mediate methane oxidation in anoxic cold seep sediments. *Proceedings of the National Academy of Sciences* 99:7663-7668.

Orphan VJ, Turk KA, Green AM, and House CH. 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. 10.1111/j.1462-2920.2009.01903.x

Pernthaler A, Dekas AE, Brown CT, Goffredi SK, Embaye T, and Orphan VJ. 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. 10.1073/pnas.0711303105

Pernthaler A, and Orphan VJ. 2010. Process for separating microorganisms. Google Patents.

Pernthaler A, Pernthaler J, and Amann R. 2002. Fluorescence In Situ Hybridization and Catalyzed Reporter Deposition for the Identification of Marine Bacteria. *Applied and Environmental Microbiology* 68:3094-3101. 10.1128/aem.68.6.3094-3101.2002

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

Rodrigue S, Malmstrom RR, Berlin AM, Birren BW, Henn MR, and Chisholm SW. 2009. Whole genome amplification and de novo assembly of single bacterial cells. *PLoS ONE* 4:e6864.

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S, Larget B, Liu L, Suchard MA, and Huelsenbeck JP. 2012. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice across a Large Model Space. *Systematic Biology*. 10.1093/sysbio/sys029

Šafařík I, and Šafaříková M. 1999. Use of magnetic techniques for the isolation of cells. *Journal of Chromatography B* 722:33-53.

Sambrook J, and Russell DW. 2001. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Schattenhofer M, and Wendeberg A. 2011. Capturing Microbial Populations for Environmental Genomics. *Handbook of Molecular Microbial Ecology I*: John Wiley & Sons, Inc., 735-740.

Schreiber L, Holler T, Knittel K, Meyerdierks A, and Amann R. 2010. Identification of the dominant sulfate-reducing bacterial partner of anaerobic methanotrophs of the ANME-2 clade. *Environmental Microbiology* 12:2327-2340. 10.1111/j.1462-2920.2010.02275.x

Simon M, and Azam F. 1989. Protein content and protein synthesis rates of planktonic marine bacteria. *Marine ecology progress series Oldendorf* 51:201-213.

Stahl DA, and Amann RI. 1991. Development and application of nucleic acid probes in bacterial systematics. In: E. Stackebrandt MG, ed. *Sequencing and Hybridization Techniques in Bacterial Systematics*. Chichester, England: John Wiley & Sons Ltd.

Thompson JR, Marcelino LA, and Polz MF. 2002. Heteroduplexes in mixed-template amplifications: formation, consequence and elimination by ‚Äòreconditioning PCR‚Äô. *Nucleic Acids Research* 30:2083-2088.

Waldron PJ, Petsch ST, Martini AM, and Nüsslein K. 2007. Salinity Constraints on Subsurface Archaeal Diversity and Methanogenesis in Sedimentary Rock Rich in Organic Matter. *Applied and Environmental Microbiology* 73:4171-4179. 10.1128/aem.02810-06

Woyke T, Sczyrba A, Lee J, Rinke C, Tighe D, Clingenpeel S, Malmstrom R, Stepanauskas R, and Cheng J-F. 2011. Decontamination of MDA reagents for single cell whole genome amplification. *PLoS ONE* 6:e26161.

Yilmaz S, Haroon MF, Rabkin BA, Tyson GW, and Hugenholtz P. 2010. Fixation-free fluorescence in situ hybridization for targeted enrichment of microbial populations. *ISME J* 4:1352-1356.

Yoon HS, Price DC, Stepanauskas R, Rajah VD, Sieracki ME, Wilson WH, Yang EC, Duffy S, and Bhattacharya D. 2011. Single-Cell Genomics Reveals Organismal Interactions in Uncultivated Marine Protists. *Science* 332:714-717. 10.1126/science.1203163

C h a p t e r 2

CHARACTERIZATION OF MICROBIAL ASSOCIATIONS WITH METHANOTROPHIC ARCHAEA AND SULFATE-REDUCING BACTERIA THROUGH STATISTICAL COMPARISON OF NESTED MAGNETO-FISH ENRICHMENTS

Abstract

Methane seep systems along continental margins host diverse and dynamic microbial assemblages, sustained in large part through the microbially mediated process of sulfate-coupled Anaerobic Oxidation of Methane (AOM). This methanotrophic metabolism has been linked to a consortia of anaerobic methane-oxidizing archaea (ANME) and sulfate-reducing bacteria (SRB). These two groups are the focus of numerous studies; however, less is known about the wide diversity of other seep associated microorganisms. We selected a hierarchical set of FISH probes targeting a range of *Deltaproteobacteria* diversity. Using the Magneto-FISH enrichment technique, we then magnetically captured CARD-FISH hybridized cells and their physically associated microorganisms from a methane seep sediment incubation. DNA from nested Magneto-FISH experiments was analyzed using Illumina tag 16S rRNA gene sequencing (iTag). Enrichment success and potential bias with iTag was evaluated in the context of full-length 16S rRNA gene clone libraries, CARD-FISH, functional gene clone libraries, and iTag mock communities. We determined commonly used Earth Microbiome Project (EMP) iTAG primers introduced bias in some common methane seep microbial taxa that reduced the ability to directly compare OTU relative abundances within a sample, but comparison of relative abundances between samples (in nearly all cases) and whole community-based analyses were robust. The iTag dataset was subjected to statistical co-occurrence measures of the most abundant OTUs to determine which taxa in this dataset were most correlated across all samples. Many non-canonical microbial partnerships were statistically significant in our co-occurrence network analysis, most of which were not recovered with conventional clone library sequencing, demonstrating the utility of combining Magneto-FISH and iTag sequencing methods for hypothesis generation of associations within complex microbial communities. Network analysis pointed to many co-occurrences containing putatively heterotrophic, candidate phyla such as OD1, Atribacteria, MBG-B, and Hyd24-12 and the potential for complex sulfur cycling involving *Epsilon-*, *Delta-*, and *Gammaproteobacteria* in methane seep ecosystems.

Introduction

A central goal in microbial ecology is identifying and understanding microbial interactions in the environment. This goal can be addressed at many scales from statistical analyses of entire ecosystems (Barberán et al. 2012; Malfatti & Azam 2010; Ruff et al. 2015; Steele et al. 2011; Sunagawa et al. 2015) to high resolution image analysis of specific symbioses (Malfatti & Azam 2010; McGlynn et al. 2015; Orphan 2009; Orphan et al. 2001b; Wegener et al. 2015). Previous studies have shown that complex datasets can be distilled to determine primary ecosystem drivers, such as temperature, as main predictors of community variability (Sunagawa et al. 2015). In addition to correlating microbial patterns to environmental factors, interspecies interactions can be evaluated with methods such as co-occurrence analysis (Friedman & Alm 2012). Statistical significance of co-occurrence can be assessed at scales ranging from the entire genome to the operational taxonomic unit (OTU) (Barberán et al. 2012; Chaffron et al. 2010).

Many physical separation methods have been developed to partition complex microbial assemblages before analysis, including fluorescence-activated flow sorting (Amann et al. 1990; Yilmaz et al. 2010), optical trapping (Ashkin 1997), microfluidics (Melin & Quake 2007), and immunomagnetic beads (Pernthaler et al. 2008; Šafařík & Šafaříková 1999) that use characteristics of interest such as phylogenetic identity (Fluorescence In-Situ Hybridization; FISH) or activity (Berry et al. 2015; Hatzenpichler & Orphan 2015; Hatzenpichler et al. 2014; Kalyuzhnaya et al. 2008; Wegener et al. 2012).

Here we combine Magneto-FISH and Illumina Tag (iTag) sequencing utilizing the Earth Microbiome Project (EMP) universal primer set (Caporaso et al. 2012). The Magneto-FISH method was originally developed to enrich for and characterize multi-species microbial associations in environmental samples (Pernthaler et al. 2008). This method consists of a liquid CARD (CAtalyzed Reporter Deposition)-FISH reaction as a 16S rRNA gene identity-based selection mechanism followed by an immunomagnetic sediment matrix separation mechanism to target specific phylogenetic groups in conjunction with their physically associated microbial partners. By combining this method for phylogenetically targeted physical separation with high throughput amplicon sequencing, we can compare an array of associated microbial communities in

parallel, with replicates. This provides statistical power in deriving microbial associations from complex sediment community assemblages, and thereby improving hypothesis development.

Anaerobic methane-oxidizing (ANME) archaea and sulfate-reducing *Deltaproteobacteria* (SRB) are the predominant community members discussed in methane seep literature and form syntrophic partnerships in physical associations, termed "aggregates" or consortia (Boetius et al. 2000; Green-Saxena et al. 2014; Knittel et al. 2003; Orphan et al. 2001a; Schreiber et al. 2010). Since physical association appears to be an important element for consortia activity (McGlynn et al. 2015; Wegener et al. 2015), methods like Magneto-FISH are ideal for probing this system because target organisms are separated from the sediment matrix along with their physically associated partners. A hierarchical probe set was chosen targeting *Deltaproteobacteria* and their ANME partners to create nested Magneto-FISH enrichments from methane seep sediment incubations under methane headspace. This method allows us to examine potential physical associations between ANME and SRB taxa and other microorganisms using co-occurrence statistical methods applied to iTag sequences from nested Magneto-FISH enrichments.

ANME have been broadly divided into three separate groups, which can be further subdivided into ANME-1a, 1b, 2a, 2b, 2c, and 2d, and 3. ANME-1 archaea are a unique order-level lineage within the *Euryarchaeota,* between the *Methanomicrobiales* and the *Methanosarcinales*, known to associate with sulfate-reducing bacteria, but obligately associated lineages have yet to be defined. ANME-2 archaea, within the order *Methanosarcinales*, commonly form associations with *Desulfosarcina*/*Desulfococcus*-related (DSS) sulfate-reducing *Deltaproteobacteria* (Boetius et al. 2000; Orphan et al. 2001a; Schreiber et al. 2010). They have also been found in association with *Desulfobulbus*-related (DSB) *Deltaproteobacteria* in the same environments, where geochemical factors have been suggested as a possible explanation for partner differentiation (Green-Saxena et al. 2014). ANME-2a/b and ANME-2c both predominately associate with a subgroup of DSS, SEEP-SRB1 (Schreiber et al. 2010), but also form consortia with DSB (Green-Saxena et al. 2014; Pernthaler et al. 2008). ANME-3 has been found in association with *Desulfobulbus*-related *Deltaproteobacteria* (Niemann et al. 2006) and SEEP-SRB1 (Schreiber et al. 2010). These ANME groups have also been observed in the environment without bacterial partners (House et al. 2009; Orphan et al. 2002; Schreiber et al. 2010; Treude et al. 2007). In addition to ANME archaea, other uncultured archaeal lineages commonly recovered from methane seeps include Marine Benthic

Group-D (*Thermoplasmatales*), Deep Sea Archaeal Group / Marine Benthic Group-B (Ruff et al. 2015; Yanagawa et al. 2011), and sometimes methanogens (Orphan et al. 2001a; Ruff et al. 2015; Takano et al. 2013; Vigneron et al. 2015).

Deltaproteobacteria diversity beyond DSS and DSB has also been well described in methane seeps. In addition to SEEP-SRB1, Knittel et al. (2003) define three more *Deltaproteobacteria* clades within *Desulfobulbaceae* (SEEP-SRB2, 3 and 4). Green-Saxena et al. (2014) also described a *Desulfobulbaceae* affiliated seepDBB group in methane seep systems. Bacterial diversity surveys of methane seep habitats frequently report occurrence of other diverse *Proteobacteria* including sulfur oxidizers (*Gammaproteobacteria* and *Epsilonproteobacteria*) and putative heterotrophs (*Alphaproteobacteria* and *Betaproteobacteria*) (Pernthaler et al. 2008; Ravenschlag et al. 1999). Many other bacterial phyla have also been found such as *Firmicutes*, *Thermomicrobia*, *Bacteroidetes*, *Chlorobi*, *Nitrospira*, WS3, OD1, OP11, TM7, and WS6 (Schreiber et al. 2010); *Cytophaga* and *Flavobacteria* (Knittel et al. 2003); *Chloroflexi*, *Atribacteria* (previously Candidate Division JS1), CD12, WS1, OS-K, AC1, and *Planctomycetes* (Yanagawa et al. 2011); and *Acidobacteria* (Ravenschlag et al. 1999). Ruff et al. (2015) indentify *Methanomicrobia*, *Deltaproteobacteria*, Hyd24-12 and Atribacteria as the characteristic 'core' microbial taxa in methane seep ecosystems, as compared to *Gammaproteobacteria*, *Flavobacteria*, *Thermoplasmatales*, and MBG-B taxa that were found in high relative abundance in seeps and other marine ecosystems.

Despite the wealth of bacterial and archaeal diversity in methane seep sediments, little is known about potential associations with ANME/SRB, or associations that do not involve ANME or SRB. Our study utilizes the novel combination of targeted Magneto-FISH enrichment of specific microbial taxonomic groups and iTag sequencing to develop statistically supported co-occurrence microbial networks to address knowledge gaps in our understanding of methane seep microbial communities. Network analysis revealed many novel associations between methane seep *Proteobacteria* taxa and Candidate phyla. The significant co-occurrences observed suggest new avenues for future studies on microbial interactions involved in carbon and sulfur cycling in methane seep systems.

Materials & Methods.

Sample collection and Magneto-FISH

iTag Magneto-FISH enrichments were conducted using a large scale (1 L) incubation of methane seep sediment from Hydrate Ridge North (offshore Oregon, USA) collected in September 2011 at 44°40.02' N 125°6.00' W, from a water depth of 775 m using the ROV *JASON II* and the R/V *Atlantis*. Marine sediment was collected using a push core to sample a sulfide-oxidizing microbial mat adjacent to an actively bubbling methane vent. A sediment slurry from the upper 0–15 cm depth horizon of the push core was prepared with 1 volume N_2 sparged artificial seawater to 1 volume sediment, overpressurized with methane (3 bar) and incubated at 8^{o}C in a 1 L Pyrex bottle capped with a butyl rubber stopper until subsampling for Magneto-FISH.

In February 2015, incubation samples were immediately fixed in 0.5 ml sediment aliquots in 2% paraformaldehyde (PFA) for 3 hrs at 4^{o}C. The samples were washed in 50% phosphate-buffered saline (PBS): 50% EtOH, then 75% EtOH: 25% DI water, and resuspended in 2 volumes (1 ml) 100% ethanol. Samples were centrifuged at 1000 × *g* for 1 min between wash steps. After fixation, the Magneto-FISH method first described by Pernthaler et al. (2008) and further optimized by Schattenhofer and Wendeberg (2011) and Trembath-Reichert et al. (2013) was used. Briefly, a liquid CARD-FISH reaction was followed by immunomagnetic bead incubation coupled with anti-fluorecsein attaching magnetic beads to CARD-FISH hybridized aggregates. Samples were then held against magnets and the sediment matrix was washed away, retaining target cells and physically associated microbes in the magnetic portion, as described in Trembath-Reichert et al. 2013. Four previously published FISH probes were used targeting a range of *Deltaproteobacteria* and *Methanomicrobia* (Table 1). A subset of three 0.5 ml aliquots was also immediately frozen before fixation (unfixed bulk sediment), and another three aliquots were frozen after fixation (fixed bulk sediment) for bulk sediment comparison with Magneto-FISH enrichments. Sediment for MSMX-Eel_932 Magneto-FISH metabolic gene analysis was fixed and washed onboard in September 2011, as described above. See methods flow chart provided in Sup Figure 1.

Table 1: FISH probes and primers used in this study. References: (Akerman et al. 2013; Blazejak et al. 2006; Boetius et al. 2000; Caporaso et al. 2012; DeLong 1992; Lane 1991; Loy et al. 2002; Macalady et al. 2006; Manz et al. 1996; Manz et al. 1992; Manz et al. 1998; Neef et al. 1998; Schreiber et al. 2010; von Wintzingerode et al. 1999; Wagner et al. 1998)

Name	*Sequence (5' -> 3')*	*Target*	*FA (%) / Annealing (^{o}C)*
PROBES for Magneto-FISH & CARD-FISH			
DSS_658	TCCACTTCCCTCTCCCAT	Desulfosarcina/ Desulfococcus, Desulfofaba, Desulfofrigus	50
Delta_495a	AGTTAGCCGGTGCTTCCT	Most Deltaproteobacteria and most Gemmatimonadetes	35
Delta_495a-comp	AGTTAGCCGGTGCTTCTT		35
Seep-1a_1441	CCCCTTGCGGGTTGGTCC	Seep-SRB1a	45
MSMX-Eel_932	AGCTCCACCCGTTGTAGT	All ANME groups	35
ANME-1_350	AGTTTTCGCGCCTGATGC	ANME-1	40
Epsi_404	AAAKGYGTCATCCTCCA	Epsilonproteobacteria	30
Gam_42a	GCCTTCCCACATCGTTT	Gammaproteobacteria	35
Gam_42a comp (Bet42a)	GCCTTCCCACTTCGTTT	Betaproteobacteria	35
Pla_46	GACTTGCATGCCTAATCC	Planctomycetes	35
Pla_886	GCCTTGCGACCATACTCCC	Planctomycetes	35
CF_319A	TGGTCCGTGTCTCAGTAC	CFB (Cytophaga, Bacteriodales, Flavobacterium, Sphingobacterium)	35
CF_319B	TGGTCCGTATCTCAGTAC	CFB (mostly Cytophaga)	35
PRIMERS for iTAG			
515F	GTGCCAGCMGCCGCGGTAA	V4 region universal 16S rRNA	55
806R	GGACTACHVGGGTWTCTAAT	V4 region universal 16S rRNA	55
PRIMERS for CLONE LIBRARIES			
Bac27F	AGAGTTTGATYMTGGCTC	Bacterial 16S rRNA	54
U1492R	GGYTACCTTGTTACGACTT	Universal 16S rRNA	54
10-30Fa	TCCGGTTGATCCTGCC	Archaeal 16S	54
Arc958R	YCCGGCGTTGAMTCCAATT	Archaeal 16S	54
DSR1F	ACSCACTGGAAGCACG	dsrAB	61-48
DSR4R	GTGTAGCAGTTACCGCA	dsrAB	61-48
APS_1F	TGGCAGATCATGATYMAYGG	APS reductase	54
APS_4R	GCGCCAACYGGRCCRTA	APS reductase	54
sox527F	TGGTWGGWCAYTGGGAATTTA	sulfate thiol esterase	46
sox1198R	AGAANGTATCTCKYTTATAAAG	sulfate thiol esterase	46

iTag Amplification

For iTag sequencing, ten Magneto-FISH enrichments were performed in parallel using the FISH probes DSS_658 (triplicate), MSMX-Eel_932 (triplicate), SEEP-1a_1441 (duplicate), Delta_495a + Delta_495a competitor (duplicate). Magneto-FISH enrichments and bulk sediment samples were resuspended in 650 µl solution PM1 and transferred to silica tubes from the PowerMicrobiome RNA Isolation Kit (MoBio). This kit was chosen based on manufacturer recommendation for formalin-fixed sediment samples, with the added capability to co-elute RNA if desired. 6.5 µl of beta-mercaptethanol was added, and samples were mechanically lysed in a bead beater (FastPrepFP120, ThermoElectronCorp.) for 45 s at setting 5.5 and incubated at 65^{o}C for 3.5 hrs. The remaining steps in the PowerMicrobiome RNA Isolation Kit were followed according to manufacturer instructions (starting at step 5) without any DNase procedures, and eluting in a final volume of 60 µl ultrapure water. DNA extracts were quantified using a Qubit Flurometer and HS dsDNA kit (Invitrogen; Sup Table 1). All but one Magneto-FISH sample had DNA concentrations below detection (<0.5 ng/µl); however, all samples yielded PCR amplicons when viewed on a gel after initial pre-barcoding PCR (30 cycles).

iTag samples were prepared with Earth Microbiome Project (EMP) primers 515f and 806r (Caporaso et al. 2012). An initial amplification of 30 cycles with primers lacking the barcode, linker, pad, and adapter was performed for all samples, in duplicate. Duplicate PCR reactions were pooled and reconditioned for 5 cycles with barcoded primers, for a total of 35 cycles. A master mix of 2X Q5 Hot Start High Fidelity Master Mix (NEB) and 10 µM forward and reverse primers was prepared for a final volume of 15 µl per sample, with 1 µl DNA template. PCRs had an initial 2 min heating step at 98^{o}C, followed by cycles of 10 s 98^{o}C, 20 s 54^{o}C, and 20 s 72^{o}C, and finished with a final extension of 2 min at 72^{o}C. PCR negative controls, substituting ultrapure water for DNA template, were amplified for 40 cycles total. We note that these are not the official recommended reagents or PCR conditions from the EMP, but internal lab tests showed that for 6 out of 9 mock community taxa, recovered sequence relative abundances were more accurate when using Q5 polymerase rather than the recommended Hot Start MasterMix (5-prime). EMP primers were chosen for iTag for cross-comparison between studies, though there is known primer bias within this universal primer set (Parada et al. 2015) and sequencing reactions will always have some inherent variability.

Mock Communities

Four mock communities were prepared with a range of relative proportions of nine common methane seep taxa (Sup Table 2). Full-length 16S rRNA gene plasmids from each taxa listed were quantified by Qubit. Taking into account the plasmid's nucleotide composition and length in order to calculate its molecular weight, plasmids were quantitatively combined in known volumetric fractions to achieve a range of desired mock community compositions. These combined plasmid mixes were diluted to ~1 ng/µL and then prepared according to the same iTag methods as all other samples.

iTag sequence processing

We followed the *mothur* Standard Operating Procedure (SOP) for Illumina MiSeq sequencing of the 16S rRNA gene V4 region, accessed May 2015 and using methods described in Kozich et al. (2013) with UCHIME chimera checking (Edgar et al. 2011). A concatenated file of the *mothur* version of separate archaeal and bacterial SILVA 119 databases (Quast et al. 2013) was used for alignment and classification. Unfixed Bulk Sediment 1 only returned 8% of the average DNA concentration of the other two samples. (Sup Table 1). This sample was removed from statistical analyses because it fails to be a representative of the unfixed bulk sediment community baseline. The mock communities were processed following the "Assessing Error Rates" section of the *mothur* SOP to compute sequencing error rates and spurious OTU rates (Sup Table 4). Additional analysis demonstrating sequence processing did not selectively remove ANME-2c sequences and relative sequence abundances recovered with iTag sequencing of mock communities are provided in Sup Table 3 and Sup Table 2, respectively.

Using *R* version 3.1.3 (R Core Team 2015), an average number of sequences per OTU was calculated from unfixed bulk sediment samples (2 and 3). All OTUs with an average relative sequence abundance below 0.1% in the unfixed bulk sediment were identified and removed from all samples using *mothur*. 135 unique OTUs remained out of 25,354. We also verified that after the 0.1% cutoff was applied, no negative control contaminant OTUs remained. The top 20 OTUs amplified from the no template negative control were classified as, in order of sequence

abundance: Sphingomonas*; Planctomyces*; Escherichia-Shigella*; Staphylococcus; Roseomonas*; Pir4_lineage; Delftia*; Macrococcus; Myxococcales;0319-6G20;unclassified; Planctomyces; Enhydrobacter; Sphingobium*; Caenispirillum; Bacillus*; Pseudoxanthomonas*; Peptoniphilus; Lysobacter; Salinicoccus; Propionibacterium.* Reagent contaminant genera discussed in Salter et al. (2014) are denoted by (*). All samples (including mock community and negative controls) were submitted to the SRA under the accession SAMN03879962, BioSample: SAMN03879962, Sample name: PC47 (5133-5137) mixed slurry.

Gene libraries of the Magneto-FISH samples were prepared as in Trembath-Reichert et al. (2013) using the primers and annealing temperatures listed in Table 1 and TOPO TA Cloning Kit for Sequencing with pCR4-TOPO Vector and One Shot Top 10 chemically competent Escherichia coli (Life Technologies). All full-length 16S rRNA gene sequences were aligned by the SINA online aligner (v 1.2.11; Pruesse et al. 2012) and added using maximum parsimony to the SILVA 119 database (Quast et al. 2013) for classification. A taxonomy-based count table was prepared (sequences per taxa, per sample) and all taxa absent from the bulk sediment library were removed from Magneto-FISH enrichment libraries (for parity with iTag contaminant removal processing). Functional gene sequences were translated using the *EMBOSS* online translation tool (Li et al. 2015), then added to *ARB* (Ludwig et al. 2004) databases for phylogenetic placement and classification. Sequences were submitted to NCBI under the following accession numbers: AprA (KT280505 - KT280517), DsrA (KT280518 - KT280533), McrA (KT280534 - KT280581), Archaeal 16S rRNA gene (KT280582 - KT280632), Bacterial 16S rRNA gene (KT280633 - KT280909), SoxB (KT280910 - KT280928). Gene trees were computed with representative sequences using PhyML 3.0 (Guindon et al. 2010) online execution with defaults on the South of France Bioinformatics platform.

Statistical Analysis

Weighted UniFrac (Lozupone & Knight 2005), Metastats (White et al. 2009), and linear discriminant analysis (LDA) effect size (LEfSe) (Segata et al. 2011) analyses were computed in *mothur* as outlined in the *mothur* SOP. Co-occurrence statistical analyses were run using the table of 135 unique OTUs in the format of sequence counts of each OTU per sample. The program *SparCC* was used to determine significant correlations (Friedman & Alm 2012). This analysis was

run 100 times with default settings, except 10 iterations were used instead of 20. OTUs with *SparCC* correlations above an absolute value of 0.6 with p-values below 0.01 were considered significant. Resulting associations that occurred in at least 50 out of 100 network iterations are provided in Sup Table 5. *Cytoscape* (Shannon et al. 2003) was used to display associations in Figure 1.

CARD-FISH microscopy

A triple CARD-FISH hybridization was performed with bacterial probes listed in Table 1, ANME-1_350 and MSMX-Eel_932. The sample preparation and CARD reaction was performed as per Green-Saxena et al. (2014). After the three CARD reactions, samples were post-stained with DAPI (25 ng/µl). CARD signal within any part of a physically attached group of cells larger than 10 µm was counted as a positive identification. For example, a large EPS matrix that contained many smaller separate ANME-1 and ANME-2 aggregates would count as one positive identification for each clade. This was done to simulate groups that would have been isolated together in a Magneto-FISH enrichment. Since the MSMX-Eel_932 probe also targets the ANME-1 population, only cells with MSMX-Eel_932 signal and no ANME-1_350 signal were recorded as an ANME-2 positive identification to comprehensively target ANME-1, -2, and a bacterial partner in a triple CARD-FISH hybridization set. ANME-3 were not recovered in the iTag dataset and were not considered as potential contributors to MSMX-Eel_932 signal.

Results.

Relative sequence abundance of seep microbiome taxa in 16S rRNA gene iTag and libraries

Relative sequence abundances of the methane seep microbiome characteristic taxa, ANME archaea, *Deltaproteobacteria*, Hyd24-12, and Atribacteria (Ruff et al. 2015), were compared two ways: 1) between iTag and gene library 16S rRNA gene samples to determine how relative sequence abundances differed between sequencing methodologies, and 2) between Magneto-FISH enrichment and bulk sediment to determine taxa-specific relative sequence abundance for each probe (Table 1).

Mock community analysis showed that ANME-2 were always underrepresented in iTag data (0.32-0.81 fold of what was expected), whereas the *Deltaproteobacteria* and ANME-1b were more faithfully represented (Sup Table 2). ANME-1a was consistently over amplified. By normalizing the relative sequence abundance of ANME-2c, -2a/b, and -1a to the abundance of ANME-1b, the most faithfully amplified archaea in the mock community data (Sup Table 2), we could compute a ratio between the average relative sequence abundance in fixed bulk sediment samples between iTag and the archaeal 16S rRNA gene library. ANME-2c (0.04 iTag:clone ratio), ANME-2a/b (0.12), and ANME-1a (0.40) were all less abundant in iTag sequences as compared to the archaeal gene clone library (calculated from values in Table 2). Similarly comparing SEEP-SRB1 to *Desulfobulbus* between the two methods in fixed bulk sediment returns a ratio of 0.41 iTag:clone. Since the iTag methodology recovers far more diversity (e.g. *Desulfobacula*, *Desulfocapsa*, *Desulfoluna*, Atribacteria, and Hyd24-12 were not recovered in the bacterial 16S rRNA gene bulk sediment library), it is expected that the relative sequence abundances of each individual taxon computed from iTag data would be less than from the domain targeted 16S rRNA gene libraries. However, the ANME-2c abundance ratio was an order of magnitude less than ANME-1a and SEEP-SRB1 ratios, and appears to be an extreme case of underestimation in iTag data. There was also variation between Magneto-FISH enrichment replicates, as indicated by the high standard deviations of Magneto-FISH samples as compared to bulk sediment samples. The degree of variation (average standard deviation across all taxa listed) correlated with the specificity of the probe; where Delta_495a had the lowest average standard deviation and Seep-1a_1441 had the highest average standard deviation.

The high relative sequence abundance taxa (>1.5 fold relative sequence abundance increase over fixed bulk sediment; Table 2) in the averaged Seep-1a_1441 iTag Magneto-FISH enrichments were *Desulfoluna* (2.20), SEEP-SRB1 (2.36), Hyd24-12 (3.44), and *Atribacteria* (1.51) (Table 2). The DSS_658 enrichment had fewer high relative sequence abundance taxa with only *Desulfoluna* (4.62), *Spirochaeta* (4.36), and *Atribacteria* (4.80). The Delta_495a enrichment also had three high relative sequence abundance taxa with *Desulfobulbus* (2.52), *Spirochaetae*-uncultured (3.70), and *Atribacteria* (3.02). The MSMX-Eel_932 enrichment had six high relative sequence abundance taxa with *Desulfococcus* (1.85), *Desulfoluna* (8.47), SEEP-SRB1 (1.67), *Spirochaeta* (1.63), Hyd24-12 (1.73), and *Atribacteria* (7.18). Gene library results showed high relative

sequence abundance (>1.5) in both ANME and *Deltaproteobacteria* with DSS_658 and MSMX-Eel_932 enrichments (Table 2). Similar to the bulk sediment, *Desulfobacula*, *Desulfocapsa*, *Desulfoluna*, *Atribacteria* and Hyd24-12 were not recovered in the bacterial 16S rRNA gene Magneto-FISH libraries. MSMX-Eel_932 enriched for SEEP-SRB1 (2.73), SEEP-SRB4 (3.28), *Desulfococcus* (3.82), *Spirochaeta* (1.64), and ANME-2a/b (2.51) in 16S rRNA gene libraries. There was also a slight enrichment of ANME-2c (1.28). The DSS_658 enrichment had high relative sequence abundance for SEEP-SRB1 (1.74), SEEP-SRB2 (2.78), ANME-2c (1.54), and ANME-2a/b (2.24) with iTag, but these same taxa did not have high relative sequence abundance in the gene library. *Spirochaeta* and SEEP-SRB1 had high relative sequence abundance in both iTag and gene libraries for MSMX-Eel_932 enrichments. Relative sequence abundances for all non-core methane seep taxa in iTag samples are included in Table 3, and where Magneto-FISH enrichments of these additional taxa support network co-occurrences they are discussed in network results.

Statistical evaluation of Magneto-FISH enrichment

To statistically compare enrichment microbial communities, we used a suite of statistical tests including: non-parametric T-tests (White et al. 2009), LEfSe (Segata et al. 2011), and UniFrac (Lozupone & Knight 2005). Using the T-test comparison, ten OTUs were significantly ($p<0.001$) different between the bulk sediment and Magneto-FISH samples (when only including OTUs with sequences present in both groups). The taxonomic assignments for these ten OTUs were: WCHB1-69, *Desulfobulbus*, *Thaumarcheota*, ANME-1a, *Bacteroidetes* (VC2.1), ANME-2c, *Caldithrix*, SEEP-SRB1, Candidate Division TA06, and *Gammaproteobacteria* (CS-B046). LEfSe was then used to determine which OTUs were significantly different between Magneto-FISH enrichments and bulk sediment. We found three OTUs were significantly (p-value <0.05) higher in relative sequence abundance in Magneto-FISH samples over bulk sediment with the taxonomies: SEEP-SRB1, *Desulfobulbus*, and *Planctomycetes* (SHA-43).

Weighted UniFrac analysis was used to compare the community composition between Magneto-FISH iTag enrichments. The UniFrac metric represents the fraction of the branch length that is unique to each sample, or unshared between samples, such that a higher ratio means less similar samples. The *Deltaproteobacteria* probe enrichment communities were more similar to each other

	16S rRNA gene (iTAG)																16S rRNA gene (Clone Library)				
	Seep1a_1441			DSS_658			Delta_495a			MSMX-Eel_932			Fixed Bulk		Unfixed Bulk		DSS_658		MSMX-Eel_932		Fixed Bulk
Taxon	Avg.	Stdev.	Rel. Fixed	Avg.	Stdev.	Rel. Fixed	Avg.	Stdev.	Rel. Fixed	Avg.	Stdev.	Rel. Fixed	Avg.	Stdev.	Avg.	Stdev.	24 arc, 41 bac	Rel. Fixed	60 arc, 87 bac	Rel. Fixed	43 arc, 95 bac
ANME-1a	0.07	0.07	0.67	0.04	0.04	0.36		0.01	0.05	0.07	0.01	0.61	0.11	0.02	0.10	0.02	0.08	0.28	0.08	0.28	0.30
ANME-1b	0.11	0.08	0.92	0.09	0.05	0.74	0.12	0.05	0.95	0.15	0.09	1.22	0.12	0.03	0.08	0.01					0.14
ANME-2a/b		0.01	0.19	0.01	0.01	0.31			0.01			0.11	0.02	0.01	0.01		0.42	2.24	0.47	2.51	0.19
ANME-2c			0.01			0.09							0.01	0.01	0.01		0.50	1.54	0.42	1.28	0.33
Desulfobacula																			0.01		
Desulfobulbus	0.08	0.06	1.01	0.11	0.05	1.30	0.20	0.14	2.52	0.03	0.01	0.36	0.08	0.01	0.12	0.01	0.05	0.66	0.06	0.78	0.07
Desulfocapsa	0.02	0.01	1.02			0.16		0.01	0.32	0.02	0.03	1.10	0.01		0.01		0.05				
Desulfococcus	0.03	0.03	0.67	0.03	0.03	0.61	0.03	0.04	0.74	0.08	0.13	1.85	0.04		0.03				0.08	3.82	0.02
Desulfoluna	0.01	0.02	2.20	0.02	0.02	4.62				0.04	0.04	8.47	0.01		0.01						
SEEP-SRB1	0.13	0.07	2.36	0.05	0.01	0.84	0.04	0.01	0.78	0.09	0.08	1.67	0.06		0.06		0.22	1.74	0.34	2.73	0.13
SEEP-SRB2	0.02	0.02	0.33	0.04	0.03	0.85	0.01	0.01	0.19	0.07	0.05	1.35	0.05	0.01	0.05		0.15	2.78	0.06	1.09	0.05
SEEP-SRB4	0.01	0.02	1.34	0.01	0.01	1.30		0.01	0.39			0.12	0.01		0.01				0.03	3.28	0.01
Hyd24-12	0.04	0.03	3.44	0.01	0.02	1.15			0.03	0.02	0.03	1.73	0.01		0.01						
Atribacteria	0.02	0.03	1.51	0.08	0.07	4.80	0.05	0.07	3.02	0.12	0.12	7.18	0.02		0.02						
Spirochaeta		0.01	0.76	0.02	0.03	4.36				0.01	0.01	1.63			0.01		0.02	1.16	0.03	1.64	0.02

Table 2: Relative sequence abundances were computed for the top 135 OTUs in the iTag dataset. These OTUs correspond to ~55% of the total sequences in the unfixed bulk sediment. Bacterial and archaeal 16S rRNA gene libraries are included for the core methane seep taxa, with the total number of clones for each library indicated above. Core methane seep taxa were based on Ruff et al. (2015) and include: Candidate Phylum *Atribacteria*, Candidate Division Hyd24-12, *Methanomicrobia*, *Caldilineales*, *Desulfobacterales*, and *Spirochaetales*. While we did recover other *Chloroflexi*, no *Caldilineales* were recovered in iTag or gene library sequencing so they are not included in Table 2. Fixed bulk sediment was chosen for baseline comparison (rather than unfixed) since it includes the potential loss of cells due to fixation and wash steps, thereby processed more similarly to the Magneto-FISH samples. An average and standard deviation for relative sequence abundance among replicates was calculated for each sample set. A ratio of the average relative sequence abundance of Magneto-FISH enrichments compared to the fixed bulk sediment value is reported (Rel. Fixed). Ratios over 1.5 are underlined. 16S rRNA gene bacteria and archaea clone libraries for two Magneto-FISH enrichments and fixed bulk sediment are also included for comparison to recovered iTag diversity.

Table 3: Relative sequence abundances were computed for the top 135 OTUs in the iTag dataset that were not included in the core methane seep microbiome. An average and standard deviation for relative sequence abundance among replicates was calculated for each sample set. A ratio of the average relative sequence abundance of Magneto-FISH enrichments compared to the fixed bulk sediment value is reported (Rel. Fixed). Ratios over 1.5 are underlined. 16S rRNA gene bacteria and archaea clone libraries for two Magneto-FISH enrichments and fixed bulk sediment are also included for comparison to iTag enrichment.

	Seep1a_1441			DSS_658			Delta_495a			MSMX-Eel_932			Fixed Bulk	
Taxon	Avg.	Stdv.	Rel. Fixed	Avg.	Stdv.	Rel. Fixed	Avg.	Stdv.	Rel. Fixed	Avg.	Stdv.	Rel. Fixed	Avg.	Stdv.
Desulfarculaceae-uncl	0.02	0.03	2.53	0.02	0.03	2.39	0.01	0.01	1.01	0.05	0.05	7.18	0.01	0.01
Spirochaetae-uncl			0.21				0.04	0.02	3.70			0.06	0.01	0.01
Desulfuromusa	0.05	0.05	4.17			0.06					0.01	0.39	0.01	
Pelobacter	0.01	0.01	2.48	0.01	0.01	1.95			0.10		0.01	0.81	0.01	
Actinobacteria-OM1	0.01	0.01	0.88	0.03	0.01	2.64	0.03	0.04	2.60	0.01	0.01	0.97	0.01	
Alpha-Ancalomicrobium	0.01	0.01	2.29				0.01	0.01	2.50					
Bacteroidetes-Actibacter	0.01	0.02	1.38		0.01	0.45	0.01	0.01	0.69			0.06	0.01	
Bacteroidetes-BD-2	0.03	0.01	1.49	0.01	0.02	0.58	0.02	0.01	0.94	0.03	0.02	1.29	0.02	
Bacteroidetes-Lutibacter													0.02	
Bacteroidetes-Marinilabiaceae			3.05		0.01	3.11								
Bacteroidetes-SB-1													0.01	
Bacteroidetes-SB-5	0.01	0.01	0.89	0.01	0.01	0.96						0.70	0.01	
Bacteroidetes-VC2.1_Bac22	0.01	0.01	0.22	0.02	0.01	0.64	0.01	0.02	0.37			0.04	0.03	
Bacteroidetes-WCHB1-69			0.29		0.01	0.30						0.11	0.01	0.01
Chlorobi-PHOS-HE36							0.03	0.04						
Chloroflexi-Anaerolineaceae	0.02	0.02	0.73	0.01	0.01	0.43	0.01	0.01	0.23	0.02	0.02	0.69	0.03	0.01
Chloroflexi-Bellilinea	0.02	0.03	4.18							0.01	0.01	2.43		
Deferribacteres-Caldithrix	0.01	0.01	0.31	0.01	0.01	0.19				0.01	0.01	0.46	0.03	
Deferribacteres-SAR406	0.01	0.01	3.13			0.06	0.03	0.04	8.82			0.18		
Fibrobacteres-uncl		0.01	1.50				0.01		4.82		0.01	1.16		
Firmicutes-Fusibacter													0.01	
Firmicutes-Negativicoccus														
Firmicutes-other		0.01	0.59	0.01	0.01	1.15							0.01	
Gam-endosymbionts										0.01	0.01	3.28		
Gamma-other						0.40						0.34		

KB1														
MBGB			0.13	0.01	0.01	1.11	0.01	0.01	1.28			0.66	0.01	
MBGD				0.01	0.01	4.48	0.01	0.01	4.89					
Milano-WF1B-44							0.01	0.02	1.87			0.02	0.01	
OD1	0.02		0.88	0.03	0.03	1.20	0.01	0.01	0.43	0.03	0.01	1.16	0.02	
Plactomycetes-OM190														
Planctomycetes-Phycisphaerae	0.01	0.01	0.64							0.02	0.02	2.24	0.01	
Planctomycetes-Pla4														
Planctomycetes-SHA-43											0.01	1.39		
Sulfurimonas			0.87							0.01	0.01	1.61		
Sulfurovum	0.17	0.16	1.59	0.26	0.11	2.43	0.27	0.18	2.49	0.06	0.03	0.52	0.11	0.01
TA06		0.01	1.12											
Thaumarc-uncl						0.12							0.01	
Thiohalobacter				0.01	0.01									
Thiotrichaceae-uncl														
WS3	0.01	0.01	0.47	0.04	0.04	2.21	0.02	0.03	1.20	0.03	0.01	1.74	0.02	

than any of the *Deltaproteobacteria* probes compared with the MSMX-Eel_932 probe (Table 4). The most distinct communities were MSMX-Eel_932 enrichment and Delta_495a enrichment, with the highest proportion of unshared branch length (0.97; p-value <0.001). MSMX-Eel_932 enrichment and DSS_658 enrichment had less unshared branch length at 0.88 (<0.001), suggesting MSMX-Eel_932 and DSS_658 probes enrich for a more similar community than MSMX-Eel_932 and Delta_495a probes. Comparison of the MSMX-Eel_932 enrichment and SEEP-1a_1441 enrichment communities was not significant at the <0.001 cutoff. Within the *Deltaproteobacteria* probes, SEEP-1a_1441 enrichment and DSS_658 enrichment had the lowest proportion of unshared community (0.77, <0.001); the most similar community structures were recovered with these two probes. The next lowest proportion of unshared community is between DSS_658 enrichment and Delta_495a enrichment (0.81). SEEP-1a_1441 enrichment and Delta_495a enrichment are least similar, at 0.85. All of these values are highly significant (<0.001). This is consistent with the expectation that the overlap between the target microbial population of the SEEP-1a_1441 probe would be most similar to the target microbial population of the DSS_658 probe, while the Delta_495a enrichment would recover more total *Deltaproteobacteria* diversity.

Assessing community structure with co-occurrence network analysis

After determination of statistically significant differences between iTag Magneto-FISH and bulk sediment samples, we computed co-occurrence networks to observe which of the 135 most abundant OTUs were correlated in the methane seep microbial community. By combining the results from 100 separate microbial association calculations, we were able to assign confidence to each microbial association and determine the most robust associations. Significant associations are reported in Sup Table 5 and depicted as a network in Figure 1.

Focusing first on the common ANME syntrophic *Deltaproteobacteria* partner, SEEP-SRB1, this taxon had the most associations in the network including nine positive associations and one negative association (Figure 1). There are two separate sets of SEEP-SRB1 & *Planctomycetes* (AKAU3564 sediment group) positive associations that are both well supported. SEEP-SRB1 is also associated with three other heterotrophic taxa (Candidate Phylum Atribacteria, *Spirochaeta*, and *Bacteroidetes* (VC2.1_Bac22)) and one sulfur-oxidizing taxa (*Sulfurovum*). SEEP-SRB1 was

also associated with Candidate Division Hyd24-12, which has a currently unknown ecophysiology, but could be a heterotroph if the topology of heterotrophic taxa being in the center of the network holds true. Hyd24-12 and *Atribacteria* are also both associated with the second most associated taxa, Candidate Division OD1, but there was no direct association between SEEP-SRB1 and OD1.

SEEP-SRB2 has two of the same associations as SEEP-SRB1 (VC2.1_Bac22 and *Atribacteria*), but is the only *Deltaproteobacteria* associated with MBG-B, *Anaerolineaceae*, and *Desulfoluna* (another *Deltaproteobacteria*). SEEP-SRB4 is associated with *Desulfobulbus*, and the only *Deltaproteobacteria* associated with and ANME (2a/b), WS3, and *Actibacter*. WS3 had high relative sequence abundance in both DSS_658 and MSMX-Eel_932 enrichments (Table 3). *Desulfobulbus* is associated with *Desulfococcus*, the only *Deltaproteobacteria* associated with BD2-2, and SAR406. SAR406 had high relative sequence abundance in Seep1a_1441 and Delta_495a enrichments (Table 3). The heterotroph *Spirochaeta* is also included in the core methane seep microbiome and was associated with *Clostridia* and WS3, in addition to Hyd24-12 and SEEP-SRB1.

In examination of additional OTUs associated with sulfur metabolisms, we found *Sulfurovum* and *Sulfurimonas* (*Epsilonproteobacteria*) were not associated with each other, but are both associated with *Deltaproteobacteria*. *Sulfurimonas* is associated with *Desulfocapsa* and *Sulfurovum* is associated with SEEP-SRB1 and *Desulfobulbus*. *Sulfurovum* had high relative sequence abundance in MSMX-Eel_932 enrichments and *Sulfurimonas* had high relative sequence abundance in Seep-1a_1441, DSS_658, and Delta_495a enrichments (Table 3). The *Gammaproteobacteria, Thiohalobacter,* is only associated with *Anaerolineaceae* and was not elevated in any of the Magneto-FISH enrichments.

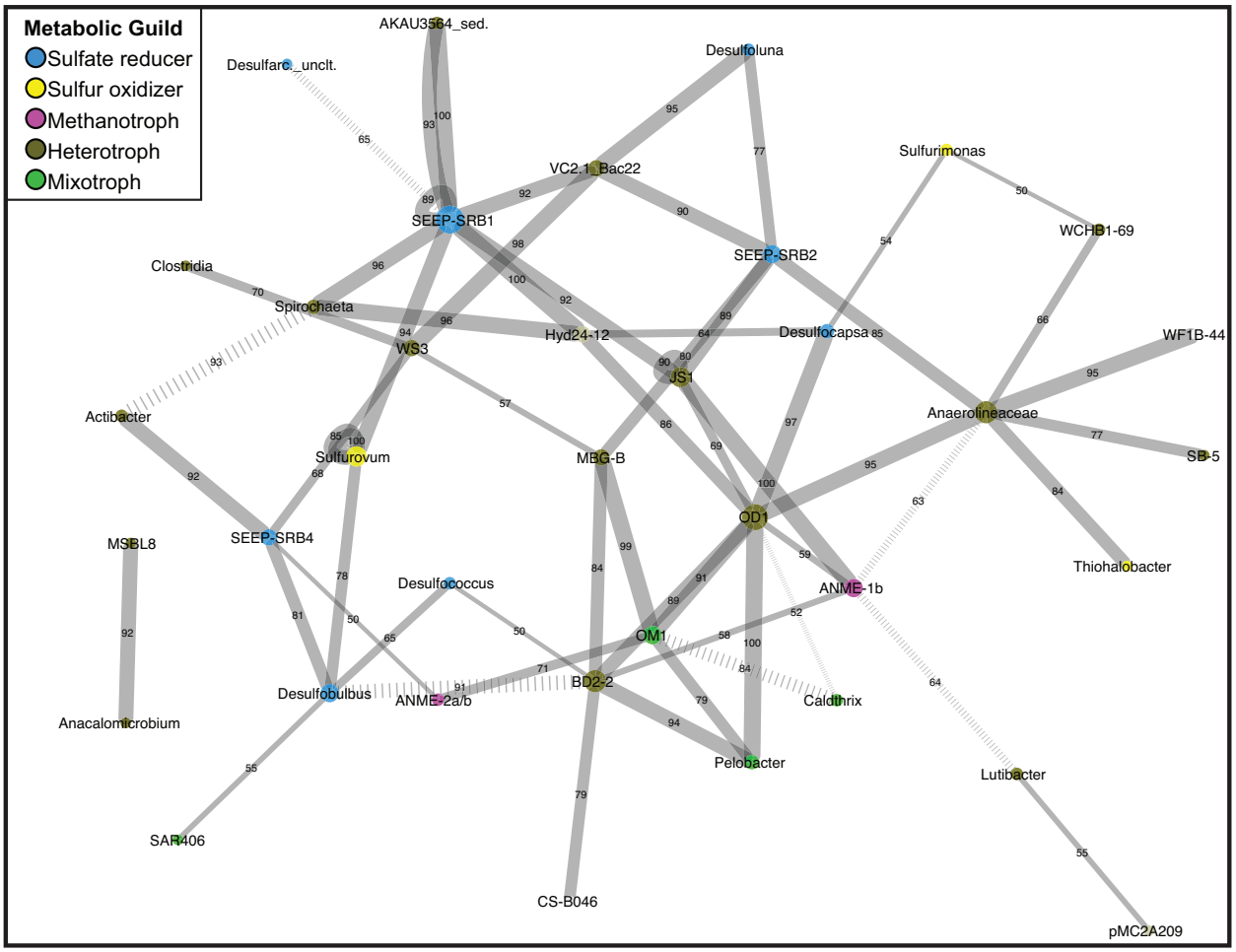
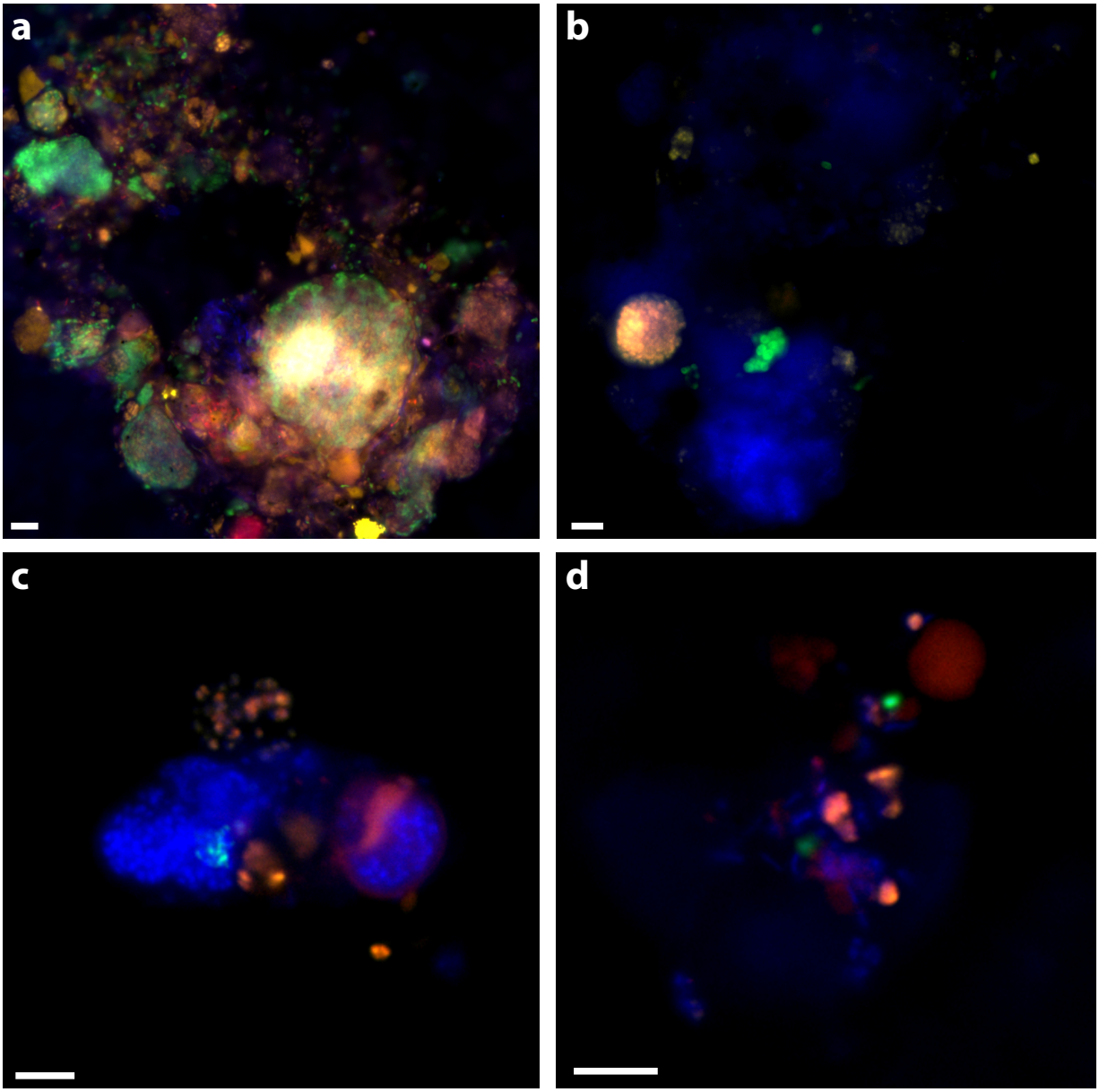


Figure 1: Co-occurrence analysis of the top 135 unique OTUs displayed in network form. Nodes represent the taxonomy of the OTUs in the network and edges are the connections between OTUs. Node size is scaled by number of connecting OTUs and colored by putative metabolic guild (blue – sulfate reducer, yellow – sulfur oxidizer, pink – archaeal methanotroph, brown – heterotroph, green – mixotroph). Edge thickness is scaled by number of occurrences of this association (from 50 to 100 times) and number of occurrences also included along the edge. Negative associations are denoted by hashed lines. The combined network is displayed using Cytoscape, with the average correlation coefficient across all runs determining the distance between nodes and the number of occurrences in 100 network iterations determining edge width. Note JS1 now C.D. *Atribacteria*.

Figure 2: Triple CARD-FISH hybridization using bacterial and archaeal probes targeting DSS_658 (A), Gam42a (B), CF319A/B (C), and Epsi404 (D) in green FITC, with ANME1-350 in red and MSMX-Eel_932 in yellow for all. Scale bar 5 μm for all. DAPI in blue.



	ANME-1_350	Eel_932	DSS_658	Epsi_404	Gam_42a	Seep-1a_1441	CF_319A/B
Total	39	70	91	5	12	29	8
With ANME-1			36	2	6	21	0
With ANME-2			63	1	9	21	4
Percent of all	39%	70%	91%	10%	24%	58%	16%
Percent ANME-1			36%	4%	12%	42%	0%
Percent ANME-2			63%	2%	18%	42%	8%

Table 4: Aggregate counts from triple CARD-FISH hybridizations with probes targeting ANME-1 (ANME-1_350), all ANME (Eel_932), DSS-type *Deltaproteobacteria* (DSS_658), *Epsilonproteobacteria* (Epsi_404), SEEP-SRB1a (SEEP-1a_1441) and *Cytophaga*, *Bacteroidetes*, *Flavobacterium*, and *Sphingobacterium* (CF_319A/B) associations described in text.

	Seep1a1441	DSS658	Delta495a	Eel932
Seep1a1441	-	0.77*	0.85*	0.91+
DSS658	-	-	0.81*	0.88*
Delta495a	-	-	-	0.97*

Table 5: Community comparison of iTag Magneto-FISH samples using weighted UniFrac analysis. Significance of relationship between communities is reported with p-values: *=<0.001, ^=0.002, +=0.030Heterotrophs are the most dominant metabolic guild in the network, and similar to sulfate-reducers, have some of the most connected taxa. The heterotroph OD1 has seven positive correlations, in addition to *Atribacteria* and Hyd24-12 listed above: *Bacteroidetes* (BD2-2), *Actinobacteria* (OM1), *Pelobacter*, ANME-1b, *Chloroflexi* (*Anaerolineaceae)*, and *Desulfocapsa. Anaerolineaceae* and *Bacteroidetes* (BD2-2) both had seven associations, but with different connectivity. BD2-2 was interconnected with other heterotrophs, sulfate-reducers, and archaeal methanotrophs in the main portion of the network, whereas *Anaerolineaceae* was connected to three taxa that share no other connections (two heterotrophs and one *Gammaproteobacteria* sulfur oxidizer). The one other ANME taxa in the network, ANME-1b, is only positively associated with heterotrophs and no known sulfate reducing groups.

Assessing ANME-bacterial partnerships by CARD-FISH

To assess ANME and DSS relative cell abundance, 100 aggregates from the same sediment incubation (see Materials & Methods) were analyzed with CARD-FISH and the DSS_658/ANME1-350/MSMX-Eel_932 probe combination. Epsi_404, Gam_42a, SEEP-1a_1441, and CF_319A/B probes were also used with the archaeal probe combination to examine non-DSS bacterial diversity recovered in the network analysis ANME associations. All probes, target populations, and references are listed in Table 1.

30% of aggregates contained an ANME-2 signal (see Materials & Methods; Table 5) and 39% of aggregates had an ANME-1 signal. ANME-1 and ANME-2 identified cells were also consistent with expected morphologies. Multiple clusters of mixed-type ANME/DSS, DSS-only, ANME-only, DSS/non-ANME, and non-DSS/non-ANME aggregates were observed with the ANME-1_350, MSMX-Eel_932, and DSS_658 probe combination (Figure 2a). There were no clear examples of aggregates with ANME/non-DSS hybridized cells, though we found many instances where both ANME and non-DSS cells were part of a larger aggregate cluster with other cell types. ANME-1 cells often occurred in the matrix surrounding tightly clustered ANME-2 aggregates. The SEEP-1a_1441 probe, targeting a subgroup of DSS, was observed to hybridize with aggregates that contained ANME-1 and ANME-2 cells, but usually with SEEP-SRB1/ANME-2 in tight association and ANME-1 in more peripheral association. Five of the SEEP-SRB1/ANME-2 aggregates did not have ANME-1 cells (10%) and three of the SEEP-SRB1/ANME-1 aggregates did not have ANME-2 cells.

Ten percent of aggregates (n=50 counted) hybridized with the Epsi_404 probe, broadly targeting members of the *Epsilonproteobacteria*. These *Epsilonproteobacteria* were mostly found in association with other bacteria and occasionally, loosely associated with some ANME. Epsi_404 hybridized cells were generally ovoid and scattered throughout an EPS matrix of cells, as depicted in Figure 2d. There was no apparent preference for *Epsilonproteobacteria* association with ANME-1 or ANME-2 aggregates (Table 4). A higher percentage of aggregates had *Gammaproteobacteria* cells (24% of 50) than *Epsilonproteobacteria* cells, and there was a slightly higher co-occurrence with ANME-2 (18%) than ANME-1 (12%) hybridized. The dominant *Gammaproteobacteria* morphology observed was a cluster or chain of large (~1 μm) ovoid cells. Gam_42a hybridizing cell clusters and chains were found both separately and associated with other bacteria, as in Figure 2b, where they are predominately an unidentified cluster stained by DAPI with a sub-aggregate of ANME-2 cells. CF319A and CF319B were used to target *Cytophaga*, *Bacteroidetes*, *Flavobacterium*, and *Sphingobacterium*. Eight percent (n=50 counted) of aggregates contained cells positively hybridizing with the CFB probe, generally observed as clustered filaments or rods (Figure 2c). Half of these aggregates also had ANME-2 hybridized cells. No CFB cells were observed to co-associate with ANME-1.

Discussion.

Evaluation of Magneto-FISH with iTag

Challenges accompanying downstream analysis of Magneto-FISH enrichments are primarily associated with low DNA yield and poor DNA quality from aldehyde fixation (for further discussion of fixation effects see Trembath-Reichert et al. 2013). Low template concentration exacerbates amplification of contaminating sequences since target and non-target templates can approach parity in a PCR reaction. Low template concentration has also been shown to create random variation in amplification products in dilution experiments (Chandler et al. 1997), which could explain the high variation seen in Magneto-FISH enrichment relative sequence abundances compared to bulk sediment samples. Despite these challenges, the DNA recovered from Magneto-FISH enrichments has been shown to increase the sequence abundance of target organisms relative to the bulk sediment by 16S rRNA gene sequencing and metagenomics on various Next Generation sequencing platforms (Pernthaler et al. 2008; Trembath-Reichert et al. 2013). In this study, conventional cloning and sequencing of full-length bacterial and archaeal 16S rRNA genes had fewer contamination issues as compared to iTag sequencing with universal primers. Our Magneto-FISH experiments were designed to mitigate as many sampling and iTag sequencing biases between samples as possible, by concurrently extracting, amplifying, and sequencing all Magneto-FISH samples in parallel, including biological and technical replicates. The relative ratio of contaminant reads to environmental OTU's were higher in Magneto-FISH enrichments than in bulk sediment samples, but bulk sediment could be used to separate indigenous community members from putative contaminants in the Magneto-FISH samples (see Materials & Methods). This provided a conservative Magneto-FISH dataset for statistical analyses and demonstrated the importance of parallel processing sequencing of bulk and separated samples.

In addition to issues with contaminating sequences, we also observed bias against some core methane seep microbiome taxa, where these taxa were consistently underrepresented by iTag when compared to gene libraries and CARD-FISH. ANME-2 was the most underrepresented taxon in iTag sequencing of the bulk sediment and mock communities, with much greater relative sequence and relative cell abundance in gene library sequencing and CARD-FISH analysis, respectively. It is most likely that iTag sequencing bias with the EMP primer set is the reason ANME-2c was not enriched in the Magneto-FISH samples and absent from microbial community network analysis. Members of the ANME-2a/b were also, to a lesser extent, underrepresented with iTag. In addition

to our gene libraries and CARD-FISH analysis, independent assays using FISH with mono labeled oligonucleotide probes from this sediment incubation further confirmed the abundance of ANME-2 aggregates; 25% of aggregates were ANME-2c and 17% of aggregates were ANME-2b, with about half of ANME-2 aggregates associating with a bacterial partner other than SEEP-SRB1 (Supplement McGlynn et al. 2015). We conclude that while expected ANME-2 associations were not recovered, they can be explained by EMP iTag bias and therefore do not reduce the validity of other non-ANME-2 associations recovered in the co-occurrence analysis (see Sup Table 2 and 3 captions for further discussion of ANME-2c bias). Although ANME-1a was not underrepresented in the iTag data, it still does not appear in the co-occurrence network. In other co-occurrence network studies dominant OTUs were not associated with the majority of the microbial community, which was thought to be due to a high degree of functional redundancy (Mu & Moreau 2015). Possible functional redundancy with other archaeal groups, or simply non-specific, loose spatial association with many taxa, as suggested by CARD-FISH analysis, could explain why ANME-1a was not recovered in our network analysis.

Despite this unanticipated methodological bias, iTag sequencing is a valid and valuable tool when combined with Magneto-FISH enrichment techniques for microbial association hypothesis development and testing. For example, we saw more bacterial OTUs, especially among *Deltaproteobacteria*, in the iTag samples compared with conventional gene libraries and the core methane seep taxon Hyd24-12 was not even observed among gene library sequences.

Magneto-FISH enrichment

This study provides a novel combination of nested Magneto-FISH enrichments and microbial community network analysis methods to develop hypotheses regarding specific lineage associations and, by inference, discusses the potential for additional metabolic interactions relating to sulfur cycling in methane seep sediments. Notwithstanding the low recovery of ANME-2 OTUs, there was statistical support for Magneto-FISH enrichments increasing the relative iTag sequence abundance of target organisms. Statistical analyses demonstrated SEEP-SRB1 and *Desulfobulbus* OTUs were significantly different in Magneto-FISH samples (t-tests), and these OTUs were significantly more enriched in Magneto-FISH samples using linear discriminant analysis (LDA) effect size (LEfSe). Additionally, weighted UniFrac analysis showed the highest

percentage of shared phylogeny was between the clade-specific SEEP-1a_1441 probe and the family-specific *Desulfobacteraceae* DSS_658 probe enrichments. Therefore these Magneto-FISH samples contain microbial community overlap consistent with probe target specificity, even when some dominant community members are not represented at expected relative sequence abundance in the iTag analysis (ANME-2).

Magneto-FISH enrichment relative sequence abundance followed expected trends for *Deltaproteobacteria* (Table 2). SEEP-SRB1 had the highest relative sequence abundance in Seep-1a_1441 and MSMX-Eel_932 enrichments, which should target this group. *Desulfobulbus* had the highest relative sequence abundance in the Delta_495a enrichment, which was the only Magneto-FISH probe that should hybridize to this group (though *Desulfobulbus* could also be retrieved via association with other target organisms). OTUs affiliated with *Desulfoluna* (within the *Desulfobacteraceae*) had the highest relative sequence abundance of all *Deltaproteobacteria* in the DSS_658 enrichment and are also targeted by the DSS_658 probe. *Desulfoluna* were not specifically targeted by MSMX-Eel_932 or Seep-1a_1441 probes, but had high relative sequence abundane in these samples and may have a potential association with ANME/DSS consortia. Also, *Atribacteria* (JS1) was recovered in all iTag sequencing of Magneto-FISH enrichments, suggesting they may associate with either DSS/ANME or DSB/ANME consortia. Members of the Hyd24-12 were only recovered in Seep1a_1441 and MSMX-Eel_932 enrichments and may preferentially associate with SEEP-SRB1a/ANME consortia.

Evaluating our iTag relative sequence abundance data with co-occurrence analysis, we developed hypotheses that were not subject to the variation between Magneto-FISH enrichment replicates; associated taxa should always co-vary, even when they are less abundant than expected. Within the core methane seep taxa, high relative sequence abundances of *Atribacteria* and Hyd24-12 with SEEP-SRB1 targeting Magneto-FISH enrichments were upheld by the network. Hyd24-12 is highly associated with SEEP-SRB1, whereas *Atribacteria* is highly associated with both SEEP-SRB1 (DSS) and SEEP-SRB2 (DSB). While *Atribacteria* have not been cultured, metagenomic sequencing suggests they are likely heterotrophic anaerobes involved in fermentation (Nobu et al. 2015). Hyd24-12 was first cloned from Hydrate Ridge (Knittel et al. 2003) and has been cited as a core methane seep microbial taxon (Ruff et al. 2015), but nothing is known about its physiology. The Hyd24-12/SEEP-SRB1 association was also one of the four unique associations that were

recovered in all the network computations (n=100). These results may aid in determining a role for these enigmatic candidate phyla of the methane seep microbiome.

Methanomicrobia and *Deltaproteobacteria* only had one co-occurrence in our network. The one statistically supported network ANME/SRB association was between ANME-2a/b and SEEP-SRB4. SEEP-SRB4, belonging to the *Desulfobulbaceae* (Knittel et al. 2003), and ANME-2a/b both had high relative sequence abundance in the ANME-targeting MSMX-Eel_932 enrichment bacterial 16S rRNA gene library. There have been FISH-confirmed physical associations between ANME-2/ANME-3 and *Desulfobulbaceae* (Green-Saxena et al. 2014; Löesekann et al. 2007; Pernthaler et al. 2008) in AOM systems. SEEP-SRB4 was also strongly associated with the candidate phyla WS3 in the network, and WS3 was enriched in both DSS_658 and MSMX-Eel_932 enrichments. Both SEEP-SRB4 associations with ANME-2a/b and WS3 warrant future study.

While expected ANME-2/*Deltaproteobacteria* associations were not recovered (see *Evaluation of Magneto-FISH with iTag*), network analysis did recover many *Deltaprotobacteria* co-occurring with bacterial groups. Almost half of all positive associations contained a *Deltaproteobacteria* OTU (30/61), suggesting a dominant role for the sulfur cycle metabolisms. Of those, 21 associations were with a non-*Proteobacteria* OTU including a number of candidate organisms as described above. The association between SEEP-SRB1 and 'AKAU3564,' a *Planctomycetes*-affiliated heterotrophic sediment group, was observed twice with two separate OTU associations in this clade that were both strongly supported (occurring 100/100 and 93/100 times, respectively, that the network analysis was run, Sup Table 5). This Planctomycete group was first described in methane hydrate bearing deep marine sediments of the Peru Margin (Inagaki et al. 2006). *Planctomycetes*-associated sequences were previously recovered in association with ANME-2c Magneto-FISH samples from the Eel River Basin, where the preferred partner was observed to be the SEEP-SRB1 group (Pernthaler et al. 2008). It follows that SEEP-SRB1 may also co-occur with *Planctomycetes*, if these organisms are affiliated (either directly or indirectly) with ANME-2 consortia. By similar logic, although it did not have high relative sequence abundance in the Seep1a_1441 enrichment, this could explain the high relative sequence abundance of this group in the MSMX-Eel_932 enrichment (Table 3). *Planctomycetes* targeted CARD-FISH hybridization using the general *Planctomycetes* probe Pla_886 was attempted; however, many cells with a

morphology similar to ANME-1 were hybridized and the results were deemed inconclusive. This ambiguity could be due to the probe's single base pair mismatch to 97% of ANME-1a, 94% of ANME-1b, and 25% of ANME-2b, even if this mismatch was centrally located (SILVA TestProbe online tool, Greuter et al. 2015). *Spirochaeta* was also associated with SEEP-SRB1, in addition to Hyd24-12 and WS3, and had high relative sequence abundance in both the DSS_658 and MSMX-Eel_932 enrichments (Table 2). In addition to being core methane seep microbial taxa, some members of the *Spirochaetes* have sulfide-oxidizing capabilities in mats with sulfidogenic bacteria (Dubinina et al. 2004) and it is possible that these organisms may be utilizing sulfide produced in seep systems as well.

Epsilonproteobacteria and *Deltaproteobacteria* were the most common intra-*Proteobacteria* association in the network and have been shown to co-occur in many sulfidic habitats (Campbell et al. 2006; Omoregie et al. 2008), where *Epsilonproteobacteria* oxidize sulfur and *Deltaproteobacteria* disproportionate or reduce sulfur species (Pjevac et al. 2014). In the network, *Sulfurovum* was associated with both SEEP-SRB1 and *Desulfobulbus*, and this was also seen in the relative sequence abundance data where *Sulfurovum* had high relative sequence abundance in all of the *Deltaproteobacteria* Magneto-FISH enrichments. *Epsilonproteobacteria* have been shown to oxidize sulfide to S^{o} or HS^{-} to sulfate in microbial mats (Pjevac et al. 2014), allowing some sulfur substrate differentiation between these *Epsilonproteobacteria* groups in this system. *Sulfurimonas* was not strongly associated with any *Deltaproteobacteria* in the network analysis and only had high relative sequence abundance in the MSMX-Eel_932 enrichment (16S rRNA gene iTag, 16S rRNA gene bacterial, and soxB gene libraries; see Sup Figure 2 for further discussion of metabolic genes). CARD-FISH analysis using probe Epsi_404 confirmed the presence of *Epsilonproteobacteria* cells within some ANME and other non-hybridized cell-containing loose aggregates, but did not appear to be in the tight physical association characteristic of ANME/SRB consortia. While cultured representatives of these *Epsilonproteobacteria* have optimum growth with some oxygen present (Inagaki et al. 2003; Inagaki et al. 2004), it is possible that these uncultured methane seep *Epsilonproteobacteria* may be able to use other oxidants such as nitrate or intermediate sulfur species while in anaerobic incubation conditions.

In comparison to *Delta*- and *Epsilonproteobacteria*, there was only one *Gammaproteobacteria* OTU in the network (*Thiohalobacter*, with one *Anaerolineaceae* association). Cultured

representatives of *Thiohalobacter* have diverse sulfur capabilities, including thiocyanate metabolism, but are not known to form associations with other sulfur cycling organisms (Sorokin et al. 2010). This differentiation between *Gamma*- and *Epsilon*-/*Deltaproteobacteria* has been seen in other systems such as sulfidic cave biofilms (Macalady et al. 2008) or in microbial mats on marine sediments (Pjevac et al. 2014). Gam_42a hybridizing cells (*Gammaproteobacteria*) were observed to form aggregates with non-ANME and non-*Desulfobulbaceae* (DSS) cells in our CARD-FISH analysis, but the identity of these organisms was not determined. While not recovered in the network, the majority of the *Gammaproteobacteria* OTUs observed by iTag from the both the bulk sediment and MSMX-Eel_932 Magneto-FISH 16S rRNA gene (Table 1) and aprA gene libraries (see Sup Figure 2 for further discussion of metabolic genes) were from the SILVA taxonomy endosymbiont clade. This endosymbiont clade houses organisms with a carbon-fixation/sulfur-oxidation metabolism (Duperron et al. 2012; Goffredi 2010) and is predicted to be an important member of the sulfur and carbon cycles in marine sediments outside of an endosymbiotic lifestyle (Lenk et al. 2011).

There were also three unique, positive *Deltaproteobacteria-Deltaproteobacteria* associations observed in the network: *Desulfobulbus*/*Desulfococcus*, *Desulfobulbus*/SEEP-SRB4, *Desulfoluna*/SEEP-SRB2. These multiple intra-*Deltaproteobacteria* associations suggest there may be further nuances to be explored in the *Deltaproteobacteria* community structure, perhaps akin to the nitrate based partitioning observed between DSB and DSS in seep sediments (Green-Saxena et al. 2014). *Desulfobulbus* was also associated with SAR406, and SAR406 had high relative sequence abundance in the Delta495a enrichments. SAR406 (Marine Group A) fosmids contained polysulfide reductase genes that may be used for dissimilatory polysulfide reduction (Wright et al. 2014). *Desulfobulbus* can also use polysulfide, in addition to a range of other sulfur sources (Fuseler & Cypionka 1995), potentially linking these two taxa.

Conclusions.

Our findings support the utilization of paired Magneto-FISH and iTag sequencing in developing and testing hypotheses to interrogate complex interactions in microbial communities. Contaminants and amplification bias can be identified and mitigated with diversity assessment by multiple means (i.e. multiple iTag primer sets, FISH surveys, or non-16S rRNA gene surveys) and

parallel processing of control samples (bulk sediment and no-template) along with Magneto-FISH enrichments. Since it may not always be known *a priori* which taxa are in an environmental sample, sequencing of a defined mock community may not be an option for assessing bias. However, in our case, prior knowledge of major seep taxa enabled assessment of amplification bias by iTag. It should also be noted that the degree of bias was more pronounced in the environmental samples than our mock samples; therefore mock community samples may not fully capture the degree of bias, but can be useful in identifying which taxa may be the most biased. We found the bulk sediment 16S rRNA gene libraries to be the most useful for determining which of the most abundant taxa were affected by amplification bias. Future studies may benefit more from bulk sediment analysis by a range of iTag primer sets or gene libraries to assess potential sequencing biases in a new microbial community.

Multiple statistical methods supported differences between Magneto-FISH enrichments and the bulk sediment. We also found variation between SparCC network computations. Therefore, we added confidence to network associations by reporting the number of times an association was recovered out of 100 co-occurrence iterations along with correlation and p-value.

Our resultant microbial community network had many statistically significant methane seep taxa correlations beyond the common ANME/SRB association. The downplay of anaerobic methanotrophs in our iTag sequencing may have had the beneficial effect of bringing fermenters to the forefront, highlighting their complex role in methane seep microbial communities. Within the core methane seep microbiome taxa, there were strong associations between *Atribacteria* and Hyd24-12 and *Deltaproteobacteria*, but no direct association between *Atribacteria* and Hyd24-12. This may indicate a different niche for these two currently uncultured groups in methane seep systems. *Sulfurovum* and *Sulfurimonas* were differentiated as either *Deltaproteobacteria*-associated or archaea-associated, respectively. There were statistically significant associations between *Deltaproteobacteria* and non-*Proteobacteria*, such as the *Planctomycetes* sediment group 'AKAU3564,' and groups that contained neither SRB nor ANME but had high statistical significance, such as MBG-B and OM1. Future development and application of more specific FISH probes will assist in further hypotheses development and testing of these associations in Hydrate Ridge methane seeps.

Some groups, such as *Gammaproteobacteria*, appeared to have associations with other microbes based on broad FISH surveys and Magneto-FISH relative sequence abundance data, but were not recovered in the network analysis. Determination of the specific *Gammaproteobacteria* involved in associations via FISH probe development or other means (Hatzenpichler et al., in review) will also aid in refining why associations might be missed in the microbial network analysis based on DNA taxa co-occurrence. In summary, a continual feedback loop between microbial identification and isolation techniques and gene based statistical analyses is required to tease apart interactions within complex microbial systems. The combination of Magneto-FISH and high throughput, parallel iTag sequencing provides an effective bridge between these two modes.

Acknowledgements.

We thank the crew of the R/V *Atlantis* and DSV *JASON II*, as well as Abigail Green-Saxena and Joshua Steele for assistance with optimization of the Magneto-FISH protocol and Stephanie Connon for assistance with sequencing. We also are grateful to Katherine Dawson, Emil Ruff, and two anonymous reviewers for providing comments on this manuscript.

References:

Akerman NH, Butterfield DA, and Huber JA. 2013. Phylogenetic diversity and functional gene patterns of sulfur-oxidizing subseafloor Epsilonproteobacteria in diffuse hydrothermal vent fluids. *Frontiers in Microbiology* 4:185. 10.3389/fmicb.2013.00185

Amann RI, Binder BJ, Olson RJ, Chisholm SW, Devereux R, and Stahl DA. 1990. Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Applied and Environmental Microbiology* 56:1919-1925.

Ashkin A. 1997. Optical trapping and manipulation of neutral particles using lasers. *Proceedings of the National Academy of Sciences* 94:4853-4860.

Barberán A, Bates ST, Casamayor EO, and Fierer N. 2012. Using network analysis to explore co-occurrence patterns in soil microbial communities. *The ISME Journal* 6:343-351. 10.1038/ismej.2011.119

Berry D, Mader E, Lee TK, Woebken D, Wang Y, Zhu D, Palatinszky M, Schintlmeister A, Schmid MC, Hanson BT, Shterzer N, Mizrahi I, Rauch I, Decker T, Bocklitz T, Popp Jr, Gibson CM, Fowler PW, Huang WE, and Wagner M. 2015. Tracking heavy water (D2O) incorporation for identifying and sorting active microbial cells. *Proceedings of the National Academy of Sciences* 112:E194-E203. 10.1073/pnas.1420406112

Blazejak A, Kuever J, Erséus C, Amann R, and Dubilier N. 2006. Phylogeny of 16S rRNA, Ribulose 1,5-Bisphosphate Carboxylase/Oxygenase, and Adenosine 5'-Phosphosulfate Reductase Genes from Gamma- and Alphaproteobacterial Symbionts in Gutless Marine Worms (Oligochaeta) from Bermuda and the Bahamas. *Applied and Environmental Microbiology* 72:5527-5536. 10.1128/aem.02441-05

Boetius A, Ravenschlag K, Schubert CJ, Rickert D, Widdel F, Gieseke A, Amann R, Jorgensen BB, Witte U, and Pfannkuche O. 2000. A marine microbial consortium apparently mediating anaerobic oxidation of methane. *Nature* 407:623-626.

Campbell BJ, Engel AS, Porter ML, and Takai K. 2006. The versatile [epsi]-proteobacteria: key players in sulphidic habitats. *Nat Rev Micro* 4:458-468.

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM, Betley J, Fraser L, Bauer M, Gormley N, Gilbert JA, Smith G, and Knight R. 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* 6:1621-1624.

Chaffron S, Rehrauer H, Pernthaler J, and von Mering C. 2010. A global network of coexisting microbes from environmental and whole-genome sequence data. *Genome Research* 20:947-959. 10.1101/gr.104521.109

Chandler DP, Fredrickson JK, and Brockman FJ. 1997. Effect of PCR template concentration on the composition and distribution of total community 16S rDNA clone libraries. *Molecular Ecology* 6:475-482.

DeLong EF. 1992. Archaea in coastal marine environments. *Proc Natl Acad Sci U S A* 89:5685-5689.

Dubinina GA, Grabovich MY, and Chernyshova YY. 2004. The role of oxygen in the regulation of the metabolism of aerotolerant spirochetes, a major component of ‚ÄúThiodendron‚Äù bacterial sulfur mats. *Microbiology* 73:621-628. 10.1007/s11021-005-0001-3

Duperron S, Rodrigues CF, Léger N, Szafranski K, Decker C, Olu K, and Gaudron SM. 2012. Diversity of symbioses between chemosynthetic bacteria and metazoans at the Guiness cold seep site (Gulf of Guinea, West Africa). *MicrobiologyOpen* 1:467-480. 10.1002/mbo3.47

Edgar RC, Haas BJ, Clemente JC, Quince C, and Knight R. 2011. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27:2194-2200. 10.1093/bioinformatics/btr381

Friedman J, and Alm EJ. 2012. Inferring correlation networks from genomic survey data.

Fuseler K, and Cypionka H. 1995. Elemental sulfur as an intermediate of sulfide oxidation with oxygen byDesulfobulbus propionicus. *Archives of Microbiology* 164:104-109. 10.1007/bf02525315

Goffredi SK. 2010. Indigenous ectosymbiotic bacteria associated with diverse hydrothermal vent invertebrates. *Environmental Microbiology Reports* 2:479-488. 10.1111/j.1758-2229.2010.00136.x

Green-Saxena A, Dekas AE, Dalleska NF, and Orphan VJ. 2014. Nitrate-based niche differentiation by distinct sulfate-reducing bacteria involved in the anaerobic oxidation of methane. *ISME J* 8:150-163. 10.1038/ismej.2013.147

Greuter D, Loy A, Horn M, and Rattei T. 2015. probeBase,Äîan online resource for rRNA-targeted oligonucleotide probes and primers: new features 2016. *Nucleic Acids Research*. 10.1093/nar/gkv1232

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, and Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 59:307-321. 10.1093/sysbio/syq010

Hatzenpichler R, and Orphan V. 2015. Detection of Protein-Synthesizing Microorganisms in the Environment via Bioorthogonal Noncanonical Amino Acid Tagging (BONCAT). Totowa, NJ: Humana Press, 1-13.

Hatzenpichler R, Scheller S, Tavormina PL, Babin BM, Tirrell DA, and Orphan VJ. 2014. In situ visualization of newly synthesized proteins in environmental microbes using amino acid tagging and click chemistry. *Environmental Microbiology* 16:2568-2590. 10.1111/1462-2920.12436

House CH, Orphan VJ, Turk KA, Thomas B, Pernthaler A, Vrentas JM, and Joye SB. 2009. Extensive carbon isotopic heterogeneity among methane seep microbiota. *Environmental Microbiology* 11:2207-2215. 10.1111/j.1462-2920.2009.01934.x

Inagaki F, Nunoura T, Nakagawa S, Teske A, Lever M, Lauer A, Suzuki M, Takai K, Delwiche M, Colwell FS, Nealson KH, Horikoshi K, D,ÄôHondt S, and J√∏rgensen BB. 2006. Biogeographical distribution and diversity of microbes in methane hydrate-bearing deep marine sediments on the Pacific Ocean Margin. *Proceedings of the National Academy of Sciences of the United States of America* 103:2815-2820. 10.1073/pnas.0511033103

Inagaki F, Takai K, Kobayashi H, Nealson KH, and Horikoshi K. 2003. Sulfurimonas autotrophica gen. nov., sp. nov., a novel sulfur-oxidizing Œμ-proteobacterium isolated from hydrothermal sediments in the Mid-Okinawa Trough. *International Journal of Systematic and Evolutionary Microbiology* 53:1801-1805. 10.1099/ijs.0.02682-0

Inagaki F, Takai K, Nealson KH, and Horikoshi K. 2004. Sulfurovum lithotrophicum gen. nov., sp. nov., a novel sulfur-oxidizing chemolithoautotroph within the Œμ-Proteobacteria isolated from Okinawa Trough hydrothermal sediments. *International Journal of Systematic and Evolutionary Microbiology* 54:1477-1482. 10.1099/ijs.0.03042-0

Kalyuzhnaya M, Lidstrom M, and Chistoserdova L. 2008. Real-time detection of actively metabolizing microbes by redox sensing as applied to methylotroph populations in Lake Washington. *The ISME Journal* 2:696-706.

Knittel K, Boetius A, Lemke A, Eilers H, Lochte K, Pfannkuche O, Linke P, and Amann R. 2003. Activity, distribution, and diversity of sulfate reducers and other bacteria in sediments above gas hydrate (Cascadia Margin, Oregon). *Geomicrobiology Journal* 20:269-294.

Kozich JJ, Westcott SL, Baxter NT, Highlander SK, and Schloss PD. 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. 10.1128/aem.01043-13

Lane DJ. 1991. 16S/23S rRNA sequencing. In: Stackebrandt EaG, M., ed. *Nucleic acid techniques in bacterial systematics*. Chichester, England: John Wiley & Sons, 115-175.

Lenk S, Arnds J, Zerjatke K, Musat N, Amann R, and Mussmann M. 2011. Novel groups of Gammaproteobacteria catalyse sulfur oxidation and carbon fixation in a coastal, intertidal sediment. *Environ Microbiol* 13:758-774. 10.1111/j.1462-2920.2010.02380.x

Li W, Cowley A, Uludag M, Gur T, McWilliam H, Squizzato S, Park YM, Buso N, and Lopez R. 2015. The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Research*. 10.1093/nar/gkv279

Löesekann T, Knittel K, Nadalig T, Fuchs B, Niemann H, Boetius A, and Amann R. 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.

Loy A, Lehner A, Lee N, Adamczyk J, Meier H, Ernst J, Schleifer K-H, and Wagner M. 2002. Oligonucleotide Microarray for 16S rRNA Gene-Based Detection of All Recognized Lineages of Sulfate-Reducing Prokaryotes in the Environment. *Applied and Environmental Microbiology* 68:5064-5081. 10.1128/aem.68.10.5064-5081.2002

Lozupone C, and Knight R. 2005. UniFrac: a New Phylogenetic Method for Comparing Microbial Communities. *Applied and Environmental Microbiology* 71:8228-8235. 10.1128/aem.71.12.8228-8235.2005

Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar, Buchner A, Lai T, Steppi S, Jobb G, Förster W, Brettske I, Gerber S, Ginhart AW, Gross O, Grumann S, Hermann S, Jost R, König A, Liss T, Lüßmann R, May M, Nonhoff B, Reichel B, Strehlow R, Stamatakis A, Stuckmann N, Vilbig A, Lenke M, Ludwig T, Bode A, and Schleifer K-H. 2004. ARB: a software environment for sequence data. *Nucleic Acids Research* 32:1363-1371. 10.1093/nar/gkh293

Macalady JL, Dattagupta S, Schaperdoth I, Jones DS, Druschel GK, and Eastman D. 2008. Niche differentiation among sulfur-oxidizing bacterial populations in cave waters. *ISME J* 2:590-601.

Macalady JL, Lyon EH, Koffman B, Albertson LK, Meyer K, Galdenzi S, and Mariani S. 2006. Dominant Microbial Populations in Limestone-Corroding Stream Biofilms, Frasassi Cave System, Italy. *Applied and Environmental Microbiology* 72:5596-5609. 10.1128/aem.00715-06

Malfatti F, and Azam F. 2010. Atomic force microscopy reveals microscale networks and possible symbioses among pelagic marine bacteria. *Aquatic Microbial Ecology* 58:1-14. doi: 10.3354/ame01355

Manz W, Amann R, Ludwig W, Vancanneyt M, and Schleifer K-H. 1996. Application of a suite of 16S rRNA-specific oligonucleotide probes designed to investigate bacteria of the phylum cytophaga-flavobacter-bacteroides in the natural environment. *Microbiology* 142:1097-1106. doi:10.1099/13500872-142-5-1097

Manz W, Amann R, Ludwig W, Wagner M, and Schleifer KH. 1992. PHYLOGENETIC OLIGODEOXYNUCLEOTIDE PROBES FOR THE MAJOR SUBCLASSES OF PROTEOBACTERIA - PROBLEMS AND SOLUTIONS. *Systematic and Applied Microbiology* 15:593-600.

Manz W, Eisenbrecher M, Neu TR, and Szewzyk U. 1998. Abundance and spatial organization of Gram-negative sulfate-reducing bacteria in activated sludge investigated by in situ probing with specific 16S rRNA targeted oligonucleotides. *Fems Microbiology Ecology* 25:43-61. 10.1111/j.1574-6941.1998.tb00459.x

McGlynn SE, Chadwick GL, Kempes CP, and Orphan VJ. 2015. Single cell activity reveals direct electron transfer in methanotrophic consortia. *Nature* 526:531-535. 10.1038/nature15512

Melin J, and Quake SR. 2007. Microfluidic large-scale integration: the evolution of design rules for biological automation. *Annu Rev Biophys Biomol Struct* 36:213-231.

Mu A, and Moreau JW. 2015. The Geomicrobiology of CO2 geosequestration: a focused review on prokaryotic community responses to field-scale CO2 injection. *Frontiers in Microbiology* 6. 10.3389/fmicb.2015.00263

Neef A, Amann R, Schlesner H, and Schleifer K-H. 1998. Monitoring a widespread bacterial group: in situ detection of planctomycetes with 16S rRNA-targeted probes. *Microbiology* 144:3257-3266. doi:10.1099/00221287-144-12-3257

Niemann H, Loesekann T, de Beer D, Elvert M, Nadalig T, Knittel K, Amann R, Sauter EJ, Schluter M, and Klages M. 2006. Novel microbial communities of the Haakon Mosby mud volcano and their role as a methane sink. *Nature* 443:854-858.

Nobu MK, Dodsworth JA, Murugapiran SK, Rinke C, Gies EA, Webster G, Schwientek P, Kille P, Parkes RJ, Sass H, Jorgensen BB, Weightman AJ, Liu W-T, Hallam SJ, Tsiamis G, Woyke T, and Hedlund BP. 2015. Phylogeny and physiology of candidate phylum /`Atribacteria/' (OP9/JS1) inferred from cultivation-independent genomics. *ISME J*. 10.1038/ismej.2015.97

Omoregie EO, Mastalerz V, de Lange G, Straub KL, Kappler A, R√∏y H, Stadnitskaia A, Foucher J-P, and Boetius A. 2008. Biogeochemistry and Community Composition of Iron- and Sulfur-Precipitating Microbial Mats at the Chefren Mud Volcano (Nile Deep Sea Fan, Eastern Mediterranean). *Applied and Environmental Microbiology* 74:3198-3215. 10.1128/aem.01751-07

Orphan VJ. 2009. Methods for unveiling cryptic microbial partnerships in nature. *Current Opinion in Microbiology* 12:231-237.

Orphan VJ, Hinrichs KU, Ussler W, Paull CK, Taylor LT, Sylva SP, Hayes JM, and Delong EF. 2001a. Comparative analysis of methane-oxidizing archaea and sulfate-reducing bacteria in anoxic marine sediments. *Applied and Environmental Microbiology* 67:1922-1934.

Orphan VJ, House CH, Hinrichs K-U, McKeegan KD, and DeLong EF. 2001b. Methane-consuming archaea revealed by directly coupled isotopic and phylogenetic analysis. *Science* 293:484-487.

Orphan VJ, House CH, Hinrichs KU, McKeegan KD, and DeLong EF. 2002. Direct phylogenetic and isotopic evidence for multiple groups of archaea involved in the anaerobic oxidation of methane. *Geochimica Et Cosmochimica Acta* 66:A571-A571.

Parada A, Needham DM, and Fuhrman JA. 2015. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time-series and global field samples. *Environmental Microbiology*:n/a-n/a. 10.1111/1462-2920.13023

Pernthaler A, Dekas AE, Brown CT, Goffredi SK, Embaye T, and Orphan VJ. 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. 10.1073/pnas.0711303105

Pjevac P, Kamyshny A, Dyksma S, and Mußmann M. 2014. Microbial consumption of zero-valence sulfur in marine benthic habitats. *Environmental Microbiology* 16:3416-3430. 10.1111/1462-2920.12410

Pruesse E, Peplies J, and Glöckner FO. 2012. SINA: Accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* 28:1823-1829. 10.1093/bioinformatics/bts252

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

R Core Team. 2015. R: A Language and Environment for Statistical Computing. *Available at* http://www.R-project.org/.

Ravenschlag K, Sahm K, Pernthaler J, and Amann R. 1999. High bacterial diversity in permanently cold marine sediments. *Applied and Environmental Microbiology* 65:3982-3989.

Ruff SE, Biddle JF, Teske AP, Knittel K, Boetius A, and Ramette A. 2015. Global dispersion and local diversification of the methane seep microbiome. *Proceedings of the National Academy of Sciences* 112:4015-4020.

Šafařík I, and Šafaříková M. 1999. Use of magnetic techniques for the isolation of cells. *Journal of Chromatography B* 722:33-53.

Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, Turner P, Parkhill J, Loman NJ, and Walker AW. 2014. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol* 12:87. 10.1186/s12915-014-0087-z

Schattenhofer M, and Wendeberg A. 2011. Capturing Microbial Populations for Environmental Genomics. *Handbook of Molecular Microbial Ecology I*: John Wiley & Sons, Inc., 735-740.

Schreiber L, Holler T, Knittel K, Meyerdierks A, and Amann R. 2010. Identification of the dominant sulfate-reducing bacterial partner of anaerobic methanotrophs of the ANME-2 clade. *Environmental Microbiology* 12:2327-2340. 10.1111/j.1462-2920.2010.02275.x

Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, and Huttenhower C. 2011. Metagenomic biomarker discovery and explanation. *Genome Biol* 12:R60. 10.1186/gb-2011-12-6-r60

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, and Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13:2498-2504. 10.1101/gr.1239303

Sorokin DY, Kovaleva OL, Tourova TP, and Muyzer G. 2010. Thiohalobacter thiocyanaticus gen. nov., sp. nov., a moderately halophilic, sulfur-oxidizing gammaproteobacterium from hypersaline lakes, that utilizes thiocyanate. *Int J Syst Evol Microbiol* 60:444-450. 10.1099/ijs.0.012880-0

Steele JA, Countway PD, Xia L, Vigil PD, Beman JM, Kim DY, Chow C-ET, Sachdeva R, Jones AC, Schwalbach MS, Rose JM, Hewson I, Patel A, Sun F, Caron DA, and Fuhrman JA. 2011. Marine bacterial, archaeal and protistan association networks reveal ecological linkages. *ISME J* 5:1414-1425.

Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, Djahanschiri B, Zeller G, Mende DR, Alberti A, Cornejo-Castillo FM, Costea PI, Cruaud C, d'Ovidio F, Engelen S, Ferrera I, Gasol JM, Guidi L, Hildebrand F, Kokoszka F, Lepoivre C, Lima-Mendez G, Poulain J, Poulos BT, Royo-Llonch M, Sarmento H, Vieira-Silva S, Dimier C, Picheral M, Searson S, Kandels-Lewis S, coordinators TO, Bowler C, de Vargas C, Gorsky G, Grimsley N, Hingamp P, Iudicone D, Jaillon O, Not F, Ogata H, Pesant S, Speich S, Stemmann L, Sullivan MB, Weissenbach J, Wincker P, Karsenti E, Raes J, Acinas SG, and Bork P. 2015. Structure and function of the global ocean microbiome. *Science* 348. 10.1126/science.1261359

Takano Y, Kaneko M, Kahnt Jr, Imachi H, Shima S, and Ohkouchi N. 2013. Detection of coenzyme F430 in deep sea sediments: A key molecule for biological methanogenesis. *Organic Geochemistry* 58:137-140. http://dx.doi.org/10.1016/j.orggeochem.2013.01.012

Trembath-Reichert E, Green-Saxena A, and Orphan VJ. 2013. Chapter Two - Whole Cell Immunomagnetic Enrichment of Environmental Microbial Consortia Using rRNA-Targeted Magneto-FISH. In: Edward FD, ed. *Microbial Metagenomics, Metatranscriptomics, and Metaproteomics*. San Diego, CA: Academic Press, 21-44.

Treude T, Orphan V, Knittel K, Gieseke A, House CH, and Boetius A. 2007. Consumption of Methane and CO2 by Methanotrophic Microbial Mats from Gas Seeps of the Anoxic Black Sea. *Applied and Environmental Microbiology* 73:2271-2283. 10.1128/aem.02685-06

Vigneron A, L'Haridon Sp, Godfroy A, Roussel EG, Cragg BA, Parkes RJ, and Toffin L. 2015. Evidence of Active Methanogen Communities in Shallow Sediments of the Sonora Margin Cold Seeps. *Applied and Environmental Microbiology* 81:3451-3459. 10.1128/aem.00147-15

von Wintzingerode F, Selent B, Hegemann W, and Göbel UB. 1999. Phylogenetic Analysis of an Anaerobic, Trichlorobenzene-Transforming Microbial Consortium. *Applied and Environmental Microbiology* 65:283-286.

Wagner M, Roger AJ, Flax JL, Brusseau GA, and Stahl DA. 1998. Phylogeny of Dissimilatory Sulfite Reductases Supports an Early Origin of Sulfate Respiration. *Journal of Bacteriology* 180:2975-2982.

Wegener G, Bausch M, Holler T, Thang NM, Prieto Mollar X, Kellermann MY, Hinrichs K-U, and Boetius A. 2012. Assessing sub-seafloor microbial activity by combined stable isotope probing with deuterated water and 13C-bicarbonate. *Environmental Microbiology* 14:1517-1527. 10.1111/j.1462-2920.2012.02739.x

Wegener G, Krukenberg V, Riedel D, Tegetmeyer HE, and Boetius A. 2015. Intercellular wiring enables electron transfer between methanotrophic archaea and bacteria. *Nature* 526:587-590. 10.1038/nature15733

White JR, Nagarajan N, and Pop M. 2009. Statistical methods for detecting differentially abundant features in clinical metagenomic samples. *PLoS Comput Biol* 5:e1000352. 10.1371/journal.pcbi.1000352

Wright JJ, Mewis K, Hanson NW, Konwar KM, Maas KR, and Hallam SJ. 2014. Genomic properties of Marine Group A bacteria indicate a role in the marine sulfur cycle. *ISME J* 8:455-468. 10.1038/ismej.2013.152

Yanagawa K, Sunamura M, Lever MA, Morono Y, Hiruta A, Ishizaki O, Matsumoto R, Urabe T, and Inagaki F. 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.

Yilmaz S, Haroon MF, Rabkin BA, Tyson GW, and Hugenholtz P. 2010. Fixation-free fluorescence in situ hybridization for targeted enrichment of microbial populations. *ISME J* 4:1352-1356.

Supplemental Table 1: Extracted DNA concentration per sample measured by fluorometer.

Sample	Extracted DNA (ng/µl)
eel932BC1	BD
eel932BC2	BD
eel932BC3	BD
Seep1a1441BC1	BD
Seep1a1441BC2	BD
DSS658BC1	BD
DSS658BC2	BD
DSS658BC3	BD
Delta495aBC1	BD
Delta495aBC2	0.05
FixedBulk1	0.52
FixedBulk2	0.55
FixedBulk3	0.70
UnfixedBulk2	0.80
UnfixedBulk3	1.34

Supplemental Table 2: Expected and recovered sequence abundances among the mock communities show differential taxonomic biases. *Fold Change* is calculated by dividing the experimentally recovered relative abundance by the expected relative abundance. Four mock communities were designed with a selection of common methane seep bacterial and archaeal taxa at different relative abundance ratios. Mock community analysis revealed that relative abundances of *Helicobacteraceae* (*Sulfurovum*), *Desulfobacteraceae* (Seep-SRB1) and *Desulfobulbaceae* (*Desulfobulbus*) had little amplification bias as compared to other mock community taxa (fold change ranges 0.93-1.42, where 1.00 means expected relative abundance was returned). ANME-1b plasmids were also overall well represented (fold change 0.64 to 1.42) across the range of expected relative abundances (1% to 20%). In contrast, ANME-2a/b and ANME-2c plasmids were always under amplified in all of the mock communities (fold change 0.32 to 0.81). These results do not appear to correlate to primer hits in the SILVA SSU r123 database, where 89.5% of ANME-2c sequences were hit by 515f and 87.1% by 806r, but 94.3% of ANME-2a/b were hit by 515f and 806r. ANME-2a/b was a better match to the EMP primers, but both taxa were under amplified in mock community analysis. Amplification bias was not always uniform, where some templates saw varied amplification response depending on initial relative abundance in the mock community. The ANME-1a plasmid was over-amplified (3.35-2.44 fold change; Sup Table 2) when the plasmid was at 5% relative abundance and lower (Mock Communities 1-3). However, Mock Community 4 with the highest relative abundance (20%) of ANME-1a plasmids saw templates amplified to the expected relative abundance (0.97 fold change). *Thaumarchaeota*: Miscellaneous Crenarchaeota Group followed a similar pattern to ANME-1a: where it was 1% expected relative abundance, the fold change is ~5, and where it was 10% expected relative abundance, the fold change was less pronounced (~1.5). MBG-D sequences were slightly over amplified when at 1% expected relative abundance, and slightly under amplified when at 42% relative abundance. Bias was consistent across mock community samples when the relative percentage of that group (e.g. *Thermoplasmatales*, 40%) was the same in both samples. This suggests that analysis based on relative abundance between samples can be applied as a means of comparison, as long as the environmental OTUs of interest are above the detection threshold. A study of EMP primers with a pelagic marine community also reported discrepancies between mock community bias and independently assessed environmental sample bias for a dominant community members (Parada et al. 2015). Parada et al. similarly conclude that over-amplification of certain community members, in their case *Gammaproteobacteria*, was the cause of lower

than expected recovery, rather than lack of SAR11 and SAR116. Our ANME-2c results, therefore, serve as yet another example of how key community members can be under-represented when exploring unknown microbial systems. The severity of this issue for future studies is dependent on the research question, interpretation approach, and the phylogenic bias imparted on community members of interest. The phenomenon of less pronounced bias when templates are at higher starting relative abundances could be explained by the reannealing inhibition affect of high copy number templates in mock samples (Suzuki & Giovannoni 1996). Due to low template of Magneto-FISH samples, PCRs were done for a total of 35 cycles. Since bias is positively correlated with number of cycles (Suzuki & Giovannoni 1996), lowering PCR amplification cycles could improve bias issues. The lack of statistically significant ANME-2c correlations is expected since this group was recovered in so few samples. ANME-1a, however, may suffer from the opposite problem where over-amplification in iTag datasets reduces the ability to determine patterns with other OTUs. As an analogy, if the ANME-2c population is an image with only a few pixels and the image of the ANME-1a population is an image with oversaturated pixels, then neither has a workable dynamic range for correlation analysis. The log transform operation performed on the data before correlation analysis can reduce the bias between high and low abundance OTUs to some degree, but may not be sufficient in all cases, such as with these two OTUs. Several approaches can ameliorate some of the issues within iTag sequencing datasets: (1) Optimization of PCR conditions and use of high-fidelity DNA polymerase for amplification in conjunction with the (2) creation and sequencing of a mock community, if there is *a priori* knowledge of the community composition; (3) Data transformation(s) before statistical analysis (i.e., square root, fourth root, or log transformations) and (4) examining the behavior of single OTUs across multiple samples/treatments may be more robust than direct comparison of OTUs within a single sample; (5) Whole-community comparisons (i.e. UniFrac, ANOVA, ANOSIM) to minimize single-taxon biases by including all taxa.

Plasmid Taxonomy		Mock 1	Mock 2	Mock 3	Mock 4
***Desulfobulbaceae* (DSB)**	*Expected*	***3.0%***	***3.0%***	***11.0%***	***11.0%***
	Experimental	3.0%	3.2%	11.6%	12.9%
	Fold Change	**0.99**	**1.05**	**1.05**	**1.17**
	Std. Dev.	0.06		0.10	
Helicobacteraceae	*Expected*	***25.0%***	***25.0%***	***1.0%***	***1.0%***
	Experimental	28.0%	23.4%	1.4%	1.2%
	Fold Change	**1.12**	**0.93**	**1.38**	**1.17**
	Std. Dev.	0.18		0.16	
***Desulfobacteraceae* (DSS)**	*Expected*	***9.0%***	***9.0%***	***31.0%***	***31.0%***
	Experimental	12.7%	11.0%	34.8%	31.5%
	Fold Change	**1.42**	**1.22**	**1.12**	**1.02**
	Std. Dev.	0.15		0.10	
ANME-1a	*Expected*	*1.0%*	*4.0%*	*5.0%*	*20.0%*
	Experimental	3.4%	9.9%	14.5%	19.4%
	Fold Change	**3.35**	**2.48**	**2.90**	**0.97**
ANME-1b	*Expected*	***1.0%***	***4.0%***	***5.0%***	***20.0%***
	Experimental	1.4%	3.4%	6.9%	12.8%
	Fold Change	**1.42**	**0.85**	**1.37**	**0.64**

ANME-2a/b	*Expected*	***6.0%***	***2.0%***	***30.0%***	***10.0%***
	Experimental	3.7%	0.6%	10.2%	6.2%
	Fold Change	**0.61**	**0.32**	**0.34**	**0.62**
ANME-2c	*Expected*	***3.5%***	***1.0%***	***15.0%***	***5.0%***
	Experimental	1.9%	0.5%	6.9%	4.1%
	Fold Change	**0.55**	**0.49**	**0.46**	**0.81**
Miscellaneous Crenarchaeota Group	*Expected*	***10.0%***	***10.0%***	***1.0%***	***1.0%***
	Experimental	14.8%	14.9%	5.9%	4.7%
	Fold Change	**1.48**	**1.49**	**5.87**	**4.69**
	Std. Dev.	0.01		0.22	
Thermoplasmatales	*Expected*	***41.5%***	***42.0%***	***1.0%***	***1.0%***
	Experimental	30.8%	30.7%	6.5%	6.1%
	Fold Change	**0.74**	**0.73**	**6.48**	**6.09**
	Std. Dev.	0.01		0.06	

Supplemental Table 3: Sequences per sample post processing. Total sequences per sample after mothur processing and 0.1% bulk sediment cutoff and total sequences remaining for the most abundant ANME-2c, SEEP-SRB1, and ANME-1a OTUs. We also performed a BLASTN (Madden 2002) search of all contigs from all samples against an in-house database of 155 ANME-2c 16S rRNA sequences of >500 bp. This yielded 1,395 iTag sequences with an e-value greater than or equal to 10^{-130}, corresponding to 99-100% sequence identity match to sequences from our ANME-2c database. We then tracked this set of BLAST match contigs through each step in the mothur pipeline, with a final result of 1,260 sequences remaining in this BLAST set from the original contig file. Thus 92% of our ANME-2c BLAST hit set remained through the mothur processing pipeline. This suggests that the lack of ANME-2c sequences in our downstream database was not due to spurious removal during sequence processing.

	after 0.1% removal	ANME-2c	SEEP-SRB1	ANME-1a
Sample ID	Sequences	Otu16818	Otu17765	Otu12964
ETR-D1-DSS658BC1	3460	0	120	67
ETR-D2-DSS658BC2	5310	0	193	78
ETR-D3-DSS658BC3	9859	0	393	804
ETR-E1-eel932BC1	6304	0	279	36
ETR-E2-eel932BC2	5253	0	100	176
ETR-E3-eel932BC3	3293	0	414	52
ETR-F1-FixedBulk1	12257	21	171	1263
ETR-F2-FixedBulk2	4957	63	70	596
ETR-F3-FixedBulk3	14400	27	215	1192
ETR-S1-Seep1a1441BC1	2562	0	200	64
ETR-S2-Seep1a1441BC2	7867	0	408	865
ETR-U2-UnfixedBulk2	14441	54	309	1124
ETR-U3-UnfixedBulk3	11447	32	244	1180
ETR-delta1-Delta495aBC1	1552	0	0	0
ETR-delta2-Delta495aBC2	3790	0	91	0
Total	171154	197	3207	7497
Percent of Unfixed Bulk Sediment	-	0.33%	2.14%	8.90%

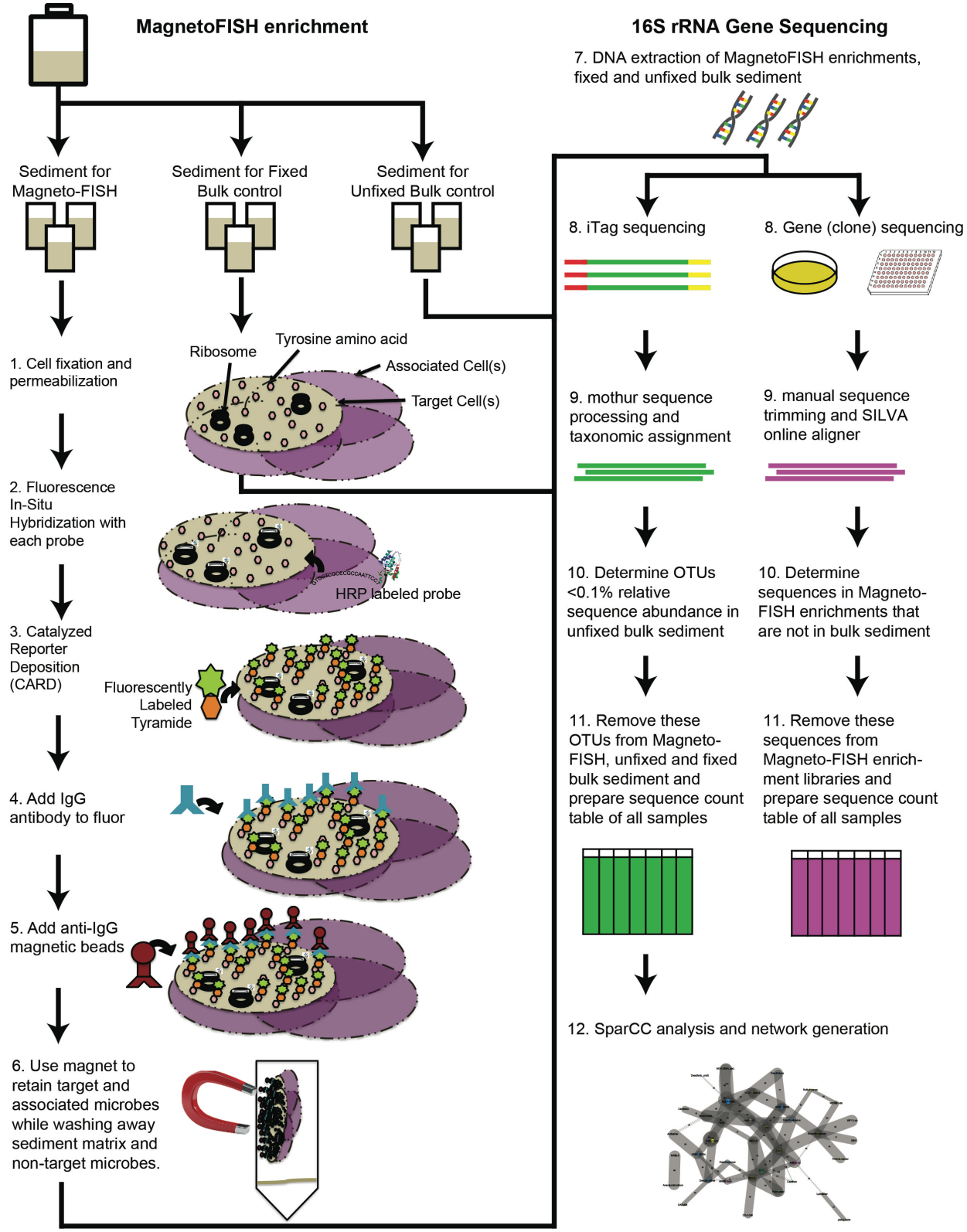
Supplemental Table 4: Mock Community sequencing error rates (0.025-0.095%; Sup Table 4) were of the same magnitude as Kozich et al. (~0.01%, 2013). Rarefaction of the mock community to 5,000 sequences shows OTU inflation rates of 3 to 4 times expected number of OTUs, after 97% OTU clustering and removal of singletons. The inflation rate is calculated by total number of OTUs recovered divided by original number of template plasmids. Since our environmental mock community only had 12 templates, the number of spurious OTUs is expected to be high (Huse et al. 2010). Experimental sediment samples have 10 to 100 times more templates, so inflation rates are expected to be much lower (10-1%).

Sample	OTU@5000	OTU@5000, singletons remv.	OTU@5000 singletons remv. inflation rate	Error Rate
Mock 1	65 (57-73)	32 (28-36)	2.7	0.025%
Mock 2	58 (50-68)	30 (26-34)	2.5	0.033%
Mock 3	113 (102-123)	38 (35-40)	3.2	0.095%
Mock 4	85 (77-94)	38 (35-41)	3.2	0.085%

Supplemental Table 5: All associations that occurred in at least 50 out of 100 networks for combined Magneto-FISH and bulk sediment samples with their number of occurrences, average correlation, and p-values.

OTU_1	Tax_1	OTU_2	Tax_2	Occurrence	Correlation	Pvalue
Otu072	ANME-1b	Otu090	JS1	100	0.746	0.002
Otu069	OD1	Otu077	Pelobacter	100	0.872	0
Otu014	AKAU3564_sed.	Otu082	SEEP-SRB1	100	0.729	0.001
Otu071	Hyd24-12	Otu083	SEEP-SRB1	100	0.765	0
Otu073	Sulfurovum	Otu087	Sulfurovum	100	0.798	0.001
Otu134	MBG-B	Otu135	OM1	99	0.737	0.001
Otu123	VC2.1_Bac22	Otu128	WS3	98	0.718	0.001
Otu106	OD1	Otu115	Desulfocapsa	97	0.705	0.002
Otu071	Hyd24-12	Otu097	Spirochaeta	96	0.708	0.002
Otu083	SEEP-SRB1	Otu097	Spirochaeta	96	0.691	0.001
Otu100	OD1	Otu107	Anaerolineaceae	95	0.709	0.001
Otu021	WF1B-44	Otu022	Anaerolineaceae	95	0.681	0.001
Otu068	Desulfoluna	Otu119	VC2.1_Bac22	95	0.692	0.001
Otu077	Pelobacter	Otu094	BD2-2	94	0.703	0.002
Otu073	Sulfurovum	Otu101	SEEP-SRB1	94	0.778	0.003
Otu097	Spirochaeta	Otu131	Actibacter	93	-0.688	0.002
Otu014	AKAU3564_sed.	Otu089	SEEP-SRB1	93	0.691	0.002
Otu122	SEEP-SRB4	Otu131	Actibacter	92	0.679	0.002
Otu112	SEEP-SRB1	Otu114	JS1	92	0.663	0.003
Otu023	Anacalomicrobium	Otu024	MSBL8	92	0.684	0.001
Otu102	SEEP-SRB1	Otu116	VC2.1_Bac22	92	0.701	0.001
Otu098	Desulfobulbus	Otu108	BD2-2	91	-0.668	0.003
Otu069	OD1	Otu086	OM1	91	0.688	0.003
Otu090	JS1	Otu114	JS1	90	0.685	0.002
Otu095	SEEP-SRB2	Otu119	VC2.1_Bac22	90	0.687	0.001
Otu069	OD1	Otu094	BD2-2	89	0.683	0.002
Otu091	SEEP-SRB2	Otu114	JS1	89	0.664	0.004
Otu082	SEEP-SRB1	Otu089	SEEP-SRB1	89	0.69	0.002
Otu071	Hyd24-12	Otu100	OD1	86	0.679	0.003
Otu091	SEEP-SRB2	Otu107	Anaerolineaceae	85	0.688	0.003
Otu087	Sulfurovum	Otu099	Sulfurovum	85	0.675	0.004
Otu094	BD2-2	Otu100	MBG-B	84	0.662	0.003
Otu081	Caldthrix	Otu086	OM1	84	-0.642	0.003
Otu127	Anaerolineaceae	Otu132	Thiohalobacter	84	0.651	0.002
Otu098	Desulfobulbus	Otu122	SEEP-SRB4	81	0.66	0.003
Otu091	SEEP-SRB2	Otu134	MBG-B	80	0.663	0.003
Otu105	CS-B046	Otu108	BD2-2	79	0.656	0.003

Otu077	Pelobacter	Otu086	OM1	79	0.655	0.003
Otu073	Sulfurovum	Otu098	Desulfobulbus	78	0.677	0.004
Otu016	Anaerolineaceae	Otu092	SB-5	77	0.649	0.003
Otu068	Desulfoluna	Otu095	SEEP-SRB2	77	0.658	0.002
Otu103	ANME-2a/b	Otu135	OM1	71	0.647	0.004
Otu126	WS3	Otu129	Clostridia	70	0.653	0.003
Otu090	JS1	Otu100	OD1	69	0.647	0.004
Otu122	SEEP-SRB4	Otu128	WS3	68	0.641	0.004
Otu027	Anaerolineaceae	Otu030	WCHB1-69	66	0.657	0.001
Otu112	SEEP-SRB1	Otu121	Desulfarc._unclt.	65	-0.648	0.004
Otu084	Desulfobulbus	Otu120	Desulfococcus	65	0.663	0.006
Otu051	Lutibacter	Otu096	ANME-1b	64	-0.636	0.004
Otu005	Desulfocapsa	Otu032	Hyd24-12	64	0.648	0.003
Otu011	Anaerolineaceae	Otu072	ANME-1b	63	-0.639	0.003
Otu072	ANME-1b	Otu100	OD1	59	0.637	0.005
Otu072	ANME-1b	Otu094	BD2-2	58	0.646	0.003
Otu124	WS3	Otu134	MBG-B	57	0.631	0.003
Otu020	pMC2A209	Otu051	Lutibacter	55	0.646	0.003
Otu113	Desulfobulbus	Otu118	SAR406	55	0.643	0.003
Otu036	Desulfocapsa	Otu046	Sulfurimonas	54	0.646	0
Otu069	OD1	Otu081	Caldthrix	52	-0.633	0.004
Otu108	BD2-2	Otu120	Desulfococcus	50	0.632	0.003
Otu103	ANME-2a/b	Otu122	SEEP-SRB4	50	0.635	0.004
Otu046	Sulfurimonas	Otu054	WCHB1-69	50	0.636	0.001



Supplemental Figure 1: Methods flow diagram.

Functional Gene	dsrA	aprA	soxB
Desulfobacteraceae (DSB)	15	7	
Desulfobulbaceae (DBB)	1	0	
Gammproteobacteria (Endosymbiont)		5	
Other (Cluster B)		1	
Epsilonproteobacteria (Sulfurimonas)			20
Epsilonproteobacteria (Sulfurovum)			2

Supplemental Figure 2: Phylogenetic trees of SoxB, AprA, and DsrA functional genes from a MSMX-Eel_932 Magneto-FISH enrichment. aLRT SH-like values above 50% displayed for branch support. Similar sequence clusters represented by one sequence are indicated in parentheticals, and sequences from this study are in bold. Clones recovered are also summarized in table form. As another method of assessing Magneto-FISH diversity, we examined functional genes relating to the sulfur cycle. This method can also provide insight into phylogenetic connections between 16S rRNA and sulfur cycling functional genes. Clone libraries were constructed from an MSMX-Eel_932 Magneto-FISH capture and performed with sediment as iTag libraries (see Materials & Methods). The following genes relating to sulfur cycling pathways were chosen for this analysis: soxB (sulfur oxidation, protein-S-thiocysteine sulfate hydrolase), aprA (sulfur oxidation and reduction, adenylylsulfate reductase α subunit), and dsrA (sulfur oxidation and reduction, dissimilatory sulfite reductase). Phylogenetic analysis of soxB clones from the MSMX-Eel_932 Magneto-FISH returned only *Epsilonproteobacteria*sequences from both *Sulfurovum* (2 clones) and *Sulfurimonas* (20 clones) clades (Sup Figure 1). From a total of 13 aprA clones, 7 were retrieved from the *Desulfobacteraceae* clade (SEEP-SRB1 containing), none from the *Desulfobulbaceae*, 5 from *Gammaproteobacteria* Endosymbiont clade, and 1 from the "Cluster B" GoM clone clade (Meyer & Kuever 2007) . 15 of 16 dsrA clones were from the *Desulfobacteraceae* clade, with one clone from the *Desulfobulbaceae* clade (Müller et al. 2015) . Functional gene clone libraries were not only successful in providing another means to assess Magneto-FISH enrichment, but provide an example of how this technique can be utilized to target specific 16S rRNA populations and the metabolic diversity contained. This is particularly useful in cases where 16S rRNA and functional gene phylogenies are not well aligned.

Chapter 3

JUST ADD (DEUTERATED) WATER:

PASSIVE TRACER AND MINIMAL CARBON AND NITROGEN AMENDED SIP-NANOSIMS INCUBATIONS PROVIDE AUTHENTIC ESTIMATES OF DEEP BIOSPHERE ACTIVITY IN 2,000 MBSF COALBEDS

Abstract

The past decade of marine scientific drilling has led to the discovery of seemingly ubiquitous microbial life in a range of deep biosphere habitats. In a quest for possible depth limits to deep life, IODP Expedition 337 successfully recovered core down to a record-breaking 2.5 km below seafloor from a deeply buried coalbed system with low thermal alteration. Isotopic signatures and taxonomic profiles suggested a typical a microbial assemblage, partnering fermentive bacteria and methanogenic archaea, which degrades coal into end products of methane, carbon dioxide, and hydrogen. The previously established global depth trend for subseafloor microbial abundance suggests ~ 10^5 cells/cm^3 at 2 kmbsf, instead cell abundances were 10 cells/cm^3 or less in most samples below 1.5 km suggesting life-limiting conditions may have been reached. The coalbeds, however, were comparative "hot spots" with cell concentrations 10 to 100 × higher than surrounding lithologies. Methy-compounds derived from coal decomposition may serve as potential carbon sources for this coalbed microbial community. To determine general microbial activity and methyl-compound utilization, 2.5 year-long Stable Isotope Probing (SIP) incubations were carried out at in situ temperatures (45 ^oC) with deuterated water (passive tracer) and minimal ^{13}C-carbon and ^{15}N-nitrogen amendments. Incubation geochemistry was non-destructively monitored during incubation to track activity, with NanoSIMS analysis performed on the most methane-producing incubations at the end of the incubation period. While cells were scarce, ranging from 50 to 2000 cells/cm^3 in the most active incubations, we estimated average hydrogen and nitrogen biosynthesis-based turnover times of less than a year to 63 years and less than a year to 2,020 years, respectively. NanoSIMS carbon biosynthesis-based turnover time estimates were orders of magnitude lower than previous deep biosphere studies, but bulk catabolic were rates similar to shallower hydrocarbon seep sediments. Our results support the concept of a deep biosphere community focused predominantly on maintenance over growth, but with much faster turnover times than previous estimates.

Introduction

Advances in deep sea drilling technology and microbiological methods have led to the discovery of microbial life in a range of deep biosphere habitats, from Earth's most oligotrophic sediments (IODP Expedition 329) to its largest aquifer (IODP Expedition 336). While approximations of total deep life cell abundance continue to evolve, currently estimated to be ~ 10^{29} cells in marine sediments (Parkes et al. 2014), the deep biosphere undoubtedly constitutes a massive percentage of life on Earth. With their continued detection on missions extending into older, deeper, and hotter habitats over the past decade of deep sea drilling, deep life has been surprisingly ubiquitous given their limited resources. While we have yet to determine what parameters limit deep biosphere cells, they are still less abundant and metabolize more slowly than cultured or surface microbes (Hoehler & Jorgensen 2013). Given that generation times are slower, and opportunities for microbe-microbe interaction are fewer, in the deep marine subsurface (low cell density, lack of fluid flow, expense of flagellar mobility, etc.), it is expected that genetic adaptation and horizontal gene transfer rates are minimal (Hoehler & Jorgensen 2013). It is therefore more likely that persistence in deep biosphere environments is achieved with pre-existing capacity, rather than cells undergoing a period of genetic evolution and adaptation (Biddle et al. 2011; Jørgensen & Marshall 2016). Though as we further constrain biosynthesis rate-based turnover times and cell distribution in the deep biosphere, these genetic evolution assumptions may require revision.

Deep biosphere cell abundance has been described by a global depth trend (log cells/cm^3 = 8.05 - 0.68 log depth (m); Parkes et al. 2014) that predicts a three order of magnitude drop from sediment surface (10^9) to 1 km below seafloor (bsf) (10^6). This trend likely describes factors that co-vary with depth, such as organic carbon availability (Lipp et al. 2008). Similarly connected to substrate availability, cell abundance can also be spatially correlated to distance from land and sedimentation rate (Kallmeyer et al. 2012). However, it is not known how cell enumeration relates to cell activity, of import for understanding the ability of deep life to recycle fossil organic carbon back to the surface biosphere. Subseafloor cell activity is often traced geochemically through the utilization of substrates, such as sulfate, per cell per unit time. Activity can be further parsed into catabolic rates (energy-generating substrate use) and anabolic

rates (biomass-generating substrate use). It is assumed substrate utilization will be diverted more toward energy generation to maintain cell viability than biomass generation since substrates are more limiting at depth. Carbon catabolic rates computed from sulfate reduction coupled to acetate oxidation at 1.5 mbsf are ~ 10^{-4} fmol C/cell/day when measured by ^{35}Sulfate incubation for Peru Margin sediments (Parkes et al. 1990) and ~ 10^{-6} fmol C/cell/day when measured by porewater fluxes for Pacific Gyre sediments (D'Hondt et al. 2002). These rates can be compared to ~ 10^{-2} fmol C/cell/day measured in shallow marine sediments of Aarhaus Bay by the same method (Leloup et al. 2009). From the radio-labeled sulfate reduction technique, it appears catabolic rates decrease with depth, but absolute rate estimates vary depending on study location.

Anabolic rates have been estimated by multiple means and are often converted to turnover times for a given element in biomass (H, C, or N). Bulk sediment D:L amino acid racemization modeling in eastern tropical Pacific sediments (~ 200 mbsf) resulted in degradation-based turnover times of 2,000-3,000 years, with total organic carbon turnover estimated at 43 million years (Lomstein et al. 2012). Estimates from biosynthesis based-turnover of microbial lipids have a similar lower estimate (1,600 years), but a much higher upper estimate (73,000 years; Xie et al. 2013). Natural abundance carbon single-cell-Secondary ion mass spectrometry (SIMS) combined with diagenetic models and a pure culture (*Acetobacterium*) slow metabolism proxy for required cellular maintenance energy, yield carbon degradation rate-based turnover times of 70 to 2,150 years (Biddle et al. 2006). Single-cell NanoSIMS carbon and nitrogen biosynthesis rates from (stable isotope probing) SIP incubations with Shimokita Peninsula sediments (219 mbsf) determined much faster turnover times, from 63 to 192 days (Morono et al. 2011). In these experiments, ^{13}C-glucose addition had the highest number of carbon incorporating cells (76%), suggesting most cells were viable when high-energy substrates were provided. Since these NanoSIMS single-cell anabolic rates were based on conditions with higher substrate concentrations than would have been accessible to the in situ microbial community, these results may inform viability more than in situ rates of metabolism.

To probe the depth limit of life, IODP Expedition 337 sampled a deeply buried (~ 2 kmbsf) coalbed system of terrestrial origin and low thermal alteration (~ 25 mya lignite to sub-bituminous, as described in Gross et al. 2015). In addition to increased depth (previous record

1.9 kmbsf; Ciobanu et al. 2014), this site is warmer, has more fossil organic carbon, and stands in contrast to previous marine drilling operations that have focused more on electron acceptor availability (D'Hondt et al. 2015; D'Hondt et al. 2004). It has been hypothesized that the slow temperature rise with burial, and concurrent thermogenic and biotic breakdown of recalcitrant organic matter, may provide a continued biomaterial source for deep life (Fry et al. 2009; Horsfield et al. 2006). The increased temperature may also reduce activation energy barriers, allowing reactions that may not have yielded energy at the sediment surface to become exergonic (Parkes et al. 2007). It is possible that microbial life has been living off this organic carbon since burial, since Shimokita coal has never been subjected to extended thermal alteration through its geologic history (< 60^{o}C; Gross et al. 2015; Inagaki et al. 2015; Konyukhov 2010). The temperatures (~ 45^{o}C) and pressures (~30 MPa) of the 2 km coalbed are also well within known limits of life for piezo-thermophilic life (120 MPa and 108oC; Zeng et al. 2009), but other parameters such as porosity or water availability may still be limiting. A study of the effect pore size has on cell habitability and mobility found particle diameters smaller than 2 μm create trapped cells, and the combined effects of particle size and pressure should render cells dead at 1 kmbsf (Rebata-Landa & Santamarina 2006).

Initial results from Expedition 337 showed cell abundances no longer tracked the global depth trend (log cells/cm^3 = 8.05 - 0.68 log depth (m); Parkes et al. 2014), suggesting that life-limiting conditions had been reached (Inagaki et al 2015). The coalbed, however, was a comparative "hot spot" for microbes with 10 to 1000 times higher cell concentrations than the adjacent shale and sandstone environments (≤ 10 cells/cm^3; Inagaki et al. 2015). Though even the coalbed cell abundances were much lower than the global depth trend estimate of ~ 5.5 × 10^5 cells/cm^3 for 2 kmbsf. Based on geochemistry, carbon and hydrogen should not be limiting at this depth. Since temperature, pressure, and substrate availability are well within life limits, pore space and/or water availability may be the limiting factors. Despite these low cell concentrations, geochemical and genomic data suggest an active microbial coalbed assemblage was not only present, but phylogenetically resembled the microbial community of its paleo-surface expression (terrestrial swamp; Inagaki et al. 2015). While no archaeal 16S rRNA genes were recovered with Illumina Tag (iTag) sequencing, geochemical and biomarker analyses supported the existence of an active methanogenic community, as well. Since recovered cells appear to be

from an indigenous community rather than later colonization of allochthonus microbes, the microbial community profile supports the idea of a microbial bioreactor buried and sustained for millions of years post deposition.

Many questions remain as to how these still active, coalbed cells are compared to other deep biosphere and surface environments. How is activity is distributed across the viable microbial assemblage? How is activity split between anabolic and catabolic modes? How are resources split between carbon and nitrogen acquisition? Answering these questions requires single-cell techniques to address sub-population dynamics and is most informative if the natural system is minimally perturbed. Measuring activity in the deep biosphere requires a balance between stimulating the population enough to observe a metabolic signal and over-stimulation of the system such that experimental concentrations no longer approximate in situ conditions, or are even detrimental (Postgate & Hunter 1963).

D_2O is a passive (non-nutrient), universal tracer with an extremely low natural abundance (0.015 %) for high detection rates. These combined factors are ideal for activity detection in slow growing, low biomass systems where the metabolic diversity may not be known a priori, such as the deep biosphere. D_2O utility can be extended by use in conjunction with other isotopically labeled or unlabeled substrates to characterize substrate specific activity along with basal anabolic rates for mixed-community, environmental samples. Utilization of D_2O as a passive tracer of anabolic activity has been successfully applied in pure culture and mammalian microbiome studies (Berry et al. 2015; Kopf et al. 2015; Kopf et al. 2016), soils (Eichorst et al. 2015), and methane-rich environments (Kellermann et al. 2016; Wegener et al. 2012). This technique can provide single-cell biosynthesis rates when used in conjunction with NanoSIMS (Kopf et al. 2015). D_2O incorporation can also be measured by other single-cell techniques such as Raman scattering (Berry et al. 2015), but this method requires much higher labeling for detection and is therefore not ideal for low activity systems.

Here, we conducted long-term (2.5 yr) SIP incubations with a passive tracer (D_2O) and minimal ^{13}C (30 μM C) and ^{15}N (3 μM N) amendments of environmentally relevant substrates to determine the in situ activity of microorganisms associated with a paleo-terrestrial swamp after

burial for millions of years. Incubations were prepared from six separate cores ranging from 1377 to 2466 mbsf and one blend of five lithologies from 1950 to 2000 mbsf. Two coalbeds were included in these seven incubations, a shallower (~ 1920 mbsf) coal-only sample deposited under more marine-influenced conditions, and a deeper coalbed (~ 2000 mbsf) deposited under more limnic conditions included in the mixed lithology sample. Fifty-two incubation carbon amendment and control conditions interrogating a range of potential deep biosphere metabolic strategies were prepared onboard (Inagaki et al. 2012). Methylamine and methanol, coalbed fermentation and degradation byproducts, and methane, terminal product from utilizing methyl-compounds (Strąpoć et al. 2011), were used as the basis for the methylotrophic subset of incubations returned to Caltech. Of these amendments, methylamine and methanol were the most active based on geochemical monitoring during incubation, and are discussed here. Our SIP incubations amended with methyl substrates, deuterated water, hydrogen and ammonium provide catabolic rates and basal and substrate-specific biosynthesis rates of single cells from the deepest samples ever retrieved by scientific ocean drilling, showing distinct responses in microbial activity and substrate utilization between coalbeds. The reciprocal of these rates also provides estimation of biosynthesis-based turnover times for hydrogen, nitrogen, and carbon that are faster than previous deep biosphere estimates by other methods.

Materials and Methods

Incubation preparation

IODP Expedition 337 operations commenced July 26 and continued through September 30, 2012 on the *D/V Chikyu*. Utilizing riser drilling, we penetrated a sedimentary sequence down to 2466 mbsf. The borehole site is off the Shimokita Peninsula in 1180 m water (Supplemental Figure 1). With depth, the drilled sequence transitions from open marine (youngest; late Pliocene, ~ 5 Ma) to terrestrial (oldest; late Oligocene, ~ 30 Ma) (Figure 1; adapted from Gross et al. 2015). Shipboard sedimentological and geochemical analysis and microbiological methods (including contamination controls) can be found in the Expedition 337 Proceedings volume (Inagaki et al. 2012). Additional coal petrography is available in Gross et al. 2015.

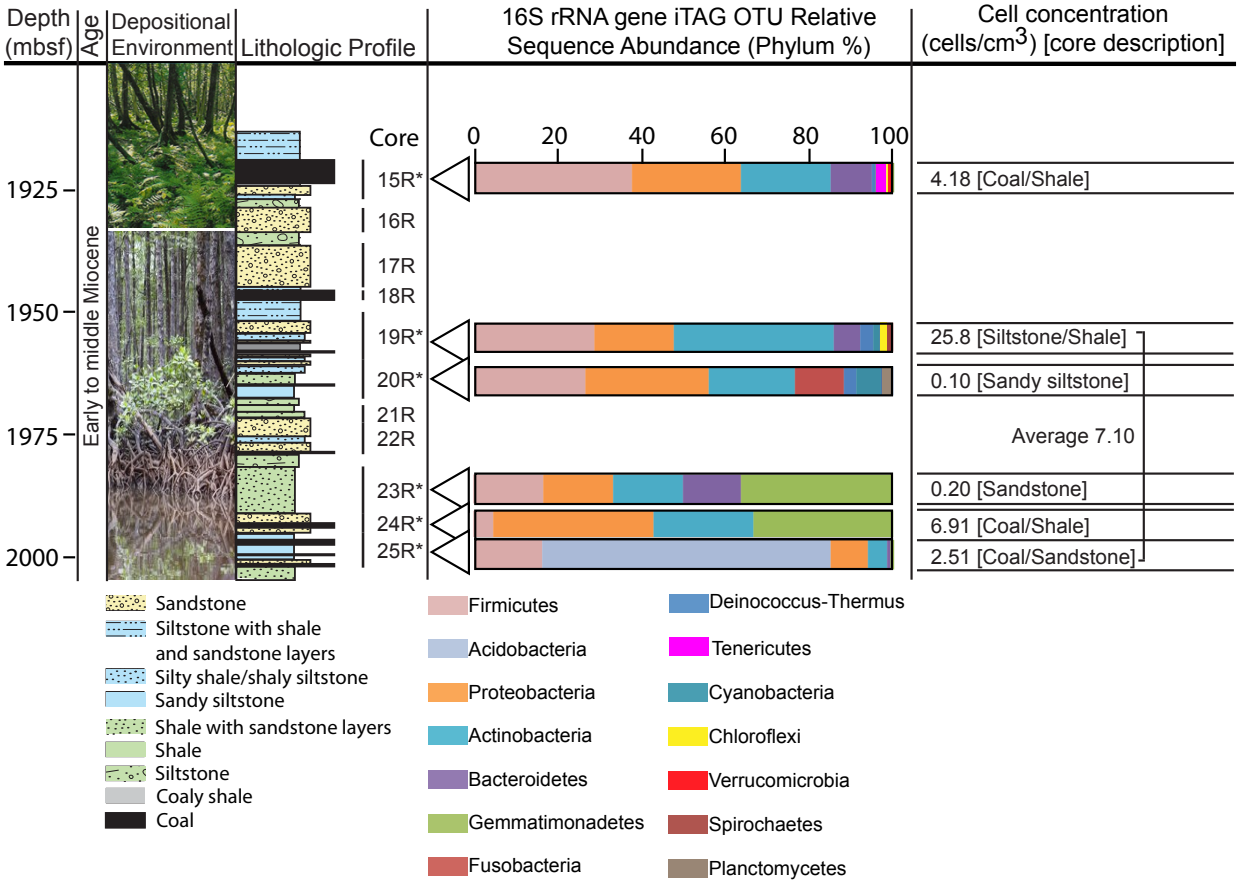


Figure 1: Description of samples collected on IODP Expedition 337 that were incubated for SIP-NanoSIMS. Sample depth, biostratigraphic age, inferred depositional environment, and lithologic profile with location of main coal bearing cores (adapted from Gross et al 2015). *Indicates cores used for SIP NanoSIMS analysis, where 15R was used for coal-only incubations and the remaining cores were combined in mixed lithology samples. 16S rRNA gene iTag relative sequence abundance profile demonstrates phylum level distinctions between lithologies sampled. Cell concentrations are most likely indigenous estimates (Inagaki et al. 2015) with broad core description in brackets.

Two sample types, of the seven prepared onboard, were analyzed by NanoSIMS in this study. The first was a coal-only sample from core 15R (1921 mbsf), termed "15R3 coal" (Figure 1). The second was a homogenized core mixture from multiple horizons (19R1, 19R5, 19R7, 20R3, 23R6, 23R8, 24R3, 25R1, 25R2, and 25R3; spanning 1950-1999 mbsf) containing lower coalbed samples, termed "mixed lithology" (Figure 1; *). All incubated samples are early to middle Miocene. In situ temperatures ranged from 46 to 48 ^{o}C in the sampled region (Supplemental Table 1). Porosity was highest in the sandstone layers and lowest in the coal/shale layers. Formation water sampling yielded pressures between 31 and 32 MPa (~ 300 atm) at the top and bottom of sample cores (Inagaki et al. 2012).

Incubations were prepared as described in the "Single-cell analyses of carbon and nitrogen assimilation rates of subseafloor microbes" section of cruise Methods (Expedition 337 Scientists 2013). Briefly, outer drill-fluid contaminated layers of core samples were removed under nitrogen and the remaining core was manually crushed into cm-sized pieces under sterile, anaerobic conditions and distributed evenly into 50 ml glass vials with butyl rubber stoppers and screw caps (NICHIDENRIKA-GLASS Co. Ltd.). Vials were flushed with argon and pressurized to 1 atm headspace. Minimal C, N, and S-free, 20% D_2O (Cambridge Isotopes) media was prepared anaerobically and 20 ml was added to each vial. 20% D_2O had little to no (depending on organism) observable effect in pure culture evaluations of toxicity (Berry et al. 2015; Kopf et al. 2015). A timepoint 1, timepoint 2 and autoclaved vial were prepared for each substrate condition. Timepoint 1 was incubated for six months. Timepoint 2 and autoclaved samples were incubated for 2.5 years and were used for NanoSIMS analysis.

Amendments and incubation conditions for the methyl-substrate Caltech subset are listed in Table 1. A full list of the incubation conditions prepared onboard are in cruise Methods (Expedition 337 Scientists 2013). All incubations listed in Table 1 were incubated at the cruise-measured coalbed in situ temperature of 45 ^{o}C. Equimolar amounts, 30 μmol C and 3 μmol N, were used across incubation conditions and were provided at 50 at. % (Cambridge Isotopes). $^{13}C/^{15}N$-Methylamine (with and without hydrogen amendment, added as 5 ml overpressure to 30 ml headspace volume) was provided as both a carbon and nitrogen source, which could be used for catabolism (methanogensis) or anabolism. ^{13}C-Methanol was provided as a carbon source for catabolism or anabolism in two conditions, one with ^{15}N-ammonium and hydrogen amendments and the other with only ^{13}C-methanol. Background ammonium was ~ 3 mM (as measured by IC) in all incubations, so the 3 μmol ammonium addition at 50 at. % was effectively diluted to 2.7 at. % by the in situ ammonium concentration.

Sampling

After 17 months of incubation (March 2014), all treatments were non-destructively sampled for geochemical analyses. 3 ml of headspace gas was removed to a vial filled with 0.1 M NaOH for methane analysis. The incubation was shaken to suspend smaller sediment particles before liquid was sampled. The liquid was filtered through a 0.1 μm 13 mm Whatman Polycarbonate

Nuclepore Track-Etched Membrane Filter (110405), aliquoting filtrate for DIC and IC analysis. IC samples were frozen and stored at -20 ^{o}C until analysis.

Dissolved Inorganic Carbon Concentration and Isotopic Analysis

DIC samples were injected into prepared He flushed exetainer vials (Labco Limited, Buckinghamshire, UK) containing 100 µl 40% phosphoric acid. Concentrations and stable carbon isotopes of DIC were measured on a Gasbench II (Thermo Scientific, Bremen, Germany) coupled to a Delta V Plus IRMS (Thermo Scientific), following the methods of Torres et al. (2005). Concentrations of DIC were determined based on comparison of the total peak area (masses 44, 45, 46) of replicate sample injections to a standard curve generated from a laboratory standard of $NaHCO_3$. DIC standards were prepared from a solution of 20 mM sodium bicarbonate at 0, 1, 2, 5, 10, 15, and 20 µmol DIC. A 3-pt correction (Coplen et al. 2006) was then applied using NBS 19, internal laboratory standards, and the average of all of the sodium bicarbonate standards. Accuracy was reported as standard deviation (± 0.5 ‰) of the measured $\delta^{13}C$ from NBS-19 and internal laboratory standards as compared to their published values and EA measurements, respectively. Using the estimated error in the volume measurement and the standard deviation of the peak area for the total carbon of bicarbonate standards, the estimated error was ± 0.009 µmol DIC. Labeled substrate solutions were also evaluated for background DIC contamination that could lead to false detection of organic substrate conversion to DIC. Labeled substrate solution DIC was determined to be insignificant.

Methane Concentration and isotopic analysis

Methane headspace concentrations were measured relative to argon (added to circum-equal partial pressure in all incubations) using a Hewlett Packard 5972 Series Mass Selective Detector and Hewlett Packard 5890 Series II Plus Gas Chromatograph at the Caltech Environmental Analysis Center (Supplemental Figure 2). $\delta^{13}C$-CH_4 was analyzed as per methods in Toki et al. (2014) and D-CH_4 was analyzed as per methods in Kikuchi et al. (2016) at Kochi Institute for Core Sample Research, Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Kochi, Japan.

Ion Chromatograph

Parallel Dionex DX-500 (Sunnyvale, CA, USA) ion chromotagraphy systems (Caltech Environmental Analysis Center) were used to measure cations and anions in the incubation filtrate. The anions chloride, acetate, propionate, butyrate, sulfate, oxalate, fumerate, phosphate, and citrate were resolved using an AS11-HC RFIC Analytical column (4 x 250 mm) with an AG11-HC guard column. The cations sodium, potassium, strontium, magnesium, calcium, ammonium, and methylamine were resolved using a CS16 Analytical column. Analyte peak area relative to chloride peak area was used for between sample comparison, rather than absolute concentrations.

Bulk Carbon and Nitrogen Isotopic Measurements with Elemental Analyzer

Total nitrogen, total carbon, and organic carbon were measured by Elemental Analyzer (EA) from 1 - 4 mg sample. Organic carbon was determined by removal of inorganic carbon, dissolved by acidification with 85% phosphoric acid for 2 hours at 50 °C. All samples were lyophilized and transferred to tin capsules (9 × 5 mm). $\delta^{13}C$ and $\delta^{15}N$ of bulk organic matter and weight percent total organic carbon and N (wt.% TOC and TON) were determined 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. Fast jump was calibrated and applied to switch between N_2 and CO_2 configurations to measure both in the same run. $\delta^{13}C$ and $\delta^{15}N$ values were measured on a Thermo Scientific Delta V Plus irMS. $\delta^{13}C$ and $\delta^{15}N$ values were corrected for sample size dependency and then normalized to the VPDB scale with a two-point calibration (Coplen et al. 2006). Error was determined by analyzing potassium nitrate (NIST ref no: 8549), sucrose (NIST ref no: 8542), and acetanilide (Costech) in combination with in-house standards (urea, glycine, and Hydrate Ridge sediment). Accuracy for C was monitored across all EA analyses and was determined to be 0.14 ‰ (n = 27) and precision was 0.34 ‰ (n = 27, 1σ). Accuracy for N was monitored across all EA analyses and was determined to be 0.16 ‰ (n = 24) and precision was 0.55 ‰ (n = 24, 1σ).

Isotope mass balance was calculated using Equations 2 and 3, where n is the number of moles, F is the fractional abundance (Equation 1) of the rare isotope, and R is the ratio of ion counts of the rare isotope over ion counts of the more abundant isotope (Hayes 2004). F_{total}, n_{total}, $F_{substrate}$, and $F_{background}$ were known. $n_{substrate}$ was solved for given a range of $F_{background}$ sources. $n_{substrate}$ was then divided by the number of days incubation and the total number of cells (cells/cm^3 × cm^3 rock incubated; ~ 7 cm^3) to calculate per cell rates. $F_{background}$ ranged from EA measurements for coal samples to cruise measurements for methane. Natural abundance, background carbon, and stable isotope standard ratios and fractional abundances used are provided in Supplemental Table 1.

$$Equation\ 1:\ F = \frac{R}{1+R}$$

$$Equation\ 2:\ n_{total} = n_{substrate} + n_{background}$$

$$Equation\ 3:\ n_{total} \times F_{total} = n_{substrate} \times F_{substrate} + n_{background} \times F_{background}$$

Sample preparation for NanoSIMS: preservation, separation, enumeration, and FAC sorting

Cell preservation, separation, enumeration, and sorting were all conducted in the clean booth and clean room facilities at Kochi Institute for Core Sample Research, Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Kochi, Japan. 3 ml of incubation headspace was sampled for methane isotope analysis from all coal samples prior to the termination of the incubation experiment (see Methane analysis methods). Half of the solid and half of the liquid portion of each sample were fixed in one 2% paraformaldehyde (PFA) : 3 × phosphate buffered saline (PBS) aliquot overnight. Samples were then subjected to two washes, incubating in 3 × PBS for 6 hrs and then 2 hrs, after each wash respectively. Samples were centrifuged (3500 × g) and supernatant was decanted after each wash. The other half of the sample was preserved in glyTE (70% glycerol, 100mM Tris, 10mM EDTA; Biglow Single Cell Genomics Center preservation protocol) and frozen by cell alive system (CAS) freezing and stored at -80 ^{o}C (Morono et al. 2015).

1 ml rock slurry and ~ 1 g rock chips were subsampled by pipet and sterile cell culture loop, respectively, from the PFA-fixed sample. Cell separation, microscopy, and sorting procedures

followed Morono et al. (2013), with the following modifications: 1) samples were sonicated (Model UH-50, SMT Co. Ltd.) in an ice bath for 20 cycles of 30 sec 200 W, 30 sec off and 2) samples were incubated in hydrofluoric acid post initial sonication, rather than after first density gradient separation. Cell detection limit was 2 cells per entire filter area (diameter 15 mm) as determined by negative controls.

Cells were stained with SYBR Green I (1:40 dilution SYBR Green : TE) and sorted using the cytometry protocol of Morono et al. (2013), with sorted cell collection on indium tin oxide (ITO) coated 0.2 μm polycarbonate filter membranes for direct transfer to NanoSIMS as per Morono et al. (2011) and Inagaki et al. (2015). These filters are termed "NanoSIMS membranes." Indium Tin Oxide (ITO) coatings on polycarbonate membranes (Isopore GTBP02500 Millipore) was developed by sputtering deposition technique at Astellatech Co. Ltd. (Kanagawa, Japan). SEM was imaged with a Zeiss 1550 VP Field Emission Scanning Electron Microscope at the Caltech GPS Division Analytical Facility and SYBR stained cells were imaged with a BX51 epifluorescence microscope (Olympus, Shinjuku, Japan) using 20x (UPlanFL N) dry, 60x (PlanApo N), and 100x (UPlanFL N) oil immersion objectives (Figure 2).

NanoSIMS instrument tuning and analysis

Cell targets were identified (by SYBR stain) and marked on NanoSIMS membranes with a laser dissection microscope (LMD6000; Leica Microsystems) for ease of rediscovery on the NanoSIMS. Samples were analyzed by raster ion imaging with a CAMECA NanoSIMS 50L at the Caltech Microanalysis Center in the Division of Geological and Planetary Sciences. A focused primary Cs^+ beam of ~ 1 pA was used for sample collection, with rasters of 256 × 256 or 512 × 512 pixels. ^{1}H (EM#1), ^{2}H (EM#2), $^{12}C_2$ (EM#3), $^{13}C^{12}C$ (EM#4), $^{12}C^{14}N$ (EM#5), and $^{12}C^{15}N$ (EM#6) were measured simultaneously (see Kopf et al. 2015 for technical development). Collection began after a pre-sputtering of equal intensity to one collection frame (~ 45 min). Recorded images and data were processed using Look@NanoSIMS software (Polerecky et al. 2012). Images were deadtime corrected and individual ion image frames were merged and aligned using the $^{12}C^{14}N$ ion image to correct for drift during acquisition. Single-cell based regions of interest (ROIs) were determined by "interactive thresholding" with the $^{12}C^{14}N$ ion image. Final ion images and counts per ROI were calculated by summation of ion counts for

each pixel over all scans, as shown in Figure 3. Outputs for ROI size and length to width ratio were used to compute cell diameters (Supplemental Figure 3).

A background correction was applied to the $^{13}C/^{12}C$ ratios of cells to correct for instrumental isotope fractionation and to address the low levels of ^{13}C enrichment in cells relative to the background carbon of the ITO coated polycarbonate membrane. In each NanoSIMS frame, an elliptical ROI was drawn in a region with no cells or particles to establish the background $^{13}C/^{12}C$ ratio. A two-point correction was applied to cells using NanoSIMS measurements of the ITO membrane and *Clostridia* spores, both of which were independently measured by EA-IRMS (filter: ^{13}R=0.01067; spores: ^{13}R = 0.01099; see Elemental Analyzer methods). A filter-only correction (no spores) was used for nitrogen, but was not necessary for deuterium because higher signal to background ratio. We also confirmed that cell ROIs had a total C to total N ratio that was distinct from the background correction ROIs and the coal C:N to ensure cell ROIs were all measuring biomass targets.

ROIs were then filtered based on the theoretical precision of the mean for the minor ion, which is equal to square root of the ion count and provided as the Poisson error by Look@NanoSIMS. As a conservative estimate, only ROIs where the deuterium ion count was more than 10 times the Poisson error are presented as violin plots in Figure 4. Data manipulation and display as violin plots of the kernel density function was done using "R" (R Core Team 2015) with the "ggplot2" (Wickham 2009), "dplyr" (Wickham & Francois 2015), "gridExtra" (Auguie 2012), and "RColorBrewer" (Neuwirth 2014) packages. The fractional abundance of all ROIs collected is presented in Supplemental Figure 4, for comparison. Violins are trimmed to the range of the data in both figures and scaled to equal width for visibility in Figure 4 and scaled to number of observations per sample in Supplemental Figure 4 to compare distribution of ROIs between samples.

Equations 5 and 6 were used for biosynthesis production rate calculations, where μ is the biosynthesis production rate (encompassing both cell maintenance and generation of new cells), T_{final} is the length of the incubation, F_{label} is the labeling strength, F_{final} is the single-cell

NanoSIMS measurement, and F_{nat} is the natural abundance. To estimate biomass turnover time, we used Equation 7 ($\tau, Equation\ 7$) as per Zilversmit et al. (1943).

$$Equation\ 5:\ {}^{2}\mu = \left(-\ln\left(1 - \frac{\left({}^{2}F_{final} - {}^{2}F_{nat}\right)}{a_w\left({}^{2}F_{label} - {}^{2}F_{nat}\right)}\right)\right)\Big/ T_{final}$$

$$Equation\ 6:\ {}^{15}\mu = \left(-\ln\left(1 - \frac{\left({}^{15}F_{final} - {}^{15}F_{nat}\right)}{\left({}^{15}F_{label} - {}^{15}F_{nat}\right)}\right)\right)\Big/ T_{final}$$

$$Equation\ 7:\quad \tau = \mu^{-1}$$

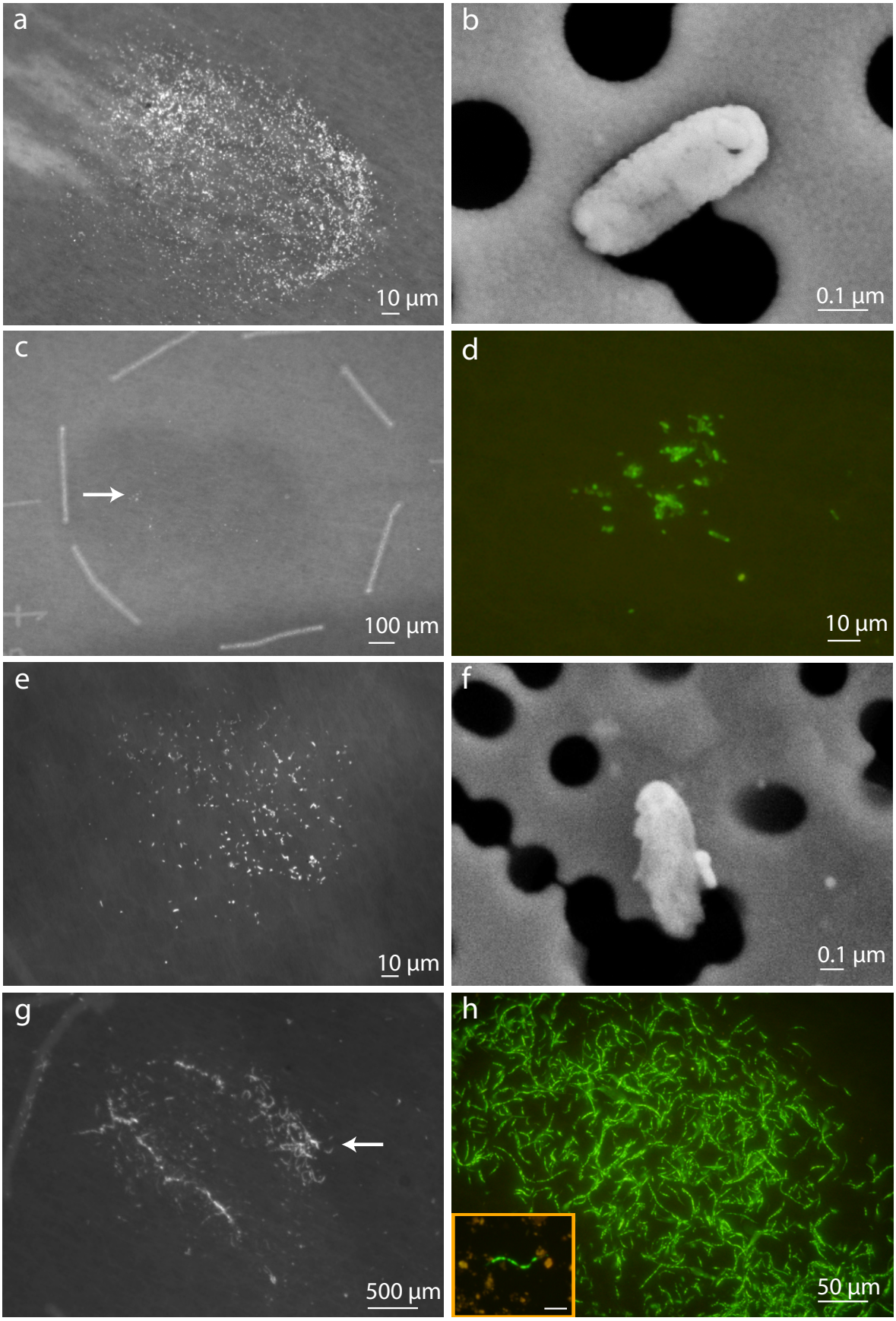
NanoSIMS ${}^{2}F$ and ${}^{15}F$ values were multiplied by a conversion factor determined for single cell to bulk isotope measurements in Kopf et al. (2015) of 0.67 (^{2}H) and 0.94 (^{15}N). Since the metabolic pathways utilized by cells in these incubations were unknown, and the proportion of water derived hydrogen in lipids is related to metabolism (Zhang et al. 2009), ^{2}F-based turnover times were calculated using Equation 5 from Kopf et al. (2015) and the full range of lipid water assimilation values (a_w; 0.44 to 0.83) from Zhang et al. (2009). Nitrogen turnover rate calculations assumed all nitrogen was derived from the substrate (methylamine or ammonium), therefore the assimilation constant was excluded from Equation 6. To determine per cell rates, we used values of 5.41 × 10^{-15} moles C per cell and 1.43 × 10^{-15} moles N per cell, which were calculated from 86 fg C/cell and 20 fg N/cell, respectively (Whitman et al. 1998).

Results

Cell Enumeration and Microscopy

Final cell concentrations were computed from cell separates of all incubations (Table 1). By comparison with average cell abundance from bulk core analysis (Figure 1), five (of 8) 15R3 coal incubations had cell concentrations above the method detection limit (1 cell/cm^3) and three incubations were above the background cell concentration estimate (4 cells/cm^3). Mixed lithology had two incubations above detection (of 8), and only one incubation was above background cell concentration (7 cells/cm^3). The highest incubation cell concentration was from

15R3 coal amended with methylamine (1921 cells/cm^3). An extrapolation for maximal doublings if growth began at onset of incubation is also provided for incubations where final cell abundance was above bulk core cell abundance (Estimated doubling time; Table 1). Estimates range between 3 and 8 months, out of the 29-month incubation period.


a
10 µm
b
0.1 µm
c
100 µm
d
10 µm
e
10 µm
f
0.1 µm
g
500 µm
h
50 µm

To overcome technical challenges for NanoSIMS analysis of low biomass samples, cell separation and FAC sorting were used to directly concentrate cells in a small analysis area (~ 1 to 0.5 mm^2). This reduces instrument time required to search for microbes by 15-20 times and increased the number of cells per NanoSIMS field of view (0.1 mm^2 out of entire region of 15 mm^2). NanoSIMS membranes (see Methods) were prepared from paraformaldehyde-fixed cell separates of 15R3 coal and mixed lithology samples after 894 days incubation. Examples of concentrated cell separation preparations for the highest cell count samples of 15R3 coal and mixed lithology incubations are provided in Figure 2. When viewed with SYBR staining, most cell morphologies were coccoid (singlets or doublets) or rods (Figure 2 d, e/f), with the exception of incubations that also contained filamentous cells (Figure 2a, b/c). Figure 3c inset is an image of unsorted cell separation, demonstrating how cell sorting both concentrates cells and removes non-target (orange) particles. NanoSIMS cell ROIs were also used to compute cell diameter (length) and length to width ratios (Supplemental Figure 4). Cell lengths fell within the deep biosphere cell size range of 0.2 to 2.1 μm (Kallmeyer et al. 2012), except for the methanol amended mixed lithology incubations that were dominated by filaments. These filamentous cells were also seen in the JAMSTEC collection of mixed lithology incubations with non-methyl substrates (data not shown). Length to width ratios were also within deep biosphere estimates of less than 3:1, with the same filamentous cell exclusion.

Bulk Catabolic Rates

Methane concentration relative to Ar headspace was measured for all coal and mixed lithology samples to determine methanogenic activity (Supplemental Figure 2). While exact concentrations were not determined due to variation between bottles in headspace partial pressures from incubation gas production, 15R3 coal with methylamine had the highest relative

←Figure 2: SYBR stained cells after separation and FACS concentration on ITO coated 0.2 μm polycarbonate NanoSIMS membranes from the two highest cell abundance conditions in 15R3 coal and mixed lithology. SYBR and SEM images highlight some distinctions in cell morphology between the two methanol amended samples. White arrows indicate region of SYBR image on larger membrane target area. a) 15R3 coal amended with methylamine, b) SEM image of a., c) 15R3 coal amended with methanol, d) c. magnified with false color, e) mixed lithology amended with methanol + H_2 + ammonium, f) SEM image of e., g) mixed lithology amended with methanol, and h) mixed lithology amended with methanol magnified with false color (inset is the same sample filtered before cell sorting to demonstrate cell density without concentration, inset bar is 20 μm).

methane concentration of all incubations. Lower, but detectable, levels of methane production were also measured in mixed lithology incubations. Since methane is the expected end product of methylotrophy from the labeled carbon substrates provided, 15R3 coal and mixed lithology were selected for NanoSIMS analysis based on their methane production. Methane ^{13}C and ^{2}H enrichment was also measured for all coal incubations ($\delta^{13}C$-CH_4 and δD-CH_4; Table 1). While incubation methane values (-42.8 to -57.5‰) were enriched in ^{13}C relative to the natural abundance methane values sampled onboard (~ -60 ‰; Inagaki et al., 2015), the minimal ^{13}C enrichment indicated that methane generated in the incubations was primarily derived from sources other than ^{13}C methylated substrates. Incubations were similarly slightly enriched in deuterium (-96.5 to -198.2‰) relative to in situ methane (~ -190 ‰; Inagaki et al., 2015), but unlikely high enough to have been exclusively produced in incubations containing 20% deuterated water.

^{13}C-Methane amended incubation $\delta^{13}C$-DIC ranged from -9 to -3‰. Therefore, enriched ^{13}C-DIC was not seen in any ^{13}C methane incubations that would be expected from active methanotrophy, and these samples were not chosen for further analysis. Methylamine and methanol, however, can be disproportionated to both methane and DIC. Some enrichment in ^{13}C-DIC was observed with methyl-substrates ($\delta^{13}C$-DIC; Table 1), though similar to methane enrichment, ^{13}C-DIC enrichment did not indicate significant DIC was derived from the provided ^{13}C substrate. ^{13}C enrichment above the D_2O-only control condition was higher in methanol-amended conditions than methylamine amended conditions for both 15R3 coal and mixed lithology incubations. DIC concentration was also measured and converted to a per cell production rate (Table 2) for samples with DIC concentration above the D_2O-only condition concentration (0.016 mM DIC 15R3 coal, 0.16 mM mixed lithology). DIC production was 1-3 orders of magnitude higher in mixed lithology than the 15R3 coal samples. Mixed lithology incubations amended with methanol ± hydrogen and ammonium had DIC production of 12 - 0.88 pmol C/cell/day above the mixed D_2O-only condition and were the most enriched in $\delta^{13}C$-DIC of all 15R3 and mixed lithology incubations. 15R3 coal amended with methylamine was the only 15R3 coal sample with DIC production (0.01 ± 0.001 pmol C/cell/day) above the 15R3 coal D_2O-only condition, but it did not produce significant enrichment in the bulk $\delta^{13}C$-DIC pool. While these ^{13}C enrichments would be significant for natural abundance measurements, they are

not significant considering the 50 at. % label used. Since DIC production was not solely from labeled substrates, isotope mass balance calculations were done with a range of potential alternative organic substrates that may also be available in the incubations (methane, -60 ‰; coal derived carbon source, -40 ‰; coal -24 ‰; Supplemental Table 1 for ^{13}R and ^{13}F) that could have been used in conjunction with the labeled substrate to produce the final DIC concentration and isotopic ratios (Table 2). The resulting catabolic rates range from ~ 2.1 to ~ 0.01 fmol C per cell per day for mixed lithology and coal samples, respectively.

Table 1: Incubation lithology, ^{13}C source, ^{15}N source, indication of autoclaved or H_2 added, final cell abundance, extrapolation of maximal days per cell doubling given background and final cell concentrations, and incubation $\delta^{13}C$-DIC (‰), $\delta^{13}C$-CH_4 and δD-CH_4. The cell detection limit was 2 cells per filter area, or ~ 1 cell/cm^3. Errors on DIC measurements are ±0.5 ‰ $\delta^{13}C$-DIC and ±0.0086 µmol DIC. Days incubation at time of DIC measurement was 864 days. *Cell abundances are not from D_2O-only condition but are averaged most likely cell concentration estimates across whole cores from Inagaki et al. (2015) to account for potential heterogeneity between bottles in determining a background cell concentration estimate. MeAm – Methylamine, MeOH – Methanol, Am – Ammonium.

Lith.	^{13}C Source	^{15}N Source	Autocl.	H_2	Abund. (cells cm^{-3})	Est. doubling time (mo.)	$\delta^{13}C$-DIC (‰)	$\delta^{13}C$-CH4 (‰)	δD-CH_4 (‰)
Coal	MeAm	MeAm	Y	Y	BD		-6.0	-42.8	
Coal	MeAm	MeAm	N	Y	4		-3.8	-46.1	-170.4
Coal	MeAm	MeAm	Y	N	BD		-16.2	-57.2	
Coal	MeAm	MeAm	N	N	1921	3.3	-7.2	-55.7	-198.2
Coal	MeOH	Am	Y	Y	BD		19.1	-49.7	
Coal	MeOH	Am	N	Y	2		-6.5	-52.9	-96.5
Coal	MeOH	Am	Y	N	48	8.4	-15.1	-57.5	
Coal	MeOH	Am	N	N	56	7.9	26.7	-56.6	
Coal	*D_2O-only*		*Y*	*N*	*4**		*-10.3*		*-189.0*
Mixed	MeAm	MeAm	Y	Y	BD		5.6		
Mixed	MeAm	MeAm	N	Y	BD		3.3		
Mixed	MeAm	MeAm	Y	N	BD		5.7		
Mixed	MeAm	MeAm	N	N	BD		6.8		
Mixed	MeOH	Am	Y	Y	BD		6.3		
Mixed	MeOH	Am	N	Y	4		14.9		
Mixed	MeOH	Am	Y	N	BD		9.4		
Mixed	MeOH	Am	N	N	129	8.0	44.6		
Mixed	*D_2O-only*		*Y*	*N*	*7**		*4.3*		

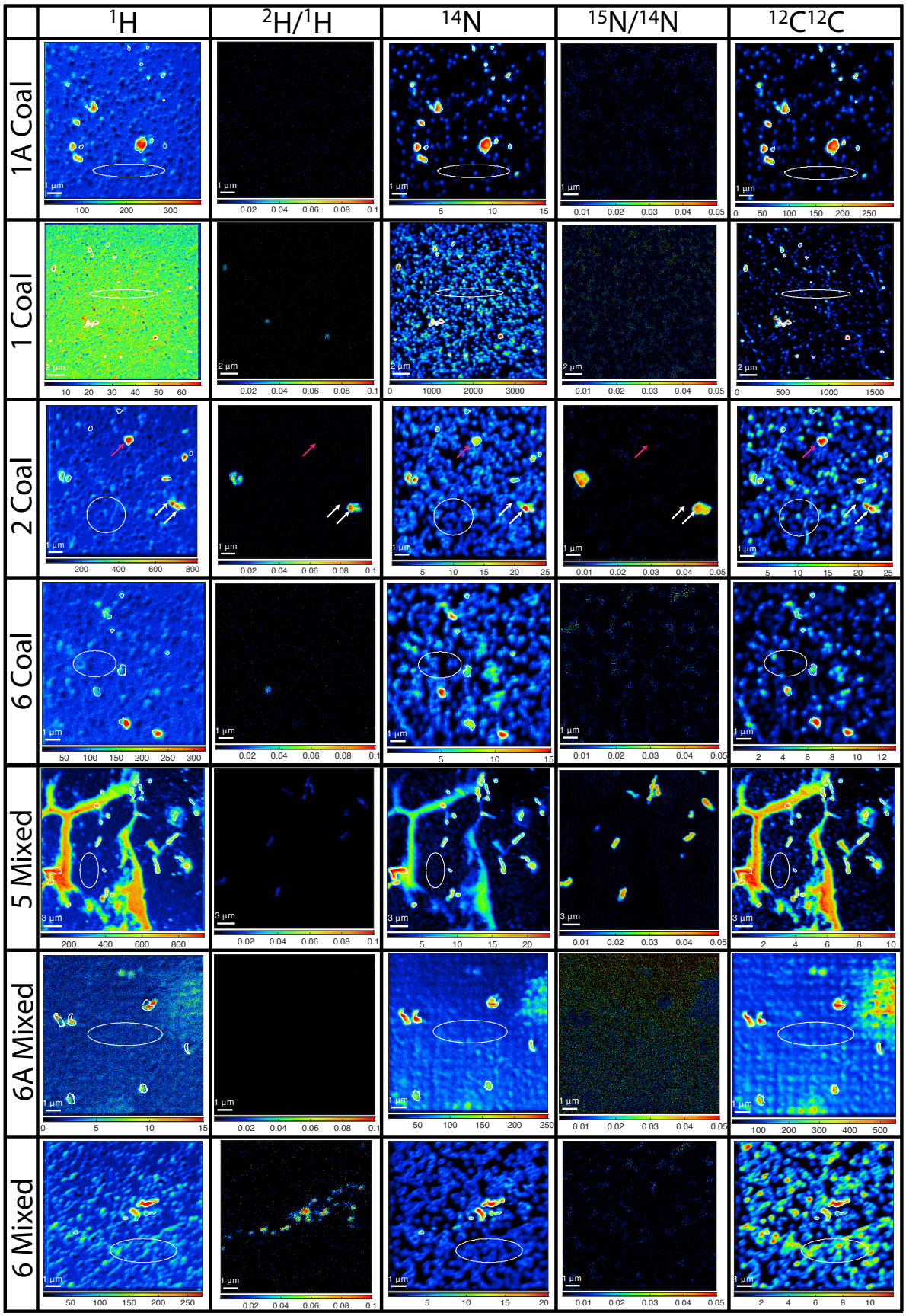
Table 2: Isotope mass balance calculations of bulk dissolved inorganic carbon (DIC) production (catabolism) from potential organic carbon sources available in incubations. *Estimated, all other sources from in situ or incubation measurements (Supplemental Table 1). Calculated fmol DIC produced per cell per day with standard deviation (~ 7 cm^3 inoculum per bottle). The error in DIC production is from propagation of μmol DIC measurement.

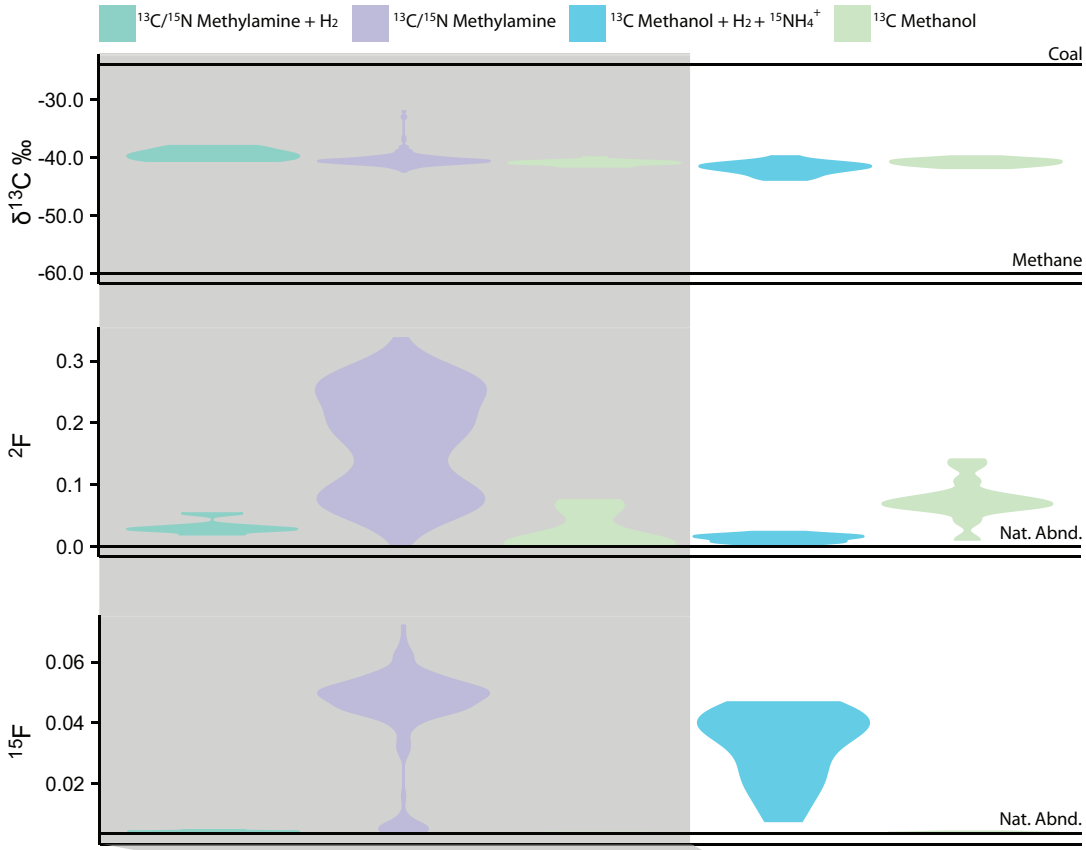
Lith.	Amendment	DIC prod (pmol C $cell^{-1}$ d^{-1})	σ DIC prod (pmol C $cell^{-1}$ d^{-1})	fmol C/cell/day catabolized from substrate mixed with:		
				Methane	****Coal Derv.***	***Coal***
Coal	MeAm.	0.01	0.001	0.015	0.006	0.005
Mixed	MeOH+H_2+Am.	12.01	0.2			
Mixed	MeOH	0.88	0.01	2.1	1.5	1.3

Single-Cell Anabolic Rates

Representative NanoSIMS ion count (total ^1H, total $^{12}C^{14}N$, and total $^{12}C_2$) and ion ratio ($^2H/^1H$ and $^{15}N/^{14}N$) maps of each incubation showed biomass labeled isotope incorporation in both 15R3 coal (3 out of 8) and mixed lithology (2 out of 8; Figure 3) incubations. Ion ratio maps were not shown for carbon because ^{13}C enrichment above background was not visible. No ^2H or ^{15}N biomass incorporation above background was detected in any autoclaved samples (A), 15R3 coal amended with methanol + ammonium + hydrogen (5 Coal), or mixed lithology amended with methylamine ± hydrogen (1 Mixed and 2 Mixed; Figure 3). Autoclaved 15R3 coal amended with methylamine + hydrogen (1A Coal) and autoclaved mixed lithology amended with methanol (6A Mixed) are provided as examples of NanoSIMS membranes with positive cell identification, but no labeled isotope incorporation, for comparison with active samples. 15R3 coal amended with methylamine (2 Coal) also had evidence of putative germination or cell division (white arrows; Figure 3), where the region of high ^2H and ^{15}N enrichment indicates the daughter cell.

→Figure 3: NanoSIMS analysis of the ^1H, ^{14}N, and $^{12}C_2$ ion counts with $^2H/^1H$ and $^{15}N/^{14}N$ ratio images for 15R3 coal and mixed lithology samples with ^2H and ^{15}N enriched cells. 2 (methylamine) Coal showed evidence for active cell division or germination during the 2.5-year incubation (white arrows), where the parent cell with little ^2H or ^{15}N enrichment is the top arrow and enriched daughter cell is the bottom arrow. Only the daughter cell is evident in the ratio images. Pink arrow provides example of cell that shows no enrichment in the same image. Large ellipses represent background filter correction ROI (see Methods). Two autoclaved samples that have cells with no enrichment are also shown (1A Coal and 6A Mixed). A - autoclaved, Coal – 15R3 coal, Mixed – mixed lithology, 1 - $^{13}C/^{15}N$-methylamine + hydrogen, 2 - $^{13}C/^{15}N$-methylamine, 5 - ^{13}C-methanol + hydrogen +^{15}N-ammonium, 6 – ^{13}C-methanol.


^{1}H
$^{2}H/^{1}H$
^{14}N
$^{15}N/^{14}N$
$^{12}C^{12}C$
1A Coal
1 Coal
2 Coal
6 Coal
5 Mixed
6A Mixed
6 Mixed



	15R3 Coal			Mixed Lithology	
Number of ROIs	6	176	18	18	15
			Deuterium		
^2F min.	0.01955	0.00222	0.00141	0.00149	0.04199
^2F med.	0.02860	0.18448	0.00651	0.01415	0.07072
^2F max.	0.03332	0.33737	0.07517	0.02387	0.10638
			aw 0.83		
^2F min. Turn. (yr)	20	196	323	302	8
^2F med. Turn. (yr)	13	<1	63	28	4
^2F max. Turn. (yr)	11	<1	4	16	2
			aw 0.44		
^2F min Turn. (yr)	10	103	171	160	4
^2F med. Turn. (yr)	6	<1	33	14	2
^2F max. Turn. (yr)	5	<1	1	8	<1
			Nitrogen		
^{15}F min	0.00405	0.03348	0.00369	0.00738	0.00353
^{15}F median	0.00429	0.04856	0.00377	0.03700	0.00387
^{15}F max	0.00474	0.06280	0.00389	0.04714	0.00429
^{15}F min Turn. (yr)	3408	40	NA	14	NA
^{15}F min Turn. (yr)	2020	26	NA	<1	NA
^{15}F min Turn. (yr)	1153	19	NA	<1	NA
mol N/cell/day	1.15E-21	9.59E-20	NA	2.49E-19	NA
mol N/cell/day	1.94E-21	1.44E-19	NA	2.25E-18	NA
mol N/cell/day	3.39E-21	1.90E-19	NA	2.94E-18	NA

All NanoSIMS cellular ROIs with significant deuterium ion counts (see Methods) are summarized in violin plots (kernel density estimation) of $\delta^{13}C$, ^{2}F, and ^{15}F (Figure 4). Similar to the bulk incubation methane and DIC data, cellular ^{13}C biomass enrichment ranged from -41.85 ‰ to -31.99 ‰ and did not indicate significant labeled substrate assimilation. Using isotope mass balance calculations, we estimated the biosynthesis rate from ^{13}C substrate in conjunction with potential native carbon sources for a range of carbon isotope values that encompass NanoSIMS ROIs (^{13}F 0.01063 to 0.01076; Figure 5). The NanoSIMS values correspond to 0.5 to 4.5×10^{-6} fmol C from substrate per cell per day, i.e. zeptomole range.

While 15R3 coal amended with methylamine did have some cell ROIs (4 out of 233) that were slightly more enriched in ^{13}C, the majority of ROIs for this sample had ^{2}H and ^{15}N biomass enrichment. A total of 212 cellular 15R3 coal amended with methylamine ROIs were defined, 175 of which met the Poisson filter for deuterium ion counts (Figure 4). 99 of the 212 ROIs (47%) have a ^{2}F greater than ~ 0.001, corresponding to biosynthesis turnover time shorter than length of the incubation. The mean $^{2}F_{biomass}$ was ~ 1000 × natural abundance ($^{2}F_{nat}$ ~ 0.00015). The paired 15R3 coal amended with methylamine and hydrogen was ~ 200 × natural abundance. Methanol amendment had the second highest activity in 15R3 coal incubations, and the highest activity in the mixed lithology incubations. Mixed lithology amended with methanol and hydrogen + ammonium had the lowest activity, ~ 10 × natural abundance. The same relative activity between incubations remains without the Poisson error filter that was applied for cell ROIs in Figure 4 (see Supplemental Figure 4 with all ROIs). Some ROIs are above the 20 at. % D_2O label provided, but are considered within error of ≤ 20 at. % when accounting for the high label and long incubation variability. Kopf et al. (2015) showed that higher isotopic spikes and

← Figure 4: Violin plots displaying the kernel density distribution of (a) deuterium and (b) ^{15}N fractional isotope abundance from NanoSIMS ROIs that had D ion counts > 2 × shot noise. All cells are above natural abundance ^{2}F and ^{15}F, depicted as horizontal lines. 15R3 coal methylamine has a bimodal distribution in both ^{2}F and ^{15}F enrichment. The table includes number of ROIs in violin plots; the minimum, median, and maximum enrichment values for ^{2}F and turnover times in years for the range of ^{2}F values at three different water hydrogen assimilation constants (a_w); the minimum, median, and maximum enrichment values for ^{15}F and turnover times in years for the range of ^{15}F values. Methanol + hydrogen + ammonium had a ^{15}N isotope spike of 0.027 and methylamine samples had a ^{15}N isotope spike of 0.5. Gray shading differentiates 15R3 coal from mixed lithology samples.

longer cell retention times resulted in a larger spread of cellular anabolism values. For incubations with high levels of $^{2}H/^{1}H$ as used in this study, this amplifies the error in the conversion factor Kopf et al. calculated between bulk NanoSIMS and lipid-only deuterium (see Methods). Deuterium enrichment did not display a normal distribution across analyzed ROIs, but instead had a bimodal distribution as was also seen by Kopf et al. (2015). When comparing water assimilation within a chemostat monoculture grown with different turnover times (2 hrs versus 19 hrs), it was demonstrated that cells had higher water assimilation (~0.6 versus ~0.8) when turnover times were longer (Kopf 2015). Extending these findings to deep biosphere conditions, the longer turnover times based on the higher water assimilation constant-based (0.83) may be the most applicable to our system.

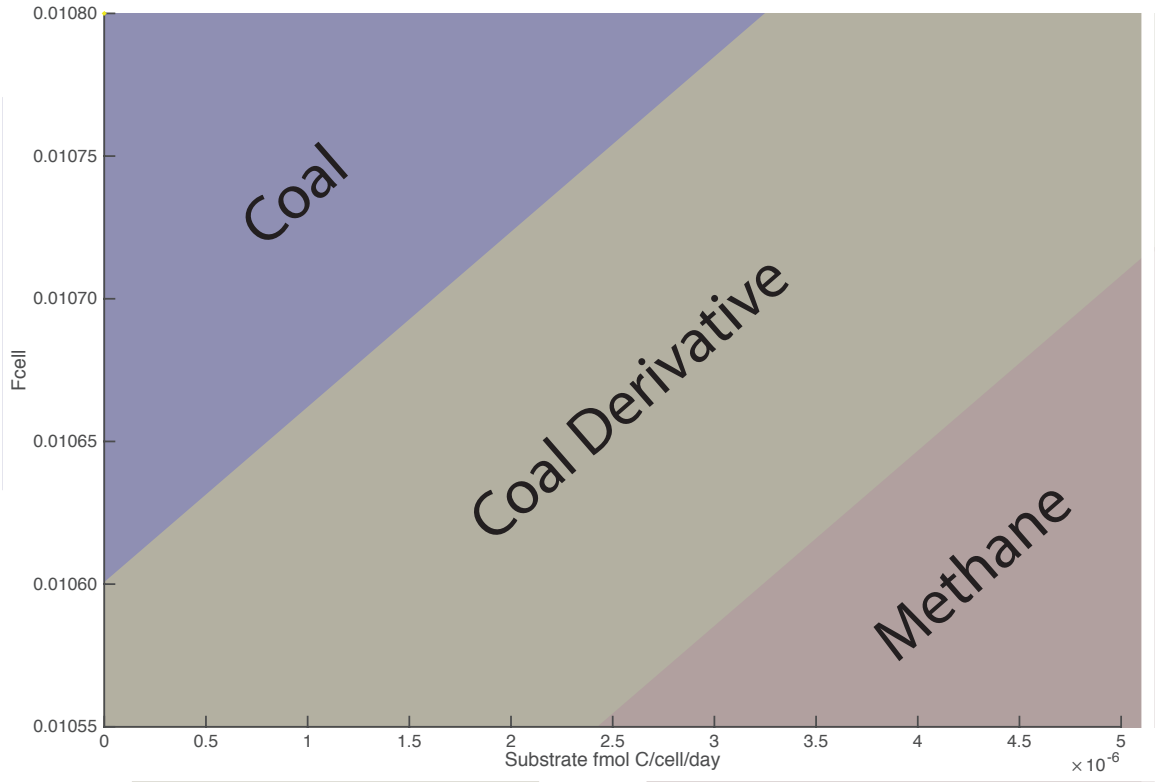


Figure 5: Carbon isotope mass balance to determine the range of substrate utilization rates (fmol C/cell/day) given the range of ^{13}C enrichment in cellular ROIs measured on the NanoSIMS (^{13}F 0.01063 to 0.01076) and the a range of native organic carbon substrates available and their in situ isotopic compositions (Supplemental Table 1).

^{15}N biomass incorporation mirrors ^{2}H for 15R3 coal amended with methylamine, with the highest anabolic activity observed in incubations without hydrogen amendment ($^{15}F_{biomass}$ is ~ 10

× natural abundance; $^{15}F_{nat}$ ~ 0.003) and lower, but detectable, anabolism when hydrogen was added. No ^{15}N activity was detected in mixed lithology amendments with methylamine ($^{15}F_{biomass}$ ~ natural abundance), but activity was detected with methanol + hydrogen + ammonium. No ^{15}N incorporation was seen in any autoclaved samples above the background threshold of natural abundance ^{15}N. There were no ROIs with ^{15}N incorporation without corresponding ^{2}H incorporation.

Single-cell biosynthesis rate-based turnover times were calculated based on ^{2}H and ^{15}N cellular enrichment (see Methods), where turnover times longer than 2.45 years are longer than the length of the incubation. Cells in 15R3 coal amended with methylamine had median deuterium-based turnover times of less than one year, regardless of a_w (Figure 4). 15R3 coal amended with methylamine + hydrogen had longer median turnover times of 6 to 13 years. 15R3 coal amended with methanol had the longest 15R3 coal turnover times of 33 to 63 years. Mixed lithology amended with methanol had a much shorter medium turnover time of 2 to 4 years, and even when amended with hydrogen (14 to 28 years) was still shorter than 15R3 coal with methanol only.

Nitrogen-based turnover times were slower than deuterium-based estimates for the methylamine (median 26 years) and methylamine + hydrogen (median 2,020 years) conditions, and faster than deuterium-based for ammonium condition (<1 year). Average anabolic rates were 1.99×10^{-6}, 1.45×10^{-4}, and 2.43×10^{-3} fmol N/cell/day for 15R3 coal amended with methylamine + hydrogen, 15R3 coal amended with methylamine, and mixed lithology amended with methanol + hydrogen + ammonium, respectively (Figure 4).

Discussion

Coal composition and phylotypes

Despite cell abundances that were substantially lower than expected from global depth trends and sediment surface concentrations, single-cell SIP-NanoSIMS successfully detected an active microbial assemblage in two terrestrial coalbeds buried for millions of years. Stable isotope incubations with the 15R3 coal and mixed lithology samples represent two coal types and

compare different methods of incubation preparation (intact coal chips versus homogenized mixed sample), labeled carbon sources (methylamine versus methanol), and labeled nitrogen sources (methylamine versus ammonium). This matrix of conditions were all amended with a passive anabolic activity tracer (D_2O) and incubated at in situ temperature (45 ^{o}C).

The 15R (upper) and 24/25R (lower) coals had different depositional histories, petrographic properties (Gross et al. 2015), and bacterial communities (Inagaki et al. 2015). The upper coal seams were predominantly coniferous in origin and deposited in a paralic environment (coastal, with marine and freshwater influence). The lower coal seams were dominantly angiosperm in origin and deposited in a limnic environment (freshwater). The 15R coal was also a less thermally altered lignite (dominant maceral humotelinite), whereas all other coal seams were in the transition between lignite and sub-bituminous coal (dominant maceral humodetrinite; Gross et al. 2015). Comparing the 16S rRNA gene phylogenic diversity of the 15R coal and 24/25R coalbeds, the shallower 15R coal had a higher relative abundance of Bacillales, whereas Gemmatimondales and Acidobacteria dominated the deeper coalbed. These coal samples, and the 16S rRNA genes recovered from the terrestrial samples as a whole, were dominated by heterotrophic bacteria. Heterotrophic bacteria, while able decompose coal to methyl-substrates, are not known to utilize methyl-substrates for catabolism under anaerobic conditions, such as this deeply buried coalbed. Anaerobic methyl metabolism is only known to archaea, which were not a significant proportion of the recovered 16S rRNA gene diversity. There was one archaeal 16S rRNA gene amplicon from a sample near the 25R coalbed related to *Methanosarcina* and sequences related to *Methanobacterium* were identified in methanogenic enrichments from 15R coal (Inagaki et al. 2015). Geochemical and biomarker evidence also points to an active methanogenic microbial coalbed community (Inagaki et al. 2015) that could be capable of our target methyl-metabolisms.

Carbon catabolism and anabolism

15R3 coal carbon catabolic rates ranged from 10^{-2} to 10^{-3} fmol C/cell/day from the provided ^{13}C-methyl-substrate, depending on which natural carbon source (methane, coal derivative, or coal) was mixed with the substrate. Comparing our carbon catabolic rates with those estimated from ^{35}S sulfate reduction (Hoehler & Jorgensen 2013), 15R3 coal incubation carbon catabolic rates

are between the lower end of marine surface sediment and higher end of deep subsurface marine sediment rates, and mixed lithology incubation carbon catabolism rates were closer to pure cultures. This difference between coal and mixed lithology incubations could be due to higher accessibility of fresh surfaces in the macerated mixed sample, relative to the intact pieces in 15R3 coal incubations, and/or a wider range of non-carbon substrates (e.g. iron minerals in sandstones and shales) from the mixing of multiple lithologies. Using NanoSIMS single-cell carbon enrichment and isotopic mass balance calculations, carbon anabolism (~ 10^{-6} fmol C/cell/day from substrate) was much lower than catabolism, and lower than any previously published deep biosphere anabolic rates. This is consistent with theories that deep biosphere cells have minimal, maintenance-only turnover of cellular biomass for energy conservation (Jørgensen & Marshall 2016), potentially aided by biomolecule recycling (Takano et al. 2010).

Considering the high percentage of label added (50 at. %), the DIC, CH_4, and single-cell NanoSIMS data indicate the methyl substrates added to the incubation were not the predominant carbon sources used to build the cellular biomass detected by ^{2}H and ^{15}N enrichment in our incubations. Instead, single-cell NanoSIMS and bulk geochemistry results suggest use of indigenous organic carbon substrates as opposed to in situ methane (-60 ‰) or DIC (-10 to 5 ‰). These findings are consistent with results from ODP Leg 201 Peru Margin where single-cell natural abundance carbon isotope measurements indicated a predominantly fossil organic carbon source (Biddle et al. 2008). However, NanoSIMS ^{13}C biomass measurements (-38 ‰) are about 15 ‰ more depleted than the coal (-25 ‰). One possibly is that microbes are accessing more depleted coal derivatives, such as the methoxyl group in vanillin. In NMR analysis of vanillin derived from microbial fermentation of lignin, the methoxyl group was 10 ‰ more depleted than the bulk vanillin molecule (Tenailleau et al. 2004). Alternatively, pressure dependent stable isotope fractionations have been observed for bulk biomass and fatty acids in heterotrophic bacteria (ΔδFA-glucose 15-18‰; Fang et al. 2006) that could allow for coalbed cells to be utilizing organic carbon sources with coal-carbon isotopic abundance (-25 ‰) while producing biomass that is -38 ‰, as seen here. While our incubations were not maintained at the in situ pressures that would produce these effects, there was little evidence of carbon anabolism from labeled substrates or significant carbon metabolic rates from any substrates, that would have significant altered the biomass in our incubations from its in situ composition. Therefore the

NanoSIMS is likely measuring the isotopic enrichment that would have been produced in situ and could explain the coal-biomass offset observed here.

Non-methyl carbon heterotrophy is further supported by results from the JAMSTEC incubations with a wider range of multi-carbon substrates. Out of all 15R3 incubations prepared onboard, ^{13}C-glucose had the highest enrichment in ^{13}C-DIC (~ 1,000 ‰; data not shown). Incubations with more coalbed relevant heterotrophic substrates, such as lignin monomer acids, also had labeled DIC production (~ 800 ‰ p-Coumaric acid ^{13}C (1, 2, 3); ~ 200 ‰ ferulic acid ^{13}C (1, 2, 3); data not shown) that was significantly higher than our methyl substrates (20 ‰; Table 1). This suggests heterotrophic coal degradation to methanogenesis substrates may occur at a faster rate than methylotrophic methanogenesis in this system. High in situ hydrogen concentrations (1 to 500 μM) indicate hydrogenotrophic methanogenesis is also operating far below the thermodynamic maximum (Inagaki et al. 2015), and multi-carbon organic substrates are more likely what supports deep life in the Shimokita coalbeds.

Hydrogen biosynthesis

Based on background 15R cell abundances of ~ 5 cells/cm^3 (~ 1920 mbsf), and similarly low cell abundances in samples that had no other signs of microbial activity, microorganisms in the 15R3 coal amended with methylamine had some growth during the 894 day incubation period (final cell concentration of 1920 cells/cm^3). This increase in cell abundance is supported by NanoSIMS data showing the cellular enrichment in both ^{2}H/^{1}H and ^{15}N/^{14}N, and the greatest number of isotopically enriched cells (Figure 4). Estimates from final incubation cell counts and NanoSIMS average cellular ^{2}F both independently calculated biosynthesis rate-based turnover times of less than one year for this incubation. 15R3 amended with methylamine also had a contiguous biomass doublet in NanoSIMS ^{1}H, ^{14}N, and ^{12}C ion counts (2 Coal; Figure 3), where only half of the sample had enrichment in ^{2}H (4.4 at. %) and ^{15}N (3.5 at. %) suggesting biomass existing prior to isotopic label exposure produced additional biomass after onset of incubation. This could be evidence of a dividing or sporulating cell. While most ROIs were active in this incubation, only about half of the cells were fully labeled (^{2}F 0.2). This could be the result of a microbial assemblage split into two modes, one that is more active and dividing and another that is undergoing maintenance-only biosynthesis. Kopf et al. (2015) also saw label uptake

heterogeneity became more pronounced as generation time increased (slower rates of growth), even in a monoculture chemostat. Therefore, this interesting physiological phenomenon does not appear limited to deep biosphere environments and does not require substrate partitioning (i.e. autotrophic versus heterotrophic populations) to occur, and should be explored further in both laboratory and environmental conditions.

Hydrogen inhibition

In cases where samples could be compared with and without hydrogen, hydrogen addition appeared to depress both ^{2}H and ^{15}N biosynthesis. Since hydrogen is generally produced by fermenters and removed by methanogens in coalbed methane systems (Strąpoć et al. 2011), it is likely the fermentive microbes that are inhibited by the addition of hydrogen if it is not kept under thermodynamic control by methanogens. Given that the hydrogen in our incubations should also be labeled from hydrogen exchange with deuterated water (Campbell et al. 2009), hydrogen derived methane produced in our system should have detectable $\delta D\text{-}CH_4$. The $\delta D\text{-}CH_4$ in 15R3 coal incubations did not indicate production via labeled hydrogen, and suggests hydrogenotrophic methanogens were not removing hydrogen from the incubations. Since we see little evidence of hydrogenotrophic methanogenesis, or other metabolisms that could remove hydrogen, our incubation hydrogen concentrations could inhibit fermenters. These findings demonstrating hydrogen as an inhibitor in our coalbed system are in contrast to the canonical idea that hydrogen is an important energy source for the deep biosphere (Adhikari et al. 2016; Pedersen 2000).

Nitrogen biosynthesis

There appeared to be a distinction in the nitrogen metabolism between the two sample types, 15R3 coal and mixed lithology. Biosynthesis turnover times based on 15R3 coal amended with ^{15}N-methylamine ranged from 19 to 40 years (Figure 4). While ^{2}F and ^{15}F were both highest in 15R3 coal amended with methylamine, N-based turnover times were fastest with ammonium amended incubations from mixed lithology (methanol + hydrogen + ammonium incubations; <1 to 3 years). This is due to turnover time calculations accounting for label dilution from background ammonium concentrations, and demonstrates the highest absolute rate ^{15}N assimilation was with ammonium. Surprisingly, no anabolism of ^{15}N was detected in 15R3 coal

incubations amended with ^{13}C-methanol, hydrogen, and ^{15}N-ammonium or mixed lithology samples amended with ^{15}N-methylamine ± hydrogen (Figure 4). Methylated amine concentrations are higher in estuarine environments than freshwater, and appear to sorb to the sediment fraction more than the aqueous fraction (Zhuang 2014). While unlikely that methylamine would survive deposition for millions of years in the coal, it is possible that a community buried under conditions where it was available might contain a microbial assemblage more predisposed to its utilization when provided as a substrate in our incubations. *Bacillus* sp. have also been shown to use methylamine as their sole nitrogen source (Bicknell & Owens 1980). Their higher relative abundance in 15R core 16S rRNA diversity screens is consistent with these nitrogen assimilation results suggesting fermentive bacteria may be anabolizing the nitrogen from methylamine in 15R3 coal incubations.

N versus C biosynthesis

NanoSIMS-based carbon anabolism rates were estimated to be ~10^{-6} fmol C/cell/day for the range of potential organic carbon sources used with the labeled substrate provided, across the full range of cellular ROI carbon isotope values from all incubations. 15R3 ^{15}N-methylamine nitrogen anabolism rates differed by two orders of magnitude for with and without hydrogen (10^{-6} fmol N/cell/day and 10^{-4} fmol N/cell/day, respectively), where the + hydrogen condition had the same anabolism rate as the carbon-based estimate for all incubations (10^{-6}). Methylamine utilization as a sole nitrogen source has been shown to be widespread among non-methylotrophic and non-methanogenic bacteria, even under anaerobic conditions, with apparent disregard of the carbon moiety of the molecule (Bicknell & Owens 1980). This suggests that only one population (i.e. methylotrophs) may be assimilating methylamine-derived carbon, but multiple community members may be capable of utilizing the nitrogen from methylamine (i.e. fermenters and methylotrophs). These results are consistent with hydrogen amendment not affecting the methylotrophic methanogen population present in both methylamine-amended 15R3 coal incubations, but hindering the dominant fermenting population utilizing methylamine for nitrogen but not carbon. Estimated DIC production rates from methylamine (catabolism) are all higher than nitrogen assimilation rates, such that the ^{13}C-labeled methyl group that was not anabolized from methylamine could have been catabolized and transferred to the DIC pool.

Mixed lithology amended with methanol + hydrogen + ammonium had the highest average nitrogen assimilation rate (2.43×10^{-3} fmol N/cell/day from ammonium). This ^{15}N-ammonium assimilation rate is 13-20% of the single-cell rates reported from NanoSIMS analysis of ^{13}C-bicarbonate and ^{13}C-acetate amendments with ^{15}N-ammonium from shallower (219 mbsf) sediment at this site (Morono et al. 2011). However, our carbon anabolic rates (~ 10^{-6} fmol C/cell/day) differ by three orders of magnitude from our ammonium assimilation rates, and four orders of magnitude from Morono et al. (~ 10^{-2} fmol C/cell/day). These discrepancies could be explained methodologically since the Morono et al. (2011) incubation amendments were different substrates and provided at higher concentrations, or could be a function of the difference in microbial community 2 km deeper into the Shimokita coalbed system. Despite the absolute rate differences, our findings are in agreement with Morono et al. (2011) that nitrogen, in the form of ammonium, is assimilated at a much higher rate than carbon in deep biosphere systems.

Conclusion

We provide the first deep biosphere application of a universal, passive tracer with extremely high sensitivity (D_2O) and single-cell resolution (NanoSIMS), as compared to previous methods based on bulk projections of sulfate reduction (D'Hondt et al. 2002; Parkes et al. 1990), single-cell natural abundance carbon (Biddle et al. 2006), or D:L racemization (Lomstein et al. 2012). SIP-NanoSIMS-based rates (months to years) appear much faster than these other methods (thousands of years), which may point to the necessity of single-cell techniques to determine accurate rates in low activity systems. Despite extremely low cell abundances, our incubations with deeply buried terrestrial coal and associated lithologies were successful in determining average deuterium-based turnover times ranging from less than a year to 63 years after an extended incubation time of 2.5 years at in situ temperatures. The bimodal distribution of single-cell deuterium enrichment for the most active incubation (15R3 coal amended with methylamine) supports the idea of a deep coalbed microbial assemblage that is capable of activation and growth when resources become available, while simultaneously sustaining a community dominated by maintenance over growth. We find that nitrogen was assimilated from both methylamine and ammonium, and the pattern of methylamine utilization was consistent

with the coal depositional setting and microbial phylotypes. We also support previous findings that nitrogen from ammonium is assimilated more readily than carbon in deep biosphere SIP-NanoSIMS incubations and that most deep biosphere cells appear to be viable (Morono et al. 2011). Methyl-substrate specific carbon anabolic rates were much slower than any previously published carbon rates (~ 10^{-6} fmol C/cell/day) where the main carbon source anabolized appeared to be in situ fossil organic carbon, as has been seen by others (Biddle et al. 2006), in a system where bacterial coal degradation may dominate archaeal methanogenesis.

References

Adhikari RR, Glombitza C, Nickel JC, Anderson CH, Dunlea AG, Spivack AJ, Murray RW, D'Hondt S, and Kallmeyer J. 2016. Hydrogen utilization potential in subsurface sediments. *Frontiers in Microbiology* 7. 10.3389/fmicb.2016.00008

Auguie B. 2012. gridExtra: functions in Grid graphics. R package version 0.9.1 ed.

Berry D, Mader E, Lee TK, Woebken D, Wang Y, Zhu D, Palatinszky M, Schintlmeister A, Schmid MC, Hanson BT, Shterzer N, Mizrahi I, Rauch I, Decker T, Bocklitz T, Popp Jr, Gibson CM, Fowler PW, Huang WE, and Wagner M. 2015. Tracking heavy water (D2O) incorporation for identifying and sorting active microbial cells. *Proceedings of the National Academy of Sciences* 112:E194-E203. 10.1073/pnas.1420406112

Bicknell B, and Owens J. 1980. Utilization of methyl amines as nitrogen sources by non-methylotrophs. *Microbiology* 117:89-96.

Biddle JF, Fitz-Gibbon S, Schuster SC, Brenchley JE, and House CH. 2008. Metagenomic signatures of the Peru Margin subseafloor biosphere show a genetically distinct environment. *Proceedings of the National Academy of Sciences* 105:10583-10588. 10.1073/pnas.0709942105

Biddle JF, Lipp JS, Lever MA, Lloyd KG, S√ ∏rensen KB, Anderson R, Fredricks HF, Elvert M, Kelly TJ, Schrag DP, Sogin ML, Brenchley JE, Teske A, House CH, and Hinrichs K-U. 2006. Heterotrophic Archaea dominate sedimentary subsurface ecosystems off Peru. *Proceedings of the National Academy of Sciences of the United States of America* 103:3846-3851. 10.1073/pnas.0600035103

Biddle JF, Sylvan JB, Brazelton WJ, Tully BJ, Edwards KJ, Moyer CL, Heidelberg JF, and Nelson WC. 2011. Prospects for the Study of Evolution in the Deep Biosphere. *Frontiers in Microbiology* 2:285. 10.3389/fmicb.2011.00285

Campbell BJ, Li C, Sessions AL, and Valentine DL. 2009. Hydrogen isotopic fractionation in lipid biosynthesis by H2-consuming Desulfobacterium autotrophicum. *Geochimica et Cosmochimica Acta* 73:2744-2757. http://dx.doi.org/10.1016/j.gca.2009.02.034

Ciobanu M-C, Burgaud G, Dufresne A, Breuker A, Redou V, Ben Maamar S, Gaboyer F, Vandenabeele-Trambouze O, Lipp JS, Schippers A, Vandenkoornhuyse P, Barbier G, Jebbar M, Godfroy A, and Alain K. 2014. Microorganisms persist at record depths in the subseafloor of the Canterbury Basin. *ISME J* 8:1370-1380. 10.1038/ismej.2013.250

Coplen TB, Brand WA, Gehre M, Gröning M, Meijer HAJ, Toman B, and Verkouteren RM. 2006. New Guidelines for δ13C Measurements. *Analytical Chemistry* 78:2439-2441. 10.1021/ac052027c

D'Hondt S, Inagaki F, Zarikian CA, Abrams LJ, Dubois N, Engelhardt T, Evans H, Ferdelman T, Gribsholt B, Harris RN, Hoppie BW, Hyun J-H, Kallmeyer J, Kim J, Lynch JE, McKinley CC, Mitsunobu S, Morono Y, Murray RW, Pockalny R, Sauvage J, Shimono T, Shiraishi F, Smith DC, Smith-Duque CE, Spivack AJ, Steinsbu BO, Suzuki Y, Szpak M, Toffin L, Uramoto G, Yamaguchi YT, Zhang G-l, Zhang X-H, and Ziebis W. 2015. Presence of oxygen and aerobic communities from sea floor to basement in deep-sea sediments. *Nature Geosci* 8:299-304. 10.1038/ngeo2387

D'Hondt S, Jørgensen BB, Miller DJ, Batzke A, Blake R, Cragg BA, Cypionka H, Dickens GR, Ferdelman T, Hinrichs K-U, Holm NG, Mitterer R, Spivack A, Wang G, Bekins B, Engelen B, Ford K, Gettemy G, Rutherford SD, Sass H, Skilbeck CG, Aiello IW, Guèrin G, House CH, Inagaki F, Meister P, Naehr T, Niitsuma S, Parkes RJ, Schippers A, Smith DC, Teske A, Wiegel J, Padilla CN, and Acosta JLS. 2004. Distributions of Microbial Activities in Deep Subseafloor Sediments. *Science* 306:2216-2221. 10.1126/science.1101155

D'Hondt S, Rutherford S, and Spivack AJ. 2002. Metabolic Activity of Subsurface Life in Deep-Sea Sediments. *Science* 295:2067-2070. 10.1126/science.1064878

Eichorst SA, Strasser F, Woyke T, Schintlmeister A, Wagner M, and Woebken D. 2015. Advancements in the application of NanoSIMS and Raman microspectroscopy to investigate the activity of microbial cells in soils. *FEMS Microbiology Ecology* 91. 10.1093/femsec/fiv106

Expedition 337 Scientists. 2013. Methods. *Proceedings of IODP* 337. 10.2204/iodp.proc.337.102.2013

Fang J, Uhle M, Billmark K, Bartlett DH, and Kato C. 2006. Fractionation of carbon isotopes in biosynthesis of fatty acids by a piezophilic bacterium Moritella japonica strain DSK1. *Geochimica et Cosmochimica Acta* 70:1753-1760. http://dx.doi.org/10.1016/j.gca.2005.12.011

Fry JC, Horsfield B, Sykes R, Cragg BA, Heywood C, Kim GT, Mangelsdorf K, Mildenhall DC, Rinna J, Vieth A, Zink K-G, Sass H, Weightman AJ, and Parkes RJ. 2009. Prokaryotic Populations and Activities in an Interbedded Coal Deposit, Including a Previously Deeply Buried Section (1.6,Äì2.3 km) Above ‚àº 150 Ma Basement Rock. *Geomicrobiology Journal* 26:163-178. 10.1080/01490450902724832

Gross D, Bechtel A, and Harrington GJ. 2015. Variability in coal facies as reflected by organic petrological and geochemical data in Cenozoic coal beds offshore Shimokita (Japan) - IODP Exp. 337. *International Journal of Coal Geology* 152, Part B:63-79. http://dx.doi.org/10.1016/j.coal.2015.10.007

Hayes JM. 2004. An introduction to isotopic calculations. *Woods Hole Oceanographic Institution, Woods Hole, MA* 2543.

Hoehler TM, and Jorgensen BB. 2013. Microbial life under extreme energy limitation. *Nat Rev Micro* 11:83-94.

Horsfield B, Schenk HJ, Zink K, Ondrak R, Dieckmann V, Kallmeyer J, Mangelsdorf K, di Primio R, Wilkes H, Parkes RJ, Fry J, and Cragg B. 2006. Living microbial ecosystems within the active zone of catagenesis: Implications for feeding the deep biosphere. *Earth and Planetary Science Letters* 246:55-69. http://dx.doi.org/10.1016/j.epsl.2006.03.040

Inagaki F, Hinrichs K-U, Kubo Y, Bowles MW, Heuer VB, Hong W-L, Hoshino T, Ijiri A, Imachi H, Ito M, Kaneko M, Lever MA, Lin Y-S, Methé BA, Morita S, Morono Y, Tanikawa W, Bihan M, Bowden SA, Elvert M, Glombitza C, Gross D, Harrington GJ, Hori T, Li K, Limmer D, Liu C-H, Murayama M, Ohkouchi N, Ono S, Park Y-S, Phillips SC, Prieto-Mollar X, Purkey M, Riedinger N, Sanada Y, Sauvage J, Snyder G, Susilawati R, Takano Y, Tasumi E, Terada T, Tomaru H, Trembath-Reichert E, Wang DT, and Yamada Y. 2015. Exploring deep microbial life in coal-bearing sediment down to ~2.5 km below the ocean floor. *Science* 349:420-424. 10.1126/science.aaa6882

Inagaki F, Hinrichs K-U, and Kubo Ys. 2010. *Integrated Ocean Drilling Program Expedition 337 Scientific Prospectus*: Integrated Ocean Drilling Program Management International, Inc.

Inagaki F, Hinrichs K-U, Kubo Ys, and Scientists E. 2012. *Deep coalbed biosphere off Shimokita*: Integrated Ocean Drilling Program Management International, Inc.

Jørgensen BB, and Marshall IPG. 2016. Slow Microbial Life in the Seabed. *Annual Review of Marine Science* 8:311-332. doi:10.1146/annurev-marine-010814-015535

Kallmeyer J, Pockalny R, Adhikari RR, Smith DC, and D,ÄôHondt S. 2012. Global distribution of microbial abundance and biomass in subseafloor sediment. *Proceedings of the National Academy of Sciences* 109:16213-16216. 10.1073/pnas.1203849109

Kellermann MY, Yoshinaga MY, Wegener G, Krukenberg V, and Hinrichs K-U. 2016. Tracing the production and fate of individual archaeal intact polar lipids using stable isotope probing. *Organic Geochemistry* 95:13-20. http://dx.doi.org/10.1016/j.orggeochem.2016.02.004

Kikuchi S, Makita H, Konno U, Shiraishi F, Ijiri A, Takai K, Maeda M, and Takahashi Y. 2016. Limited reduction of ferrihydrite encrusted by goethite in freshwater sediment. *Geobiology*:n/a-n/a. 10.1111/gbi.12181

Konyukhov AI. 2010. Continental margins as global oil and gas accumulation belts: The West Pacific belt. *Lithology and Mineral Resources* 45:465-485. 10.1134/s0024490210050056

Kopf SH. 2015. From lakes to lungs: assessing microbial activity in diverse environments Doctor of Philosophy. California Institute of Technology.

Kopf SH, McGlynn SE, Green-Saxena A, Guan Y, Newman DK, and Orphan VJ. 2015. Heavy water and 15N labelling with NanoSIMS analysis reveals growth rate-dependent metabolic heterogeneity in chemostats. *Environmental Microbiology* 17:2542-2556. 10.1111/1462-2920.12752

Kopf SH, Sessions AL, Cowley ES, Reyes C, Van Sambeek L, Hu Y, Orphan VJ, Kato R, and Newman DK. 2016. Trace incorporation of heavy water reveals slow and heterogeneous pathogen growth rates in cystic fibrosis sputum. *Proceedings of the National Academy of Sciences* 113:E110-E116. 10.1073/pnas.1512057112

Leloup J, Fossing H, Kohls K, Holmkvist L, Borowski C, and Jørgensen BB. 2009. Sulfate-reducing bacteria in marine sediment (Aarhus Bay, Denmark): abundance and diversity related to geochemical zonation. *Environmental Microbiology* 11:1278-1291. 10.1111/j.1462-2920.2008.01855.x

Lipp JS, Morono Y, Inagaki F, and Hinrichs K-U. 2008. Significant contribution of Archaea to extant biomass in marine subsurface sediments. *Nature* 454:991-994. http://www.nature.com/nature/journal/v454/n7207/suppinfo/nature07174_S1.html

Lomstein BA, Langerhuus AT, D'Hondt S, Jorgensen BB, and Spivack AJ. 2012. Endospore abundance, microbial growth and necromass turnover in deep sub-seafloor sediment. *Nature* 484:101-104.

Morono Y, Terada T, Kallmeyer J, and Inagaki F. 2013. An improved cell separation technique for marine subsurface sediments: applications for high-throughput analysis using flow cytometry and cell sorting. *Environmental Microbiology* 15:2841-2849. 10.1111/1462-2920.12153

Morono Y, Terada T, Nishizawa M, Ito M, Hillion F, Takahata N, Sano Y, and Inagaki F. 2011. Carbon and nitrogen assimilation in deep subseafloor microbial cells. *Proceedings of the National Academy of Sciences* 108:18295-18300. 10.1073/pnas.1107763108

Morono Y, Terada T, Yamamoto Y, Xiao N, Hirose T, Sugeno M, Ohwada N, and Inagaki F. 2015. Intact preservation of environmental samples by freezing under an alternating magnetic field. *Environmental Microbiology Reports* 7:243-251. 10.1111/1758-2229.12238

Neuwirth E. 2014. RColorBrewer. R package version 1.1-2 ed.

Parkes RJ, Cragg B, Roussel E, Webster G, Weightman A, and Sass H. 2014. A review of prokaryotic populations and processes in sub-seafloor sediments, including biosphere:geosphere interactions. *Marine Geology* 352:409-425. http://dx.doi.org/10.1016/j.margeo.2014.02.009

Parkes RJ, Cragg BA, Fry JC, Herbert RA, Wimpenny JWT, Allen JA, and Whitfield M. 1990. Bacterial Biomass and Activity in Deep Sediment Layers from the Peru Margin [and Discussion]. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 331:139-153. 10.1098/rsta.1990.0061

Parkes RJ, Wellsbury P, Mather ID, Cobb SJ, Cragg BA, Hornibrook ERC, and Horsfield B. 2007. Temperature activation of organic matter and minerals during burial has the potential to

sustain the deep biosphere over geological timescales. *Organic Geochemistry* 38:845-852. http://dx.doi.org/10.1016/j.orggeochem.2006.12.011

Pedersen K. 2000. Exploration of deep intraterrestrial microbial life: current perspectives. *FEMS Microbiology Letters* 185:9-16. 10.1111/j.1574-6968.2000.tb09033.x

Polerecky L, Adam B, Milucka J, Musat N, Vagner T, and Kuypers MMM. 2012. Look@NanoSIMS – a tool for the analysis of nanoSIMS data in environmental microbiology. *Environmental Microbiology* 14:1009-1023. 10.1111/j.1462-2920.2011.02681.x

Postgate JR, and Hunter JR. 1963. Acceleration of Bacterial Death by Growth Substrates. *Nature* 198:273-273.

R Core Team. 2015. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

Rebata-Landa V, and Santamarina JC. 2006. Mechanical limits to microbial activity in deep sediments. *Geochemistry, Geophysics, Geosystems* 7:n/a-n/a. 10.1029/2006GC001355

Strąpoć D, Mastalerz M, Dawson K, Macalady J, Callaghan AV, Wawrik B, Turich C, and Ashby M. 2011. Biogeochemistry of microbial coal-bed methane.

Takano Y, Chikaraishi Y, Ogawa NO, Nomaki H, Morono Y, Inagaki F, Kitazato H, Hinrichs K-U, and Ohkouchi N. 2010. Sedimentary membrane lipids recycled by deep-sea benthic archaea. *Nature Geosci* 3:858-861.

Tenailleau EJ, Lancelin P, Robins RJ, and Akoka S. 2004. Authentication of the origin of vanillin using quantitative natural abundance 13C NMR. *Journal of agricultural and food chemistry* 52:7782-7787.

Toki T, Higa R, Ijiri A, Tsunogai U, and Ashi J. 2014. Origin and transport of pore fluids in the Nankai accretionary prism inferred from chemical and isotopic compositions of pore water at cold seep sites off Kumano. *Earth, Planets and Space* 66:1-14.

Torres ME, Mix AC, and Rugh WD. 2005. Precise δ13C analysis of dissolved inorganic carbon in natural waters using automated headspace sampling and continuous-flow mass spectrometry. *Limnology and Oceanography: Methods* 3:349-360. 10.4319/lom.2005.3.349

Wegener G, Bausch M, Holler T, Thang NM, Prieto Mollar X, Kellermann MY, Hinrichs K-U, and Boetius A. 2012. Assessing sub-seafloor microbial activity by combined stable isotope probing with deuterated water and 13C-bicarbonate. *Environmental Microbiology* 14:1517-1527. 10.1111/j.1462-2920.2012.02739.x

Whitman WB, Coleman DC, and Wiebe WJ. 1998. Prokaryotes: The unseen majority. *Proceedings of the National Academy of Sciences* 95:6578-6583.

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

Wickham H, and Francois R. 2015. dplyr: A Grammar of Data Manipulation. R package version 0.4.1 ed.

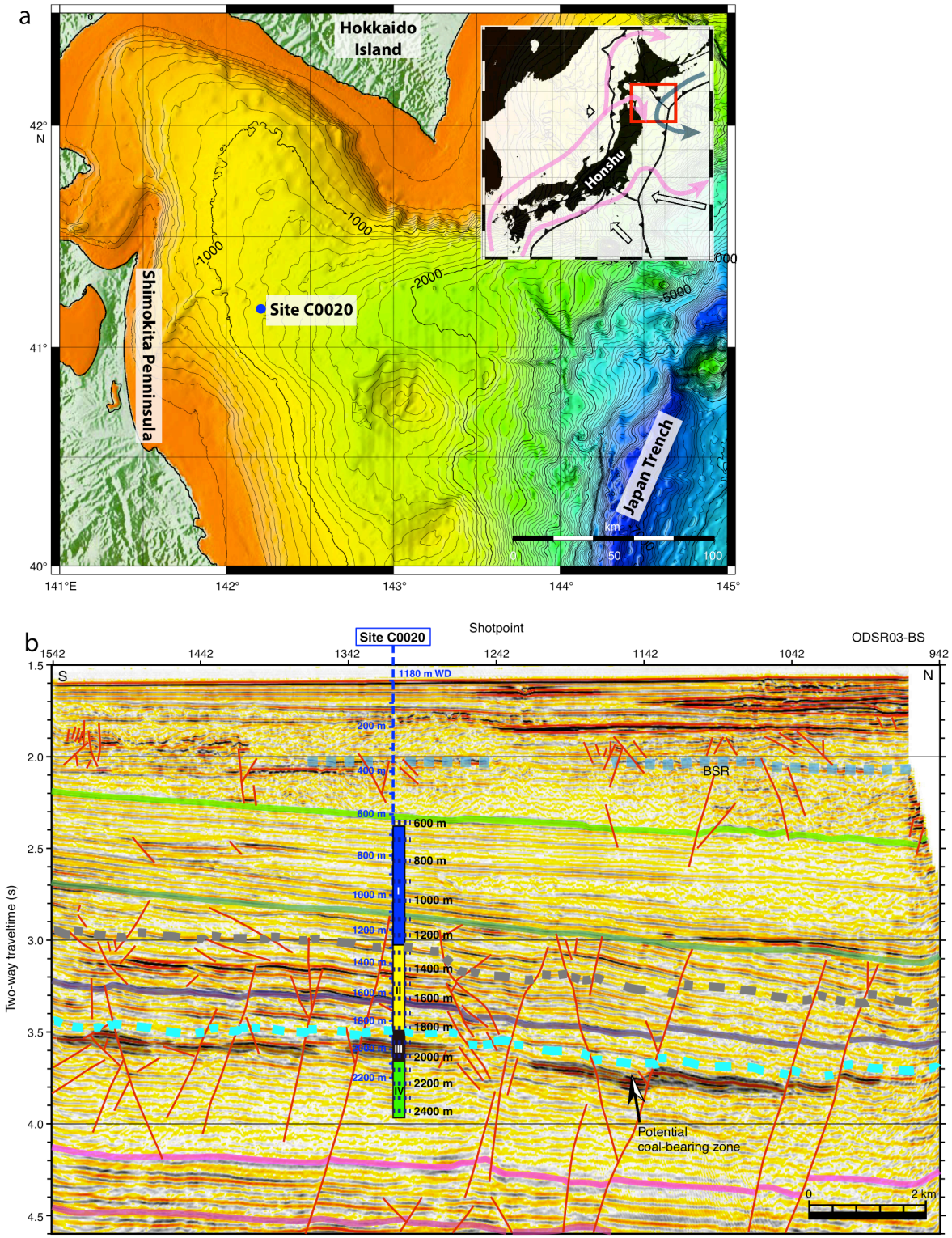
Xie S, Lipp JS, Wegener G, Ferdelman TG, and Hinrichs K-U. 2013. Turnover of microbial lipids in the deep biosphere and growth of benthic archaeal populations. *Proceedings of the National Academy of Sciences* 110:6010-6014. 10.1073/pnas.1218569110

Zeng X, Birrien J-L, Fouquet Y, Cherkashov G, Jebbar M, Querellou J, Oger P, Cambon-Bonavita M-A, Xiao X, and Prieur D. 2009. Pyrococcus CH1, an obligate piezophilic hyperthermophile: extending the upper pressure-temperature limits for life. *ISME J* 3:873-876.

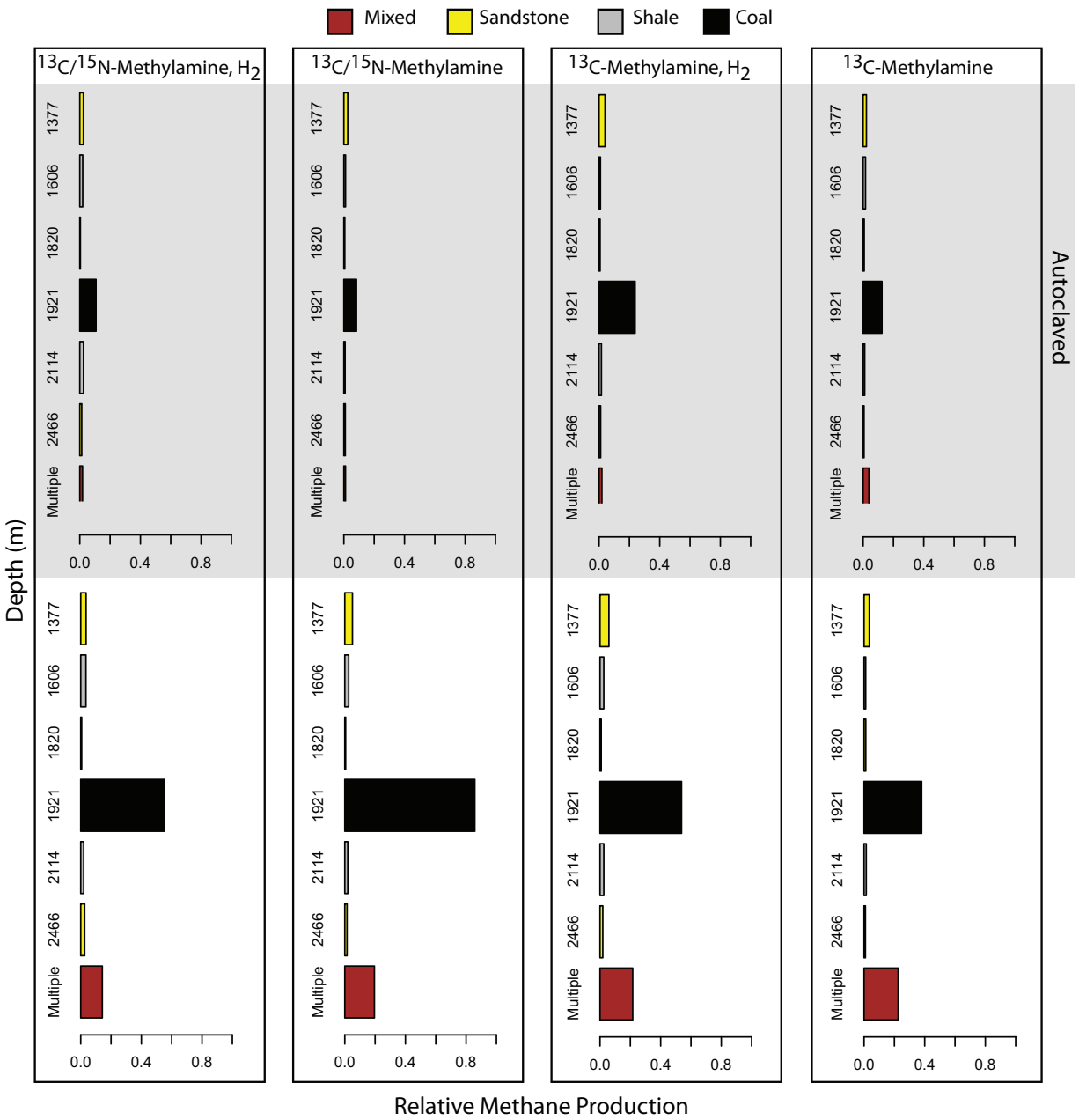
Zhang X, Gillespie AL, and Sessions AL. 2009. Large D/H variations in bacterial lipids reflect central metabolic pathways. *Proceedings of the National Academy of Sciences* 106:12580-12586. 10.1073/pnas.0903030106

Zhuang G. 2014. Methylotrophic methanogenesis and potential methylated substrates in marine sediment Doctoral. Universität Bremen.

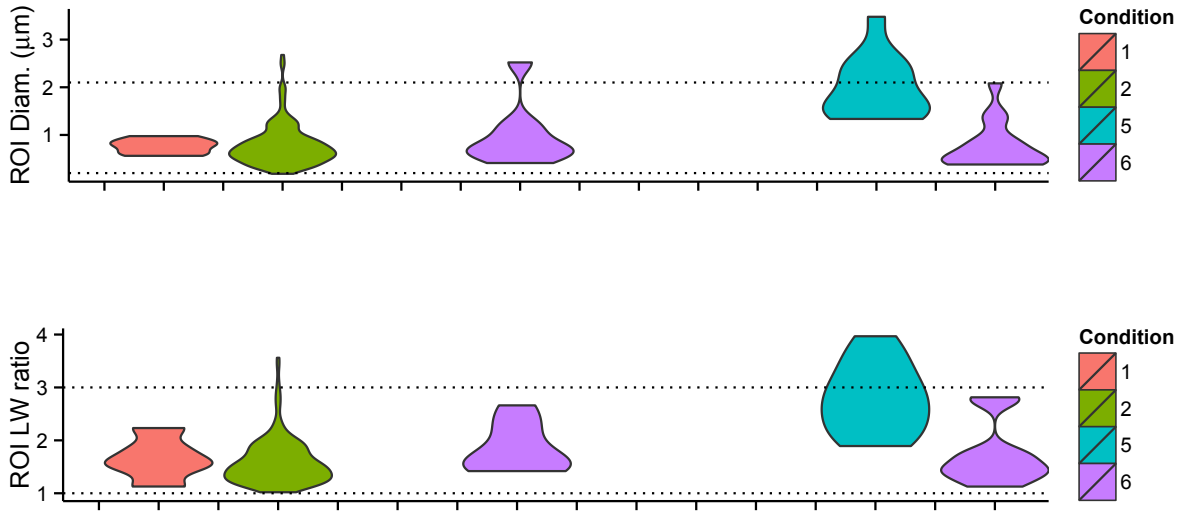
Zilversmit DB, Entenman C, and Fishler MC. 1943. ON THE CALCULATION OF "TURNOVER TIME" AND "TURNOVER RATE" FROM EXPERIMENTS INVOLVING THE USE OF LABELING AGENTS. *The Journal of General Physiology* 26:325-331.



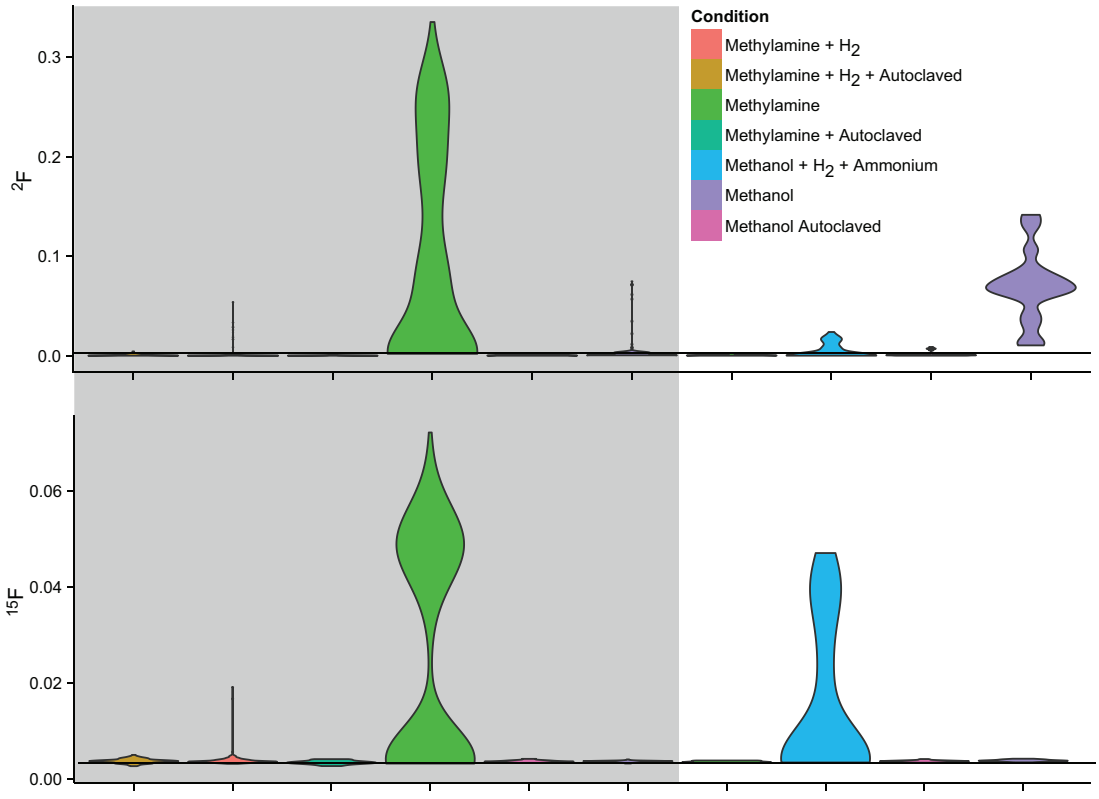
Supplemental Figure 1: a. Location of Site C0020 borehole off Shimokita Peninsula, Northern Honshu Island with bathymetry of the region. Inset contains plate configuration, dominant currents (Kuroshio and Tsushima) and bathymetric map location (red square). b. Seismic profile overlaid with IODP Expedition core recovery. Unit III and IV were used for incubations in this study. Dashed gray line indicates transition from offshore marine. Figures adapted from (2010).



Supplemental Figure 2: Relative methane production from all methylamine and methanol incubations, by lithology.



Supplemental Figure 3: Estimated maximal cell diameter (um) and length to width ratios calculated from NanoSIMS ROIs. Dashed lines indicate deep biosphere ranges (Kallmeyer et al. 2012). Conditions: 1 – methylamine + hydrogen, 2 – methylamine, 5 - methanol + hydrogen + ammonium, 6 – methanol. The first four samples are 15R3 coal and the last two are mixed lithology.



Supplemental Figure 4: Violin plots displaying the kernel density estimation of ^{2}H and ^{15}N fractional abundance from all NanoSIMS ROIs. Gray shading differentiates 15R3 coal (gray) and from mixed lithology samples. No ROIs were recovered from the autoclaved methanol + hydrogen + ammonium conditions so it is not included.

Supplemental Table 1: Isotopic ratios and fractional abundances used for reference and isotope mass balance calculations.

	R	F
^{2}H VSMOW	0.00015576	0.00015574
^{2}H Nat. Abd.	0.00011570	0.00011569
^{13}C VPDB	0.0112372	0.0111123
^{13}C Nat. Abd.	0.0110	0.0109
^{13}C 2km Coal	0.010979	0.010860
^{13}C 2km Methane	0.010563	0.010453
^{15}N Air	0.003677	0.003664
^{15}N Nat. Abd.	0.003642	0.003629
^{15}N 2km Coal	0.003703	0.003689

Supplemental Table 2: IODP Core Identification, top of core depth in meters below seafloor (mbsf), estimated in situ temperature calculated from borehole temperature gradient of 24.0^{o}C/km, porosity ranges, lithology from Site C0020 report. Raw and most likely cell abundances (cells/cm^{3}) and taxonomic information from 16S rRNA gene iTag sequencing from Inagaki et al. 2015.

IODP Core	Depth (mbsf)	Porosity (%)	Lithology	Most Likely (cells/cm^{3})	Dominant Order(s)	Unique Phyla
15R-2	1920.4	24-32	Coal/Shale	3.69	Bacillales	Verrucomicrobia, Tenericutes
15R-7	1925.8	24-32	Coal/Shale	4.66		
19R-1	1951.2	24-32	Siltstone/Shale	0.45	Lactobacillales, Micrococcales	Spirochaetes
19R-5	1954.5	24-32	Siltstone/Shale	30.54		
20R-3	1961.6	25-48	Sandy siltstone	0.10	Micrococcales, Clostridales	Planctomycetes
23R-8	1989.5	25-48	Sandstone	0.20	Gemmatimondales	Gemmatimondales
24R-3	1994.0	25-38	Coal/Shale	6.91	Gemmatimondales	
25R-3	1999.0	25-38	Coal/Sandstone	2.51	Granulicella, Clostridales	Acidobacteria

Appendix A

400 MILLION YEARS OF SILICA BIOMINERALIZATION IN LAND PLANTS

Abstract

Biomineralization plays a fundamental role in the global silicon cycle. Grasses are known to mobilize significant quantities of Si in the form of silica biominerals, and dominate the terrestrial realm today, but have relatively recent origins and only rose to taxonomic and ecological prominence within the Cenozoic Era. This raises questions regarding when, and how, the biological silica cycle evolved. To address these questions, we examined silica abundances of extant members of early-diverging land plant clades, which show silica biomineralization is widespread across terrestrial plant linages. Particularly high silica abundances are observed in lycophytes and early-diverging ferns. However, silica biomineralization is rare within later evolving gymnosperms, implying a complex evolutionary history within the seed plants. Electron microscopy and x-ray spectroscopy show that the most common silica-mineralized tissues include the vascular system, epidermal cells, and stomata—consistent with the hypothesis that biomineralization in plants is frequently coupled to transpiration. Furthermore, sequence, phylogenetic, and structural analysis of nodulin 26-like intrinsic proteins (NIPs) from diverse plant genomes points to a plastic and ancient capacity for silica accumulation within terrestrial plants. The integration of these two comparative biology approaches demonstrates that silica biomineralization has been an important process for land plants over their > 400 myr evolutionary history.

Introduction

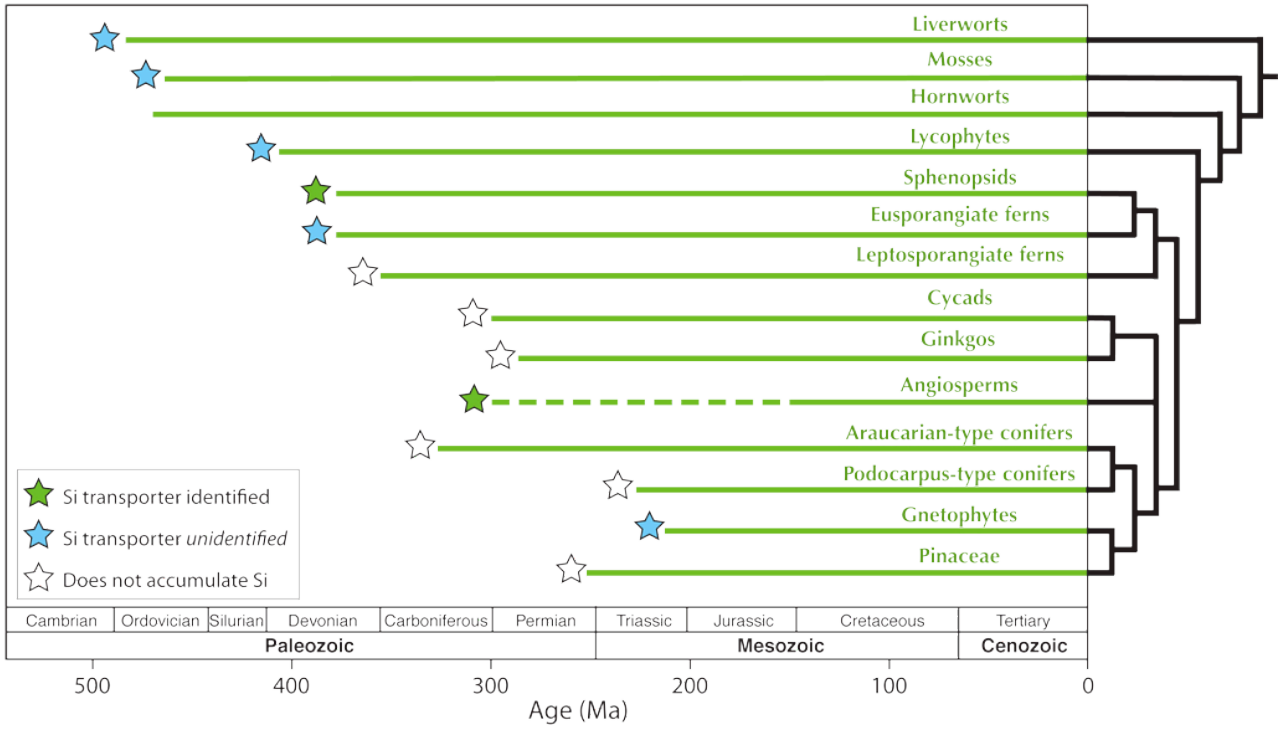
In modern ecosystems, land plants play a major role in the silica cycle through the accumulation and synthesis of silica, an amorphous biomineral composed of SiO_2, known as phytoliths or silica bodies. It is widely appreciated that actively accumulating plants such as grasses are important components of the terrestrial biological pump of silica (Carey & Fulweiler 2012; Conley 2002; Epstein 1994). Plant silica also plays a key role in connecting the terrestrial and marine carbon cycles, because silica is an important nutrient for marine silica-biomineralizing primary producers (i.e. diatoms) (Conley 2002; Epstein 1994; Falkowski et al. 2004; Frings et al. 2014; Raven 1983; Raven 2003). However, both grasses and diatoms evolved in the latter part of the Mesozoic Era (Edwards et al. 2010; Harper & Knoll 1975; Philippe et al. 1994) and rose to ecological dominance within the Cenozoic Era (Edwards et al. 2010; Falkowski et al. 2004; Katz et al. 2004; Stromberg 2004; Stromberg 2005; Stromberg & Feranec 2004). Determining precisely when, and how, the terrestrial-marine silica teleconnections evolved remain obstacles to reconstructing the history of the silica cycle.

Direct analysis of silica bodies in the fossil record provides limited insight into this problem. When fossiliferous material is macerated, it is often challenging to identify whether residual silica bodies are the result of primary biomineralization or secondary diagenetic processes, and if a living plant origin is suspected, it is often difficult to assign taxonomic identity to the phytolith producer. Additionally, with rare exceptions (e.g. Prasad et al. 2005) lagerstätten that preserve exceptional anatomical detail in fossils—and might therefore be expected to preserve silica bodies—tend to be oversaturated with respect to silica (e.g. Kidston & Lang 1920) or extremely undersaturated with respect to silica (e.g. Hatcher et al. 1982; Scott et al. 1996). To account for this, efforts to understand the history of silica biomineralization in terrestrial plants have taken a comparative biology approach (Epstein 1994; Raven 2003).

Silica is widely employed within plants for structural support and pathogen defense (Cooke & Leishman 2011; Ma 2002; Ma & Yamaji 2008), but remains a poorly understood aspect of plant biology. Recent work on the angiosperm *Oryza sativa* demonstrated that silica accumulation is facilitated by transmembrane proteins expressed in root cells (Ma 2006; Ma & Yamaji 2008; Mitani & Ma 2005; Mitani et al. 2008). Phylogenetic analysis revealed that these silicon transport

proteins were derived from a diverse family of modified aquaporins that include arsenite and glycerol transporters (Cooke & Leishman 2011; Liu & Zhu 2010; Ma & Yamaji 2008; Ma et al. 2008). A different member of this aquaporin family was recently identified that enables silica uptake in the horsetail *Equisetum*—an early-diverging fern known to accumulate substantial amounts of silica (Gregoire et al. 2012). However, despite a growing number of fully sequenced genomes, angiosperm-type silicon transporters are not found within the gymnosperms or in spore-bearing plants (Anderberg et al. 2012; Liu & Zhu 2010), including plant lineages that are known to contain many weight-percent silica (Figure 1)**.** A more complete understanding of the distribution and mechanisms of silica accumulation within these early-diverging lineages is a necessary precondition for assessing the evolutionary history of silica biomineralization in terrestrial plants.

Figure 1: Stratigraphic ranges and evolutionary relationships between major terrestrial plant lineages. Although the angiosperm macrofossil record only extends to the Early Cretaceous Period (Sun et al. 2002), a strict interpretation of their position as sister group to all other seed plant clades implies an earlier origin, shown here with a dashed line (Frohlich & Chase 2007; Mathews 2009). For the purposes of this paper, we define Araucarian-type conifers as Araucariaceae and extinct relatives, and Podocarpus-type conifers as Podocarpaceae and extinct relatives. Filled stars mark clades that accumulate > 1 wt.% silica (dry matter), color coded for identified and unidentified silicon transport proteins. Stratigraphic ranges: (Taylor et al. 2008). Evolutionary relationships: (Doyle 2006; Mathews 2009; Qiu et al. 2006).



Methods

Bulk plant silica analysis using dry ashing:

Samples were collected in and around Southern California. Source locations include: Caltech grounds; The Huntington Library, Art Collections, and Botanical Gardens; and Rancho Santa Ana Botanic Garden, the private collection of Loran M. Whitelock, and commercial sources. Plant material (~ 1 g wet weight leaf, sporophyte, or photosynthetic surface sample) was rinsed and dried, and then combusted at 500˚C (Parr et al. 2001). The sample ashes were subsequently washed in 10% HCl at 70˚C, incubated again in 15% H_2O_2 at 70˚C, and dehydrated in ethanol. Final SiO_2 masses are presented as percent initial sample dry weight. Typical uncertainty of the dry ashing method is less than 0.1 wt.% SiO_2 (Ali et al. 1988; Jorhem 1995)—much smaller than the natural variation between tissues of a single plant (e.g. Carnelli et al. 2001). Previously published silica abundance data whose primary sources could be verified (e.g. Hodson et al. 2005) were combined with our results. The complete list of all silica abundances are reported as SiO_2 wt.% in **Supplementary Table S1**.

Imaging and elemental mapping of silica bodies:

A representative subset of washed ash powders was selected for imaging and elemental analysis via electron microscopy and energy dispersive spectroscopy. Samples were pressed gently on to a carbon tape coated SEM stub and either carbon or palladium sputter-coated then imaged with a Zeiss 1550 VP Field Emission Scanning Electron Microscope to observe microstructures. Chemistry was also confirmed by creating spectral element maps with an Oxford INCA Energy 300 X-ray Energy Dispersive Spectrometer system.

NIP phylogeny and structure prediction:

Sequences were collected from the NCBI nr/nt database using the NIP homolog XP_002986711.1 as a query. 1000 sequences were retrieved and aligned and manipulated with Jalview (Waterhouse et al. 2009) and CLUSTALO (Sievers et al. 2011) with a full distance matrix for each iteration and 10 iterations. The alignment was manually trimmed to obtain an alignment block, and a tree constructed with Fasttree (Price et al. 2010). This first tree was used to identify the NIP group. NIP sequences were then collected along with three closely related bacterial homologs, leaving 686 unique NIP protein sequences. These were then re-aligned with CLUSTALW(Larkin et al. 2007)

using default gap extension and opening penalties, and the Gonnet substitution matrix. Prottest (Darriba et al. 2011) was then used to identify an appropriate evolutionary model for tree construction. Fasttree was again used to construct a tree, and this tree was then used as a starting tree for optimization by PhyML (Guindon et al. 2010). The tree was constructed with JTT+I+G, 8 rate substitution categories, the best of NNI's and SPR's, and aBayes was used to evaluate branch supports. From this phylogeny of NIP proteins, each lineage (I, II, and III) was analyzed for sequence conservation using WebLogo3 (Crooks et al. 2004). Residues in the NIP Ar/R filter region were then selected to display diversity at these positions. A representative subset of NIPI,II,III sequences was selected for structure prediction to visualize pore geometries. Models of NIP homologs were generated through sequence submission to the iterative threading assembly refinement (I-TASSER) server (Roy et al. 2010; Zhang 2008; Zhang 2009). The top model based on C-score was selected for further analysis (Wu & Zhang 2007). One of each NIP type was analyzed using PoreWalker (Pellegrini-Calace et al. 2009) to identify pore-lining residues from the modeled structures and observe constriction at the Ar/R gate. All structures were visualized using PyMol. Whole sequence alignments for each of the protein classes, and the phylogenetic trees used in this study can be found in the *Supplementary Material* of the online version of (Trembath-Reichert et al. 2015).

Results and Discussion

We measured SiO_2 content within and across a diverse set of terrestrial plants (88 different plants from 23 families) collected in Southern California, with a focus on lesser-studied lineages with long fossil records. Silica content was assessed gravimetrically on bulk above-ground plant tissues using a modified dry ashing technique, and the resulting silica bodies were imaged using scanning electron microscopy and microscale energy dispersive spectroscopy (see *Methods*). We combined these results with previously published observations (Hodson et al. 2005; Ma 2002; Ma 2006; Ma & Yamaji 2008) to build a coherent picture of silica biomineralization in land plants (Figure 2).

The observed pattern of silica abundance among extant plants (Figure 1, Figure 2) implies a protracted evolutionary history of silica biomineralization and indicates many plant groups with long fossil records precipitate substantial amounts of silica. Accumulation of silica is widespread among diverse land plant families, and variance within groups is also high. Consistent with

previous work, plants with high silica concentrations include members of the monocots, specifically grasses and sedges (Hodson et al. 2005; Ma 2002). However, we also observed that many members of early diverging lineages (e.g., Sellaginellacae, Equisetaceae, Marattiacaeae, and Osmundaceae; toward the left of Figure 2) contain as much or greater amounts of silica than the grasses and sedges (Ma 2002). The only groups that show consistently low silica abundances are found in the gymnosperms, including the conifers, ginkgo, and many cycads. Exceptions are *Gnetum gnemon* and *Cycas revoluta*, which have greater than 1 percent dry weight silica. Beyond gnetophytes and cycads, however, there is a general paucity of silica in gymnosperms suggesting this form of biomineralization is not an important feature of their biology (Carnelli et al. 2001; Hodson et al. 2005; Mitani & Ma 2005).

Additionally, the evolution of seed plants must then require either multiple gains or losses of silica biomineralization. The hypothesis that some lineages of seed plants (Araucarian-type conifers, Podocarpus-type conifers, Pinaceae) have lost biomineralization capacity is possible (Ma 2002; Ma & Yamaji 2008), however the observation that two gnetophytes, *Gnetum* and *Ephedra,* each accumulate silica and contain silicified cell walls (Figure 2, Figure 3) complicates this scenario. This distribution of silica abundance either implies a secondary gain of biomineralization within the gnetophytes, and a loss in the last common ancestor of all gymnosperms, or several independent losses of Si-accumulation within gymnosperms. We use this hypothetical framework to evaluate evolution of the molecular mechanisms of silica accumulation in terrestrial plants.

The most well characterized means by which plants accumulate silicic acid from soil water is via transmembrane proteins with selective pores that belong to a plant-specific subfamily of the aquaporins termed nodulin 26-like intrinsic proteins (NIPs) (see refs: Abascal et al. 2014; Johanson & Gustavsson 2002; Liu & Zhu 2010; Zardoya 2005). Our observations from electron microscopy and spectroscopy confirmed the presence of silicified cell wall structures in diverse taxa, including *Equisetum*, *Selaginella*, and *Gnetum* species (Figure 3**)**. Where we can resolve anatomical structures in the SiO_2 residues, the most heavily biomineralized structures are parts of the vascular system, epidermal cells, and stomata. This is consistent with the hypothesis of a transpiration driven transport process in these plant groups, in which silicic acid is assimilated by roots and is subsequently deposited as silica bodies throughout the plant via distillation (Ma & Yamaji 2008; Mitani & Ma 2005).

Figure 2: Violin plots of silica abundance in terrestrial plant families. Kernel density estimates, green fill (n=688) of silica abundance across terrestrial plant families. White dots are medians, top and bottom of the thick bar mark 1st and 3rd quartile ranges, respectively. Following the family name in parentheses is the number of total analyses within that group, followed by the number of analyses in that group from this study. Families are arranged from left to right in rough order of evolutionary divergence. For clarity, we did not display data from several angiosperm families that are not known to accumulate silica (see *Supplementary Methods*). Conifers and more recently diverging fern clades (e.g., Polypodiaceae) have the lowest medians, whereas liverworts, mosses, lycophytes, and eusporangiate ferns have higher weight percent silica (dry matter), with median values that approach or exceed those found in grasses (Poaceae) and sedges (Cyperaceae).

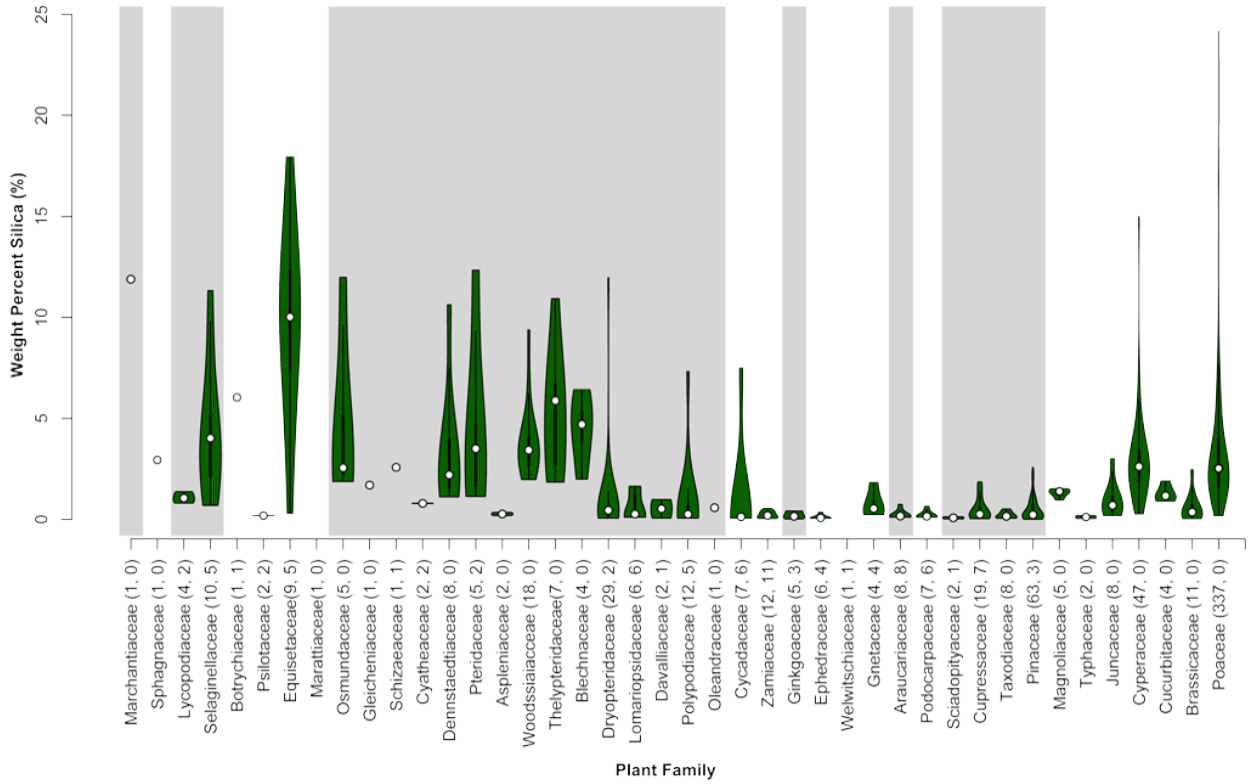
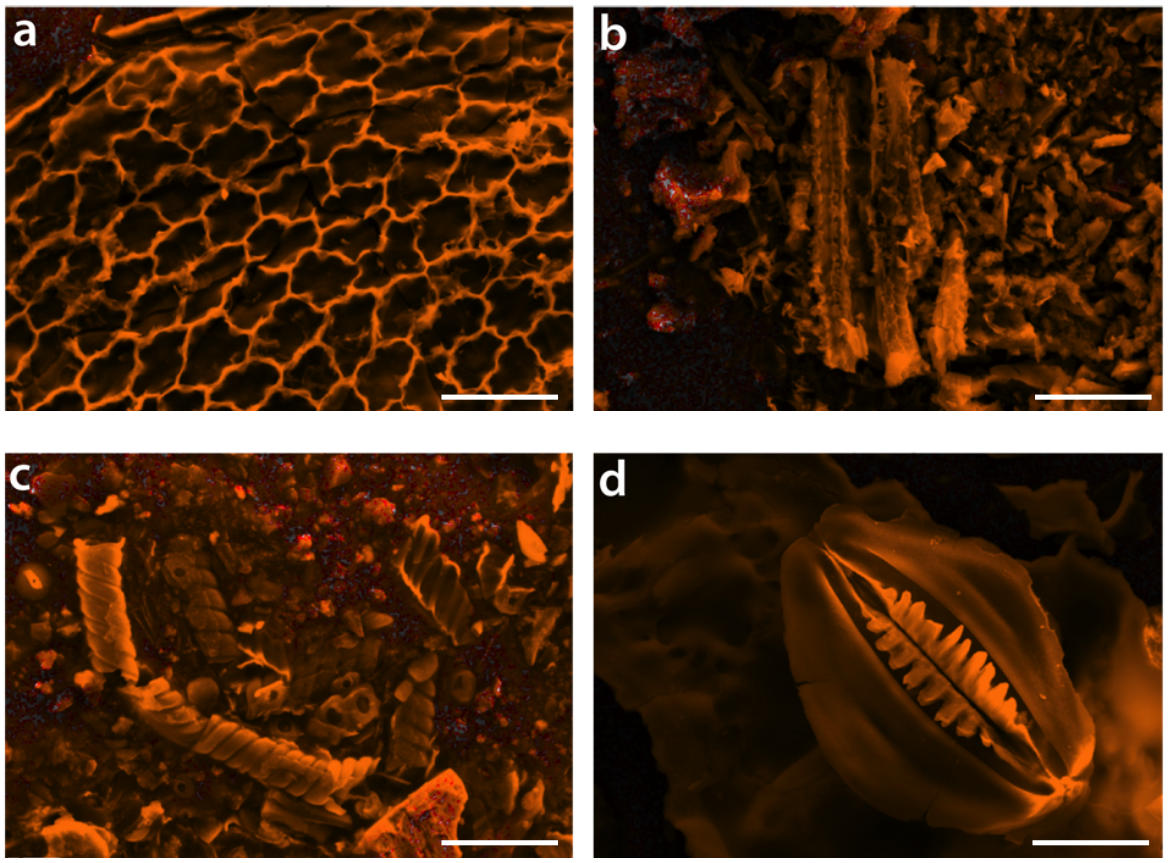


Figure 3: Secondary electron images of silica bodies (grayscale) overlaid with Si maps from energy-dispersive X-ray spectroscopy (orange). Scale bar in all images is 25 µm. a) *Selaginella sp.*, b) *Gnetum gnemon*, c) *Ephedra californicum*, and d) *Equisetum hyemale*. Some distinct mineralized plant tissues can be recognized: in a, epidermal cell walls are silicified; in c, possible silicified vascular tissue; and in d, a silicified stomatal complex. It is noteworthy that these tissues are all near the sites of transpiration.



The NIPs can be subdivided into three groups based on phylogenetic relationships (Zardoya 2005; Zardoya et al. 2002) (Figure 4a). Amino acid residues that surround the narrowest portion of the pore confer a selectivity filter responsible for the exclusion of larger molecules, termed the aromatic/arginine filter (Ar/R filter or gate) (Ar/R filter or gate, Fu et al. 2000; Liu & Zhu 2010). NIPI and NIPII groups are thought to be responsible for the movement of a range of solutes, including arsenite and glycerol (Wallace & Roberts 2005). Of the three major lineages of the NIP proteins, selective transport of orthosilicic acid has been demonstrated in members of the NIPIII (*Lsi1*) and the NIPII group (Figure 4a, arrows), where the presence of a relatively large aperture at the Ar/R filter is thought to permit the passage of silicic acid as compared to smaller constrictions

in other NIPs that would only allow smaller solutes passage (Gregoire et al. 2012; Ma 2006; Ma & Yamaji 2008; Mitani et al. 2008; Mitani-Ueno et al. 2011).

To evaluate the distribution and evolution of silica transport biochemistry in land plants, we constructed a phylogenetic tree and built structural models of key members from all three NIP subgroups (see *Methods*). The phylogenetic analyses recover the expected salient relationships between the NIP subgroups with ~25 times more sequence data then previous reports (Abascal et al. 2014; Liu & Zhu 2010; Zardoya 2005). Results show a complex pattern of functional evolution. Two of the three NIP subgroups have highly conserved residues at Ar/R gate positions (Figure 4a). NIPI is predominantly WVAR (Figure 4a, maroon). Nearly all NIPIII members display GSGR (Figure 4a, purple), with the exceptions of a CSGR bearing homolog in the Cucurbitaceae (*Cucumis melo* and *Cucumis sativus*), where silicon transporters were bred out for rind softening (Liu & Zhu 2010; Piperno et al. 2002) and also in the string bean *Phaseolus vulgaris* that has a single NIPIII homolog with ASGR, and in *Eucalyptus grandis*, which contains a homolog coding for GSPT at the Ar/R gate position. By contrast, NIPII is highly diverse (Figure 4a, orange). The earliest diverging NIPs are found in the moss *Physcomitrella patens*, the lycophyte *Selaginella moellendorffii,* and the fern *Adiantum capillus-verneris* (Figure 4a, asterisk). Sister to these are bacterial NIP-like MIPs (bNIPs, Danielson & Johanson 2010) represented here by sequences from *Ktedonobacter racemifer* and *Nitrolancea hollandica*, both members of the Chloroflexi. Both these bNIPs and the early diverging plant NIPs display the Ar/R gate residues FAAR (or NNAR in the case of the *Selaginella moellendorffii* homolog XP_002986711.1). Proteins with this motif have not been studied *in vivo*, but the prevalence of FAAR residues suggests that the last common ancestor to the plant NIPs may have had conserved function. NIPIIIs form a clade derived from NIPs with the FAAR motif, and are only found in angiosperms. Based upon their conserved Ar/R filter, NIPIIIs facilitate silicic acid uptake (Liu & Zhu 2010; Ma 2008) (Figure 4). NIPIs form a diverse clade, but with conserved pore residues and presumably function as water, glycerol, and lactic acid transporters (Liu & Zhu 2010). By contrast, NIPIIs are not only diverse, but show extreme sequence diversity at the Ar/R gate. Our structural models (Figure 4b) are consistent with the hypothesis that they may transport a range of larger molecules (Liu & Zhu 2010). Included in the NIPIIs are a group of recently identified, highly efficient (twice the silicic acid conductance of *Lsi1*) silicic acid transporters with the previously unreported Ar/R gate residues STAR from the

horsetail *Equisetum arvense*, demonstrating that porins facilitating silicic acid transport have evolved at least twice in plants (Gregoire et al. 2012). Notably, the model structure of the ANAR porins from *Selaginella moellendorffii* has similar size and chemistry to both the STAR porin found in *Equisetum arvense* and the NIPIIIs.

A reasonable evolutionary scenario that satisfies both biochemical and empirical silicic acid abundance data begins with the evolution of NIP-like proteins with an Ar/R conformation of FAAR in bacteria from an ancestral aquaporin, followed by horizontal gene transfer into early terrestrial plants, resulting in the FAAR NIPs found in mosses. The ancestral NIPs subsequently diversified into the NIPI and NIPII clades found throughout land plants, including the functional diversity of pore residues found in the NIPIIs—at least some of which enable selective silicic acid uptake (STAR porin). Despite many fully sequenced genomes, NIPIIs are rare in gymnosperms. It is possible that silica biomineralization was lost in the last common ancestor of seed plants, and angiosperm NIPIIIs constitute a secondary gain of function (Danielson & Johanson 2010; Liu & Zhu 2010), with gnetophyte silica biomineralization currently unresolved, awaiting further molecular data. The NIP phylogeny implies an adaptive radiation of metalloid (including silicic acid) transport early within land plants (Liu & Zhu 2010) and is consistent with our observations of silica biomineralization in early-diverging lineages (Zardoya et al. 2002).

Figure 4: Phylogeny and predicted structures of nodulin 26-like intrinsic protein (NIP) clades. a) Phylogenic tree of major NIP clades with NIP I (maroon), NIP II (orange), and NIP III (purple). The frequency of amino acid occurrence at the Ar/R filter are displayed for the NIPI, NIPII, and NIPIII groups by: Hydrophilic = RKDENQ (blue), Neutral = SGHTAP (green), Hydrophobic =YVMCLFIW (black). The maximum of each scale is 1.0, or 100% probability. Unclassified NIP groups with a conserved Ar/R residues FAAR are indicated by an asterisk. Verified silica transporters include the NIPIIIs and some members of the NIPIIs, shown here with arrows. b) Structural models of four representatives of each of the NIP subgroups. Pore-lining residues are identified with Porewalker. Residues of the Ar/R filter are colored by hydrophobicity on a green scale. A longitudinal transmembrane view of three *Orzya sativa* representatives of the NIPs (I, II, III) is depicted at the top of each panel. Below is the same pore as above with three additional ribbon diagrams rotated to show a transverse view of the pore and four Ar/R gate residues.

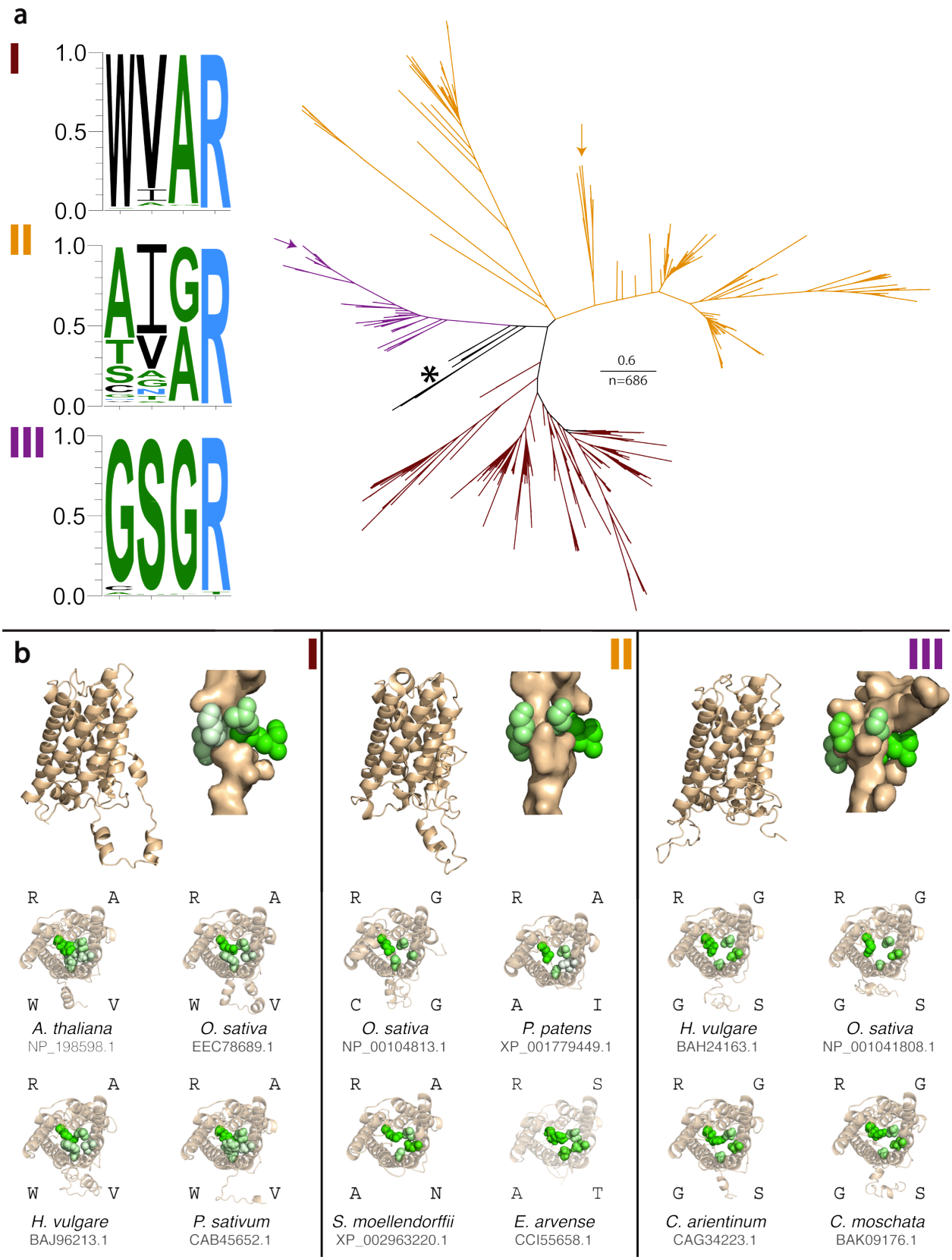
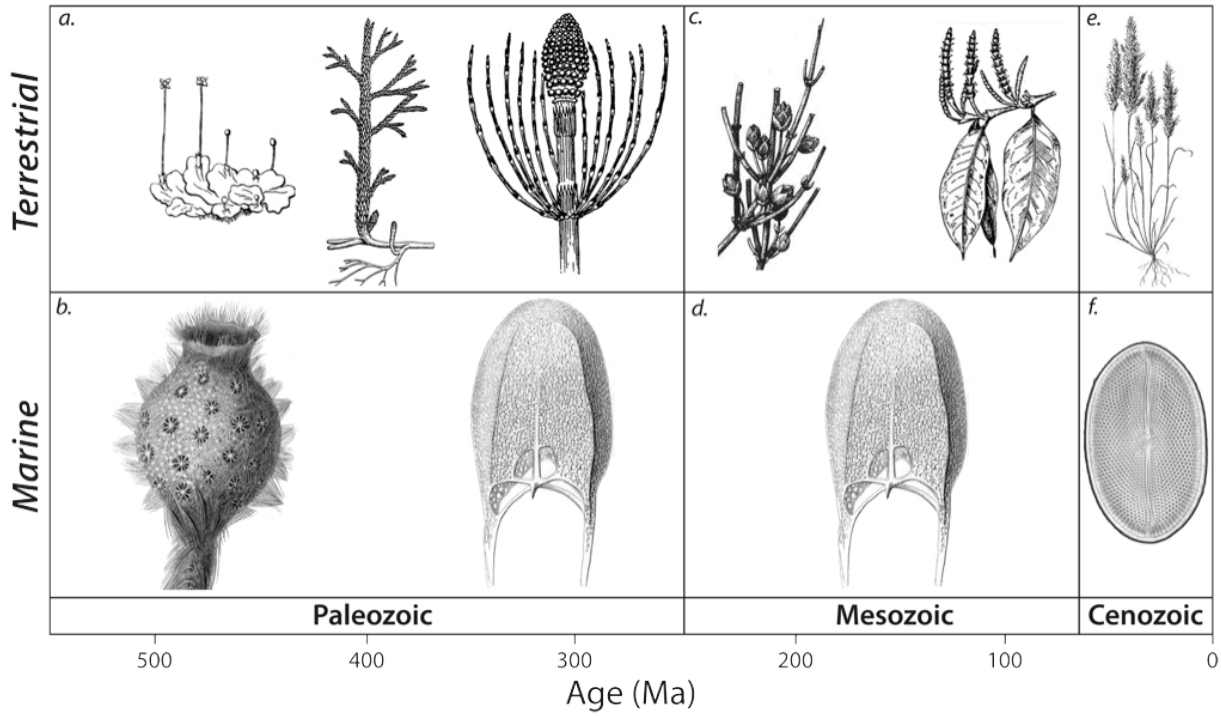

a
I
II
III
1.0
0.5
0.0
0.6
n=686
b
R A
W V
A. thaliana
NP_198598.1
O. sativa
EEC78689.1
R G
C G
O. sativa
NP_00104813.1
R A
A I
P. patens
XP_001779449.1
R G
G S
H. vulgare
BAH24163.1
O. sativa
NP_001041808.1
H. vulgare
BAJ96213.1
P. sativum
CAB45652.1
R A
A N
S. moellendorffii
XP_002963220.1
R S
A T
E. arvense
CCI55658.1
C. arientinum
CAG34223.1
C. moschata
BAK09176.1

Figure 5: Major players in the silica cycle over Phanerozoic time. a) Early-evolving nonvascular and vascular plants, including liverworts, lycophytes, and horsetails. b) Radiolaria and siliceous sponges. c) Seed plants including *Ephedra* and *Gnetum*. d) Mesozoic radiolaria lineages. e) Grasses. f) Diatoms. Fluctuating diversity and abundance of these different taxonomic groups through time, combined with uneven concentration of silica in different plant organs, suggests that plants have played a major, and dynamic, role in the silica cycle over the last 400 million years.



Summary

In order of appearance, major players in the terrestrial silica cycle include some bryophytes (liverworts), lycophytes, and early-diverging vascular plants (horsetails, eusporangiate ferns), followed by gnetophytes and grasses. Terrestrial plant lineages with roots in the Paleozoic Era, including lycophytes, horsetails, and ferns, accumulate silica at abundances comparable to or exceeding many siliceous angiosperm lineages. Combining our results with stratigraphic ranges of silica-biomineralizing plants from the fossil record, we hypothesize that a terrestrial silica cycle must have developed no later than the time of the Rhynie Chert, which contains fossilized stem group bryophytes and vascular plants (411-407 Ma, early Devonian Period; (Kidston & Lang 1920; Mark et al. 2011; Taylor et al. 2008). Consequently, plants may have had a significant impact on the terrestrial silica cycle throughout the Middle and Late Paleozoic Era, with much of the fluxes cycled through lycophytes, horsetails, and early-diverging lineages of ferns that

dominated terrestrial ecosystems at this time (Figure 5). A decrease in continental accumulation of silica may have followed throughout the Mesosoic Era as a consequence of the radiation of conifers, perhaps with some modest silica accumulation in gnetophytes and cycads. Finally, large-scale changes in the terrestrial silica cycle likely occurred with the rise of grasslands late in the Cenozoic Era (Derry et al. 2005; Stromberg 2004; Stromberg 2005; Stromberg & Feranec 2004).

Acknowledgements

The authors thank George Rossman and Victoria Orphan for laboratory equipment, Chi Ma for assistance with electron microscopy and spectroscopy, Sean Lahmeyer (The Huntington Library, Art Collections, and Botanical Gardens), Loran M. Whitelock (Eagle Rock, CA) and Lucinda McDade (Rancho Santa Ana Botanic Garden) for aid in sample collection. This project was partially supported by an O K Earl Postdoctoral Scholarship (Caltech) and a San Andreas Visiting Fellowship to The Huntington Library, Art Collections, and Botanical Gardens to J.P.W. S.E.M. was partially supported by the Agouron Institute as a Geobiology Fellow, and W.W.F. acknowledges support from the Agouron Institute.

References

Abascal F, Irisarri I, and Zardoya R. 2014. Diversity and evolution of membrane intrinsic proteins. *Biochimica et Biophysica Acta (BBA) - General Subjects* 1840:1468-1481. http://dx.doi.org/10.1016/j.bbagen.2013.12.001

Ali MW, Zoltai SC, and Radford FG. 1988. A COMPARISON OF DRY AND WET ASHING METHODS FOR THE ELEMENTAL ANALYSIS OF PEAT. *Canadian Journal of Soil Science* 68:443-447. 10.4141/cjss88-041

Anderberg HI, Kjellbom P, and Johanson U. 2012. Annotation of Selaginella moellendorffii major intrinsic proteins and the evolution of the protein family in terrestrial plants. *Frontiers in Plant Science* 3. 10.3389/fpls.2012.00033

Carey JC, and Fulweiler RW. 2012. The Terrestrial Silica Pump. *Plos One* 7:e52932. 10.1371/journal.pone.0052932

Carnelli AL, Madella M, and Theurillat J-P. 2001. Biogenic Silica Production in Selected Alpine Plant Species and Plant Communities. *Annals of botany* 87:425-434. 10.1006/anbo.2000.1355

Conley DJ. 2002. Terrestrial ecosystems and the global biogeochemical silica cycle. *Global Biogeochemical Cycles* 16:68.

Cooke J, and Leishman MR. 2011. Is plant ecology more siliceous than we realise? *Trends in Plant Science* 16:61-68. 10.1016/j.tplants.2010.10.003

Crooks GE, Hon G, Chandonia JM, and Brenner SE. 2004. WebLogo: A sequence logo generator. *Genome Research* 14:1188-1190. 10.1101/gr.849004

Danielson JAH, and Johanson U. 2010. Phylogeny of Major Intrinsic Proteins. In: Jahn TP, and Bienert GP, eds. *MIPs and Their Role in the Exchange of Metalloids*: Landes Bioscience and Springer Science+Business Media, 19-29.

Darriba D, Taboada GL, Doallo R, and Posada D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* 27:1164-1165. 10.1093/bioinformatics/btr088

Derry LA, Kurtz AC, Ziegler K, and Chadwick OA. 2005. Biological control of terrestrial silica cycling and export fluxes to watersheds. *Nature* 433:728-731.

Doyle JA. 2006. Seed ferns and the origin of angiosperms. *The Journal of the Torrey Botanical Society* 133:169-209. 10.3159/1095-5674(2006)133[169:sfatoo]2.0.co;2

Edwards EJ, Osborne CP, Stromberg CAE, Smith SA, Bond WJ, Christin PA, Cousins AB, Duvall MR, Fox DL, Freckleton RP, Ghannoum O, Hartwell J, Huang YS, Janis CM, Keeley JE, Kellogg EA, Knapp AK, Leakey ADB, Nelson DM, Saarela JM, Sage RF, Sala OE, Salamin N, Still CJ, Tipple B, and Consortium CG. 2010. The Origins of C-4 Grasslands: Integrating Evolutionary and Ecosystem Science. *Science* 328:587-591. 10.1126/science.1177216

Epstein E. 1994. The anomaly of silicon in plant biology. *Proceedings of the National Academy of Sciences* 91:11-17.

Falkowski PG, Katz ME, Knoll AH, Quigg A, Raven JA, Schofield O, and Taylor FJR. 2004. The evolution of modern eukaryotic phytoplankton. *Science* 305:354-360. 10.1126/science.1095964

Frings P, Clymans W, Jeppesen E, Lauridsen T, Struyf E, and Conley D. 2014. Lack of steady-state in the global biogeochemical Si cycle: emerging evidence from lake Si sequestration. *Biogeochemistry* 117:255-277. 10.1007/s10533-013-9944-z

Frohlich MW, and Chase MW. 2007. After a dozen years of progress the origin of angiosperms is still a great mystery. *Nature* 450:1184-1189.

Fu DX, Libson A, Miercke LJW, Weitzman C, Nollert P, Krucinski J, and Stroud RM. 2000. Structure of a glycerol-conducting channel and the basis for its selectivity. *Science* 290:481-486. 10.1126/science.290.5491.481

Gregoire C, Remus-Borel W, Vivancos J, Labbe C, Belzile F, and Belanger RR. 2012. Discovery of a multigene family of aquaporin silicon transporters in the primitive plant Equisetum arvense. *Plant Journal* 72:320-330. 10.1111/j.1365-313X.2012.05082.x

Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, and Gascuel O. 2010. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology* 59:307-321. 10.1093/sysbio/syq010

Harper HE, and Knoll AH. 1975. Silica, diatoms, and Cenozoic radiolarian evolution. *Geology* 3:175-177. 10.1130/0091-7613(1975)3<175:sdacre>2.0.co;2

Hatcher PG, Lyons PC, Thompson CL, Brown FW, and Maciel GE. 1982. Organic matter in a coal ball: peat or coal? *Science* 217:831-833. 10.1126/science.217.4562.831

Hodson MJ, White PJ, Mead A, and Broadley MR. 2005. Phylogenetic variation in the silicon composition of plants. *Annals of botany* 96:1027-1046. 10.1093/aob/mci255

Johanson U, and Gustavsson S. 2002. A new subfamily of major intrinsic proteins in plants. *Molecular Biology and Evolution* 19:456-461.

Jorhem L. 1995. Dry ashing, sources of error, and performance evaluation in AAS. *Microchimica Acta* 119:211-218. 10.1007/BF01244000

Katz ME, Finkel ZV, Grzebyk D, Knoll AH, and Falkowski PG. 2004. Evolutionary trajectories and biogeochemical impacts of marine eukaryotic phytoplankton. *Annual Review of Ecology Evolution and Systematics* 35:523-556. 10.1146/annurev.ecolsys.35.112202.130137

Kidston R, and Lang WH. 1920. On Old Red Sandstone Plants showing Structure, from the Rhynie Chert Bed, Aberdeenshire Part III. Asteroxlon mackiei, Kidston and Lang. *Earth and Environmental Science Transactions of the Royal Society of Edinburgh* 52:643-680.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, and Higgins DG. 2007. Clustal W and clustal X version 2.0. *Bioinformatics* 23:2947-2948. 10.1093/bioinformatics/btm404

Liu Q, and Zhu Z. 2010. Functional divergence of the NIP III subgroup proteins involved altered selective constraints and positive selection. *BMC Plant Biology* 10:256.

Ma JF. 2002. *Soil, Fertilizer, and Plant Silicon Research in Japan*.

Ma JF. 2006. A silicon transporter in rice. *Nature* 440:688-688.

Ma JF. 2008. Functions and transport of silicon in plants. *Cellular and molecular life sciences* 65:3049.

Ma JF, and Yamaji N. 2008. Functions and transport of silicon in plants. *Cellular and molecular life sciences* 65:3049-3057. 10.1007/s00018-008-7580-x

Ma JF, Yamaji N, Mitani N, Xu X-Y, Su Y-H, McGrath SP, and Zhao F-J. 2008. Transporters of arsenite in rice and their role in arsenic accumulation in rice grain. *Proceedings of the National Academy of Sciences of the United States of America* 105:9931-9935. 10.1073/pnas.0802361105

Mark DF, Rice CM, Fallick AE, Trewin NH, Lee MR, Boyce A, and Lee JKW. 2011. 40Ar/39Ar dating of hydrothermal activity, biota and gold mineralization in the Rhynie hot-spring system, Aberdeenshire, Scotland. *Geochimica et cosmochimica acta* 75:555-569.

Mathews S. 2009. Phylogenetic relationships among seed plants: Persistent questions and the limits of molecular data. *Am J Bot* 96:228-236. 10.3732/ajb.0800178

Mitani N, and Ma JF. 2005. Uptake system of silicon in different plant species. *Journal of Experimental Botany* 56:1255-1261. 10.1093/jxb/eri121

Mitani N, Yamaji N, and Ma JF. 2008. Characterization of substrate specificity of a rice silicon transporter, Lsi1. *Pflugers Archiv-European Journal of Physiology* 456:679-686. 10.1007/s00424-007-0408-y

Mitani-Ueno N, Yamaji N, Zhao F-J, and Ma JF. 2011. The aromatic/arginine selectivity filter of NIP aquaporins plays a critical role in substrate selectivity for silicon, boron, and arsenic. *Journal of Experimental Botany* 62:4391-4398. 10.1093/jxb/err158

Parr JF, Lentfer CJ, and Boyd WE. 2001. A comparative analysis of wet and dry ashing techniques for the extraction of phytoliths from plant material. *Journal of Archaeological Science* 28:875-886. 10.1006/jasc.2000.0623

Pellegrini-Calace M, Maiwald T, and Thornton JM. 2009. PoreWalker: A Novel Tool for the Identification and Characterization of Channels in Transmembrane Proteins from Their Three-Dimensional Structure. *PLoS Comput Biol* 5:e1000440. 10.1371/journal.pcbi.1000440

Philippe H, Sörhannus U, Baroin A, Perasso R, Gasse F, and Adoutte A. 1994. Comparison of molecular and paleontological data in diatoms suggests a major gap in the fossil record. *Journal of Evolutionary Biology* 7:247-265. 10.1046/j.1420-9101.1994.7020247.x

Piperno DR, Holst I, Wessel-Beaver L, and Andres TC. 2002. Evidence for the control of phytolith formation in Cucurbita fruits by the hard rind (Hr) genetic locus: Archaeological and ecological implications. *Proceedings of the National Academy of Sciences* 99:10923-10928. 10.1073/pnas.152275499

Prasad V, Stromberg CAE, Alimohammadian H, and Sahni A. 2005. Dinosaur coprolites and the early evolution of grasses and grazers. *Science* 310:1177-1180. 10.1126/science.1118806

Price MN, Dehal PS, and Arkin AP. 2010. FastTree 2-Approximately Maximum-Likelihood Trees for Large Alignments. *Plos One* 5. 10.1371/journal.pone.0009490

Qiu Y-L, Li L, Wang B, Chen Z, Knoop V, Groth-Malonek M, Dombrovska O, Lee J, Kent L, Rest J, Estabrook GF, Hendry TA, Taylor DW, Testa CM, Ambros M, Crandall-Stotler B, Duff RJ, Stech M, Frey W, Quandt D, and Davis CC. 2006. The deepest divergences in land plants inferred from phylogenomic evidence. *Proceedings of the National Academy of Sciences* 103:15511-15516. 10.1073/pnas.0603335103

Raven JA. 1983. THE TRANSPORT AND FUNCTION OF SILICON IN PLANTS. *Biological Reviews* 58:179-207. 10.1111/j.1469-185X.1983.tb00385.x

Raven JA. 2003. Cycling silicon – the role of accumulation in plants. *New Phytologist* 158:419-421. 10.1046/j.1469-8137.2003.00778.x

Roy A, Kucukural A, and Zhang Y. 2010. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols* 5:725-738. 10.1038/nprot.2010.5

Scott AC, Mattey DP, and Howard R. 1996. New data on the formation of Carboniferous coal balls. *Review of Palaeobotany and Palynology* 93:317-331. http://dx.doi.org/10.1016/0034-6667(95)00132-8

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soeding J, Thompson JD, and Higgins DG. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* 7. 10.1038/msb.2011.75

Stromberg CAE. 2004. Using phytolith assemblages to reconstruct the origin and spread of grass-dominated habitats in the great plains of North America during the late Eocene to early Miocene. *Palaeogeography Palaeoclimatology Palaeoecology* 207:239-275. 10.1016/j.palaeo.2003.09.028

Stromberg CAE. 2005. Decoupled taxonomic radiation and ecological expansion of open-habitat grasses in the Cenozoic of North America. *Proceedings of the National Academy of Sciences of the United States of America* 102:11980-11984. 10.1073/pnas.0505700102

Stromberg CAE, and Feranec RS. 2004. The evolution of grass-dominated ecosystems during the late Cenozoic. *Palaeogeography Palaeoclimatology Palaeoecology* 207:199-201. 10.1016/j.palaeo.2004.01.017

Sun G, Ji Q, Dilcher DL, Zheng S, Nixon KC, and Wang X. 2002. Archaefructaceae, a new basal angiosperm family. *Science* 296:899-904. 10.1126/science.1069439

Taylor TN, Taylor EL, and Krings M. 2008. *Paleobotany: The Biology and Evolution of Fossil Plants*: Academic Press.

Trembath-Reichert E, Wilson JP, McGlynn SE, and Fischer WW. 2015. Four hundred million years of silica biomineralization in land plants. *Proceedings of the National Academy of Sciences of the United States of America* 112:5449-5454. 10.1073/pnas.1500289112

Wallace IS, and Roberts DM. 2005. Distinct Transport Selectivity of Two Structural Subclasses of the Nodulin-like Intrinsic Protein Family of Plant Aquaglyceroporin Channels†. *Biochemistry* 44:16826-16834. 10.1021/bi0511888

Waterhouse AM, Procter JB, Martin DMA, Clamp M, and Barton GJ. 2009. Jalview Version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25:1189-1191. 10.1093/bioinformatics/btp033

Wu S, and Zhang Y. 2007. LOMETS: A local meta-threading-server for protein structure prediction. *Nucleic Acids Research* 35:3375-3382. 10.1093/nar/gkm251

Zardoya R. 2005. Phylogeny and evolution of the major intrinsic protein family. *Biology of the Cell* 97:397-414. 10.1042/bc20040134

Zardoya R, Ding XD, Kitagawa Y, and Chrispeels MJ. 2002. Origin of plant glycerol transporters by horizontal gene transfer and functional recruitment. *Proceedings of the National Academy of Sciences of the United States of America* 99:14893-14896. 10.1073/pnas.192573799

Zhang Y. 2008. I-TASSER server for protein 3D structure prediction. *Bmc Bioinformatics* 9. 10.1186/1471-2105-9-40

Zhang Y. 2009. I-TASSER: Fully automated protein structure prediction in CASP8. *Proteins-Structure Function and Bioinformatics* 77:100-113. 10.1002/prot.22588

Supplemental Table 1: Silica abundance data for all samples from this study combined with those from previous work, organized by species Name, Order, Family, Silica (wt. %), Reference, and collection Location. Full reference information for previous studies can be found in Appendix of Hodson et al. 2005.

Name	Silica (wt. %)	Reference	Location
Abies alba	0.08	Carnelli et al. (2001)	
Abies balsamea	0.18	Klein & Geis (1978)	
Abies fraseri	0.13	Klein & Geis (1978)	
Abies grandis	0.84	Hodson et al. (1997)	
Abies mariesii	0.12	Hodson et al. (1997)	
Abies nordmann	0.06	This Study	Huntington
Abies nordmann	0.10	This Study	Huntington

Abies nordmanniana	0.39	Hodson et al. (1997)	
Abies pectinata	11.34	Bartoli & Beaucire (1976)	
Abies pectinata	13.37	Bartoli & Beaucire (1976)	
Abies procera	0.93	Hodson et al. (1997)	
Acer ginnala	1.13	Ma & Takahashi (2002)	
Acer negundo	0.34	Geis (1973)	
Acer rubrum	1.49	Lanning & Eleuterius (1983)	
Acer saccharinum	0.98	Geis (1973)	
Acer saccharum	2.63	Geis (1973)	
Acer saccharum	1.84	Ma & Takahashi (2002)	
Achillea millefolium	4.20	Hogenbirk & Sarrazin-Delay (1995)	
Achnatherum hymenoides	2.00	Blank *et al.* (1994)	
Achnatherum hymenoides	4.00	Blank *et al.* (1994)	
Aconitum japonicum	0.75	Ma & Takahashi (2002)	
Aconitum loczyanum	0.68	Ma & Takahashi (2002)	
Acorus calamus	0.09	Ma (2002)	
Acorus gramineus	0.21	Ma (2002)	
Acrophorus stipellatus	2.42	Ma & Takahashi (2002)	
Adiantum pedatum	1.13	Ma & Takahashi (2002)	
Adiantum pedatum	4.71	Ma & Takahashi (2002)	
Aegilops squarrosa	4.51	Ma & Takahashi (2002)	
Aesculus pavia	0.30	Lanning & Eleuterius (1983)	
Agathis robusta	0.11	This Study	Huntington
Agathis robusta	0.17	This Study	Huntington
Agave americana	0.30	Ma & Takahashi (2002)	
Agropyron cristatum	1.85	Bezeau et al. (1996)	
Agropyron dasystachyum	1.81	Bezeau *et al.* (1966)	
Agropyron repens	3.50	Hogenbirk & Sarrazin-Delay (1995)	
Agropyron smithii	2.20	Bezeau *et al.* (1966)	
Agropyron smithii	2.52	Bezeau et al. (1996)	
Agropyron subsecundum	2.23	Bezeau et al. (1996)	
Agropyron trichophorum	2.20	Bezeau et al. (1996)	
Agrostis alba	6.74	Butler & Hodges (1967)	

Agrostis gigantea	8.30	Hogenbirk & Sarrazin-Delay (1995)
Agrostis palustris	2.40	Barbehenn (1993)
Agrostis scabra	1.70	Hogenbirk & Sarrazin-Delay (1995)
Agrostis stolonifera	10.48	Tyler (1971)
Agrostis tenuis	2.89	Pahkala & Pihala (2000)
Agrostis tenuis	3.04	Pahkala & Pihala (2000)
Agrostis tenuis	3.06	Pahkala & Pihala (2000)
Alangium platanifolium	0.40	Nakanishi et al. (2003)
Albizia julibrissin	0.21	Ma & Takahashi (2002)
Alhagi mannifera	0.17	Cowgill (1989)
Allium fistulosum	0.36	Ma & Takahashi (2002)
Alnus viridis	0.13	Carnelli et al. (2001)
Aloë arborescens	0.34	Ma & Takahashi (2002)
Alternanthera sessilis	0.60	Cowgill (1989)
Amaranthus albus	0.14	Cowgill (1989)
Amaranthus gracilis	0.13	Cowgill (1989)
Amaranthus graecizans	0.09	Cowgill (1989)
Amaranthus retroflexus	0.43	Cowgill (1989)
Amaranthus spp.	5.00	Bilbro et al. (1991)
Amaranthus viridis	0.32	Ma & Takahashi (2002)
Ammi visnaga	0.37	Cowgill (1989)
Amorphophallus rivieri	0.09	Ma & Takahashi (2002)
Ananas comosus	0.51	Ma & Takahashi (2002)
Anaphalis margaritacea	0.70	Hogenbirk & Sarrazin-Delay (1995)
Andropogon gerardii	6.79	Geis 1978
Andropogon gerardii	2.89	Lanning & Eleuterius (1987)
Andropogon scoparius	9.25	Lanning & Eleuterius (1987)
Anemarrhena asphodeloides	0.17	Ma & Takahashi (2002)
Angiopteris lygodiifolia	3.55	Ma & Takahashi (2002)
Anthoxanthum odoratum	0.88	Cornelissen & Thompson (1997)
Anthoxanthum odoratum	1.09	Cornelissen & Thompson (1997)
Aquilegia flabellata	0.11	Ma & Takahashi (2002)
Aralia cordata	0.17	Ma & Takahashi (2002)

Araucaria araucana	0.85	Hodson *et al.* (1997)	
Araucaria sp.	0.11	This Study	Pasadena, CA
Araucaria sp.	0.15	This Study	Huntington
Araucaria sp.	0.15	This Study	Caltech
Araucaria sp.	0.17	This Study	Pasadena, CA
Araucaria sp.	0.66	This Study	Pasadena, CA
Araucaria sp.	0.74	This Study	Pasadena, CA
Arctostaphylos uva-ursi	0.04	Carnelli et al. (2001)	
Aristida stricta	2.48	Kalisz & Stone (1984)	
Armoracia rusticana	0.09	Ma & Takahashi (2002)	
Aronia melanocarpa	3.08	Kolesnikov & Gins (2001)	
Arrhenatherum elatius	1.80	Cornelissen & Thompson (1997)	
Artemisia absinthium	0.68	Ma & Takahashi (2002)	
Artemisia cana	0.33	Bezeau et al. (1996)	
Artemisia frigida	0.99	Bezeau et al. (1996)	
Artemisia gnaphalodes	0.16	Bezeau et al. (1996)	
Artemisia maritima	0.21	Ma & Takahashi (2002)	
Artemisia tridentata	1.40	Blank et al. (1994)	
Artemisia tridentata	0.60	Blank et al. (1994)	
Arundinaria gigantea	18.16	Lanning & Eleuterius (1985)	
Arundo donax	3.20	Bilbro et al. (1991)	
Arundo donax	2.65	Ma & Takahashi (2002)	
Asparagus cochinchinensis	0.53	Ma & Takahashi (2002)	
Asparagus officinalis	1.65	Kolesnikov & Gins (2001)	
Asparagus officinalis	0.58	Ma & Takahashi (2002)	
Aspidistra elatior	0.13	Ma & Takahashi (2002)	
Asplenium cuneifolium	0.34	Höhne & Richter (1981)	
Asplenium trichomanes	0.19	Ma & Takahashi (2002)	
Aster laevis	0.45	Bezeau et al. (1996)	
Aster macrophyllus	2.60	Hogenbirk & Sarrazin-Delay (1995)	
Aster tenuifolia	0.03	Lanning & Eleuterius (1983)	
Aster tenuifolia	0.09	Lanning & Eleuterius (1983)	
Aster tripolium	0.24	de Bakker et al. (1999)	

Athyrium filix-femina	4.09	Höhne (1963)
Athyrium filix-femina	2.80	Höhne (1963)
Athyrium filix-femina	1.97	Höhne (1963)
Athyrium filix-femina	3.47	Höhne (1963)
Athyrium filix-femina	2.33	Höhne (1963)
Athyrium filix-femina	2.46	Höhne & Richter (1981)
Athyrium filix-femina	3.38	Höhne & Richter (1981)
Athyrium japonicum	2.61	Ma & Takahashi (2002)
Athyrium lobato-crenatum	4.09	Ma & Takahashi (2002)
Athyrium niponicum	2.08	Ma & Takahashi (2002)
Athyrium yokoscense	3.06	Ma & Takahashi (2002)
Atriplex canescens	0.20	Bilbro et al. (1991)
Atriplex littoralis	0.00	de Bakker et al. (1999)
Atriplex nuttallii	0.60	Bezeau et al. (1996)
Atriplex portulacoides	0.32	de Bakker et al. (1999)
Atriplex prostrata	0.00	de Bakker et al. (1999)
Atriplex rosea	0.29	Cowgill (1989)
Atropa belladonna	0.06	Ma & Takahashi (2002)
Aucuba japonica	1.24	Ma & Takahashi (2002)
Aucuba japonica	0.53	Nakanishi et al. (2003)
Avena sativa	2.72	Bertrand & Ghitescu (1934)
Avena sativa	0.58	Grosse-Brauckmann (1953)
Avena sativa	2.03	Grosse-Brauckmann (1953)
Avena sativa	2.04	Jones & Handreck (1967)
Avena sativa	4.45	Ma & Takahashi (2002)
Avena sativa	2.90	McManus et al. (1977)
Avena sativa	3.68	Saijonkari-Pahkala (2001)
Avena sativa	5.13	Schnug & v. Franck (1985)
Avena sativa	8.56	Schnug & v. Franck (1985)
Avena sativa	17.12	Schnug & v. Franck (1985)
Baccharis halimifolia	0.28	Ma & Takahashi (2002)
Baccharis trimera	0.82	Pereira & Felcman (1998)
Bacopa monnieri	1.18	Lanning & Eleuterius (1983)

Ballota undulata	0.40	Cowgill (1989)	
Batis maritima	0.27	Lanning & Eleuterius (1985)	
Benincasa hispida	0.90	Ma & Takahashi (2002)	
Betonica foliosa	2.44	Kolesnikov & Gins (2001)	
Betula pendula	0.83	Bartoli & Beaucire (1976)	
Blechnum amabile	1.99	Ma & Takahashi (2002)	
Blechnum spicant	6.42	Höhne & Richter (1981)	
Blechnum spicant	4.36	Höhne & Richter (1981)	
Bletilla striata	1.01	Ma & Takahashi (2002)	
Boltonia asteroides	2.00	Lanning & Eleuterius (1983)	
Borrichia frutescens	0.38	Lanning & Eleuterius (1983)	
Borrichia frutescens	0.39	Lanning & Eleuterius (1985)	
Botrychium virginianum	6.04	This Study	Commercial
Bouteloua curtipendula	5.64	Smith et al. (1971)	
Bouteloua gracilis	2.17	Bezeau et al. (1996)	
Bouteloua gracilis	2.51	Johnston et al. (1967)	
Bouteloua gracilis	2.45	Johnston et al. (1967)	
Bouteloua gracilis	2.60	Johnston et al. (1967)	
Bouteloua gracilis	3.24	Johnston et al. (1967)	
Bouteloua gracilis	1.30	Johnston et al. (1967)	
Bouteloua gracilis	2.26	Johnston et al. (1967)	
Bouteloua gracilis	2.51	Johnston et al. (1967)	
Bouteloua gracilis	2.12	Johnston et al. (1967)	
Bouteloua gracilis	1.77	Johnston et al. (1967)	
Bouteloua gracilis	2.73	Johnston et al. (1967)	
Bouteloua gracilis	5.72	Smith et al. (1971)	
Bouteloua hirsuta	6.68	Smith et al. (1971)	
Brachypodium pinnatum	1.97	Cornelissen & Thompson (1997)	
Brachypodium sylvaticum	6.50	Höhne (1963)	
Brachypodium sylvaticum	5.56	Höhne (1963)	
Brachypodium sylvaticum	8.41	Höhne (1963)	
Brachypodium sylvaticum	5.71	Höhne (1963)	
Brassica alba	0.15	Jones & Handreck (1967)	

Brassica napus	0.03	Bertrand & Ghitescu (1934)
Brassica napus	0.36	Saijonkari-Pahkala (2001)
Brassica rapa	0.14	Saijonkari-Pahkala (2001)
Brassica rapa	0.43	Schnug & v. Franck (1985)
Brassica rapa	1.07	Schnug & v. Franck (1985)
Brassica rapa	1.07	Schnug & v. Franck (1985)
Briza media	1.80	Cornelissen & Thompson (1997)
Bromus inermis	2.47	Bezeau et al. (1996)
Bromus inermis	4.30	Hogenbirk & Sarrazin-Delay (1995)
Bromus inermis	1.30	Robbins et al. (1987)
Bromus pumpellianus	1.86	Bezeau et al. (1996)
Bromus tectorum	1.50	Blank et al. (1994)
Bromus tectorum	3.40	Blank et al. (1994)
Bromus tectorum	2.40	Robbins et al. (1987)
Cajanus cajan	2.80	Bilbro et al. (1991)
Calamagrostis canadensis	4.90	Hogenbirk & Sarrazin-Delay (1995)
Calamagrostis epigejos	4.58	Höhne (1963)
Calamagrostis epigejos	6.33	Höhne (1963)
Calamagrostis epigejos	4.02	Höhne (1963)
Calamagrostis epigejos	4.17	Höhne (1963)
Calamagrostis inexpansa	3.80	Bezeau et al. (1996)
Calamagrostis rubescens	3.29	Bezeau et al. (1996)
Calamagrostis villosa	5.90	Carnelli et al. (2001)
Calamagrostis villosa	7.40	Höhne (1963)
Calamagrostis villosa	3.27	Höhne (1963)
Calamagrostis villosa	5.24	Höhne (1963)
Callistemon rigidus	0.19	Ma & Takahashi (2002)
Calluna vulgaris	5.35	Bartoli & Beaucire (1976)
Calluna vulgaris	0.72	Carnelli et al. (2001)
Calluna vulgaris	1.69	Höhne (1963)
Calystegia japonica	0.04	Ma & Takahashi (2002)
Calystegia sepium	0.13	Cowgill (1989)
Camellia japonica	0.13	Ma & Takahashi (2002)

Camellia japonica	0.24	Nakanishi et al. (2003)
Camellia sasanqua	0.32	Ma & Takahashi (2002)
Camellia sasanqua	0.45	Nakanishi et al. (2003)
Campsis grandiflora	0.30	Ma & Takahashi (2002)
Canna indica	0.77	Ma & Takahashi (2002)
Cannabis sativa	0.19	Saijonkari-Pahkala (2001)
Capsicum annuum	0.11	Taber et al. (2002)
Carex aquatilis	1.15	Gadallah & Jefferies (1995)
Carex atherodes	2.71	Bezeau et al. (1996)
Carex biwensis	2.61	Ma & Takahashi (2002)
Carex cinica	5.43	Ma & Takahashi (2002)
Carex curvula	1.03	Carnelli et al. (2001)
Carex dispalata	5.16	Ma & Takahashi (2002)
Carex filifolia	2.76	Bezeau et al. (1996)
Carex filifolia	1.21	Johnston et al. (1967)
Carex filifolia	3.66	Johnston et al. (1967)
Carex filifolia	2.99	Johnston et al. (1967)
Carex filifolia	3.10	Johnston et al. (1967)
Carex filifolia	1.48	Johnston et al. (1967)
Carex filifolia	3.92	Johnston et al. (1967)
Carex filifolia	2.47	Johnston et al. (1967)
Carex filifolia	2.88	Johnston et al. (1967)
Carex filifolia	2.47	Johnston et al. (1967)
Carex filifolia	3.33	Johnston et al. (1967)
Carex flacca	1.82	Cornelissen & Thompson (1997)
Carex flavicans	2.20	Gadallah & Jefferies (1995)
Carex parciflora	3.70	Ma & Takahashi (2002)
Carex sempervirens	2.31	Carnelli et al. (2001)
Carex subspathacea	2.70	Gadallah&Jefferies (1995)
Carex thunbergii	3.57	Ma & Takahashi (2002)
Carpinus caroliniana	0.62	Geis (1973)
Carya cordiformis	0.26	Geis (1973)
Carya laciniosa	0.32	Geis (1973)

Carya ovata	0.46	Geis (1973)	
Carya tomentosa	0.36	Geis (1973)	
Caryota mitis	2.76	Lanning (1966)	
Catalpa ovata	0.43	Ma & Takahashi (2002)	
Catalpa ovata	0.97	Nakanishi et al. (2003)	
Cedrus atlantica	0.09	Hodson et al. (1997)	
Celtis occidentalis	3.44	Geis (1973)	
Celtis occidentalis	8.80	Wilding & Drees (1971)	
Cenchrus longispinus	3.38	Lanning & Eleuterius (1987)	
Centaurea iberica	0.32	Cowgill (1989)	
Ceratiola ericoides	0.07	Lanning & Eleuterius (1985)	
Ceratozamia hildae	0.10	This Study	Huntington
Cercidiphyllum japonicum	2.01	Ma & Takahashi (2002)	
Cercidiphyllum japonicum	0.82	Nakanishi et al. (2003)	
Cercis canadensis	0.20	Geis (1973)	
Chaenomeles sinensis	0.62	Ma & Takahashi (2002)	
Chamaecyparis lawsoniana	1.85	Hodson et al. (1997)	
Chamaecyparis obtusa	1.09	Hodson et al. (1997)	
Chamaecyparis obtusa	0.26	Ma & Takahashi (2002)	
Chamaecyparis pisifera	0.79	Hodson et al. (1997)	
Chamaecyparis thyoides	1.38	Lanning & Eleuterius (1985)	
Chamerion angustifolium	0.13	Cornelissen & Thompson (1997)	
Chamomilla recutita	0.34	Ma & Takahashi (2002)	
Chasmanthium latifolium	7.74	Lanning & Eleuterius (1989)	
Chasmanthium sessiliflorum	0.43	Lanning & Eleuterius (1989)	
Chenopodium album	0.18	Cowgill (1989)	
Chenopodium album	0.00	Lanning & Eleuterius (1983)	
Chenopodium murale	0.36	Cowgill (1989)	
Chenopodium opulifolium	0.23	Cowgill (1989)	
Chrysanthemum coronarium	0.39	Ma & Takahashi (2002)	
Chrysanthemum morii	0.62	Reay & Bennett (1987)	
Citrullus lanatus	1.90	Taber et al. (2002)	
Cladium jamaicense	2.35	Lanning & Eleuterius (1983)	

Cladium mariscus	2.02	Cowgill (1989)	
Clethra alnifolia	1.97	Lanning & Eleuterius (1985)	
Cleyera ochnacea	0.27	Nakanishi et al. (2003)	
Cliftonia monophylla	0.30	Lanning & Eleuterius (1985)	
Coffea arabica	0.17	Lanning (1966)	
Colysis decurrens	0.64	Ma & Takahashi (2002)	
Colysis wrightii	0.06	Ma & Takahashi (2002)	
Conium maculatum	0.06	Ma & Takahashi (2002)	
Convallaria majalis	1.13	Ma & Takahashi (2002)	
Conyza canadensis	0.16	Cowgill (1989)	
Cornucopiae cucullatum	0.80	Cowgill (1989)	
Cornus stolonifera	0.20	Geis (1973)	
Cortaderia selloana	0.21	Lanning & Eleuterius (1989)	
Cortaderia selloana	1.39	Ma & Takahashi (2002)	
Corylopsis pauciflora	0.37	Nakanishi et al. (2003)	
Crataegus cuneata	0.24	Ma & Takahashi (2002)	
Crinum asiaticum	0.04	Ma & Takahashi (2002)	
Cryptomeria japonica	0.18	Hodson et al. (1997)	
Cryptomeria japonica	0.30	Ma & Takahashi (2002)	
Ctenitis subglandulosa	1.05	Ma & Takahashi (2002)	
Ctenium aromaticum	5.91	Lanning & Eleuterius (1985)	
Cunninghamia lanceolata	0.06	Hodson et al. (1997)	
Cunninghamia lanceolata	0.13	Ma & Takahashi (2002)	
Cupania oblongifolia	0.27	Pereire & Felcman (1998)	
Cupressocyparis leylandii	1.73	Hodson et al. (1997)	
Cupressus sempervirens	0.21	Ma & Takahashi (2002)	
Cuscuta planiflora	0.16	Cowgill (1989)	
Cuscuta racemosa	0.03	Pereire & Felcman (1998)	
Cyathea cooperi	0.78	This Study	Huntington
Cyathea cooperi	0.80	This Study	Huntington
Cycas revoluta	0.15	Ma & Takahashi (2002)	
Cycas revoluta	0.05	This Study	Huntington
Cycas revoluta	0.10	This Study	Huntington

Cycas revoluta	0.13	This Study	Huntington
Cycas revoluta	7.49	This Study	Huntington
Cycas revoluta (less dry)	0.08	This Study	Huntington
Cycas revoluta (less dry)	0.25	This Study	Huntington
Cyclosorus acuminatus	7.00	Ma & Takahashi (2002)	
Cyclosorus dentatus	10.93	Ma & Takahashi (2002)	
Cymbopogon citratus	1.82	Ma & Takahashi (2002)	
Cynanchum acutum	0.15	Cowgill (1989)	
Cynodon dactylon	4.47	Barbehenn (1993)	
Cynodon dactylon	7.68	Butler & Hodges (1967)	
Cynodon dactylon	3.08	Lanning (1966)	
Cynodon dactylon	2.49	Street (1974)	
Cynodon dactylon	3.85	Street (1974)	
Cynodon dactylon	0.32	Street (1974)	
Cynodon dactylon	2.49	Street (1974)	
Cynodon dactylon	3.85	Street (1974)	
Cyperus alopecuroides	0.49	Cowgill (1989)	
Cyperus alternifolius	7.53	Ma & Takahashi (2002)	
Cyperus latifolius	1.00	Cowgill (1989)	
Cyperus michelianus	1.04	Cowgill (1989)	
Cyperus microiria	1.95	Ma & Takahashi (2002)	
Cyperus papyrus	0.27	Cowgill (1989)	
Cyperus papyrus	1.20	Lanning (1966)	
Cyperus papyrus	3.74	Ma & Takahashi (2002)	
Cyperus polystachyos	1.86	Lanning & Eleuterius (1985)	
Cyperus surinamensis	2.36	Lanning & Eleuterius (1985)	
Cyrtomium falcatum	0.28	Ma & Takahashi (2002)	
Cyrtomium falcatum	0.16	This Study	Huntington
Cyrtomium falcatum	0.19	This Study	Huntington
Cyrtomium falcatum	0.59	This Study	Huntington
Cyrtomium falcatum	2.51	This Study	Huntington
Cyrtomium fortunei	0.49	Ma & Takahashi (2002)	
Dactylis glomerata	1.20	Cornelissen & Thompson (1997)	

Danthonia intermedia	3.08	Bezeau et al. (1996)	
Danthonia parryi	2.61	Bezeau et al. (1996)	
Daphne odora	0.19	Ma & Takahashi (2002)	
Daphne odora	0.24	Nakanishi et al. (2003)	
Davallia fejeensis	0.07	This Study	Huntington
Davallia mariesii	0.98	Ma & Takahashi (2002)	
Dennstaedtia scabra	4.07	Ma & Takahashi (2002)	
Deschampsia cespitosa	1.85	Bezeau et al. (1996)	
Deschampsia cespitosa	2.55	Höhne (1963)	
Deschampsia cespitosa	2.89	Höhne (1963)	
Deschampsia cespitosa	2.72	Höhne (1963)	
Deschampsia cespitosa	3.15	Höhne (1963)	
Deschampsia cespitosa	1.86	Johnston et al. (1967)	
Deschampsia cespitosa	4.53	Johnston et al. (1967)	
Deschampsia flexuosa	0.68	Cornelissen & Thompson (1997)	
Deschampsia flexuosa	0.98	Höhne (1963)	
Deschampsia flexuosa	1.41	Höhne (1963)	
Deschampsia flexuosa	1.78	Höhne (1963)	
Deschampsia flexuosa	0.51	Höhne (1963)	
Deschampsia flexuosa	0.45	Höhne (1963)	
Deschampsia flexuosa	1.95	Höhne (1963)	
Deschampsia flexuosa	1.43	Höhne (1963)	
Desmodium uncinatum	0.30	McManus et al. (1977)	
Dianthus superbus	0.13	Ma & Takahashi (2002)	
Digitaria decumbens	2.20	McManus et al. (1977)	
Digitaria ischaemum	8.09	Butler & Hodges (1967)	
Dioon spinulosum	0.06	This Study	Pasadena, CA
Diplazium hachijoense	9.39	Ma & Takahashi (2002)	
Diplazium wichurae	3.57	Ma & Takahashi (2002)	
Distichlis spicata	3.31	Lanning & Eleuterius (1981)	
Distichlis spicata	3.85	Lanning & Eleuterius (1983)	
Dryopteris bissetiana	0.13	Ma & Takahashi (2002)	
Dryopteris carthusiana	0.24	Höhne & Richter (1981)	

Dryopteris carthusiana	0.98	Höhne & Richter (1981)
Dryopteris crassirhizoma	0.30	Ma & Takahashi (2002)
Dryopteris erythrosora	0.45	Ma & Takahashi (2002)
Dryopteris filix-mas	0.13	Höhne (1963)
Dryopteris filix-mas	0.15	Höhne (1963)
Dryopteris filix-mas	0.11	Höhne (1963)
Dryopteris filix-mas	0.06	Höhne (1963)
Dryopteris filix-mas	0.13	Höhne (1963)
Dryopteris filix-mas	0.30	Höhne (1963)
Dryopteris filix-mas	0.64	Höhne & Richter (1981)
Dryopteris filix-mas	0.51	Höhne & Richter (1981)
Dryopteris lacera	0.41	Ma & Takahashi (2002)
Dryopteris sieboldii	0.58	Ma & Takahashi (2002)
Dryopteris uniformis	0.51	Ma & Takahashi (2002)
Ecballium elaterium	1.22	Ma & Takahashi (2002)
Echinochloa colona	0.62	Cowgill (1989)
Echinochloa crus-galli	2.22	Lanning & Eleuterius (1983)
Echinochloa crus-galli	3.65	Lanning & Eleuterius (1987)
Echium angustifolium	1.13	Cowgill (1989)
Eclipta erecta	0.46	Cowgill (1989)
Ehrharta erecta	1.82	Barbehenn (1993)
Elaeagnus multiflora	0.34	Ma & Takahashi (2002)
Eleocharis cellulosa	2.67	Lanning & Eleuterius (1983)
Eleocharis parvula	6.13	Lanning & Eleuterius (1983)
Eleocharis uniglumis	14.98	Tyler (1971)
Elymus cinereus	2.10	Bezeau et al. (1996)
Elymus elymoides	1.60	Blank et al. (1994)
Elymus elymoides	3.20	Blank et al. (1994)
Elymus innovatus	2.13	Bezeau et al. (1996)
Elymus junceus	2.37	Bezeau et al. (1996)
Elymus junceus	2.32	Bezeau et al. (1996)
Elymus mollis	2.03	Ma & Takahashi (2002)
Elymus virginicus	4.48	Lanning & Eleuterius (1985)

Elytrigia atherica	3.42	de Bakker *et al.* (1999)	
Empetrum nigrum	0.19	Carnelli et al. (2001)	
Encephalartos arenarius	0.54	This Study	Loran Whitelock
Encephalartos horridus	0.18	This Study	Huntington
Encephalartos lebomboensis	0.21	This Study	Loran Whitelock
Encephalartos longifolius	0.41	This Study	Loran Whitelock
Encephalartos natalensis	0.12	This Study	Huntington
Encephalartos sp.	0.31	This Study	Loran Whitelock
Encephalartos sp.	0.38	This Study	Loran Whitelock
Encephalartos villosus	0.42	This Study	Huntington
Encephalartos vondola	0.06	This Study	Loran Whitelock
Ephedra californicum	0.36	This Study	RSABG
Ephedra nevadensis	0.07	This Study	RSABG
Ephedra sinica	0.04	Ma & Takahashi (2002)	
Ephedra viridis	0.08	This Study	RSABG
Ephedra viridis	0.12	This Study	RSABG
Epilobium hirsutum	0.21	Cowgill (1989)	
Epimedium grandiflorum	0.83	Ma & Takahashi (2002)	
Equisetum arvense	9.09	Ma & Takahashi (2002)	
Equisetum arvense	12.84	Ma & Takahashi (2002)	
Equisetum hyemale	5.31	Ma & Takahashi (2002)	
Equisetum hyemale	12.02	Ma & Takahashi (2002)	
Equisetum hyemale	0.30	This Study	Pasadena, CA
Equisetum hyemale	7.39	This Study	Huntington
Equisetum hyemale	10.03	This Study	Huntington
Equisetum hyemale	12.38	This Study	Huntington
Equisetum hyemale	17.94	This Study	Huntington
Eragrostis curvula	1.20	Bilbro et al. (1991)	
Eranthis giganteus	3.55	Lanning & Eleuterius (1985)	
Erica canaliculata	0.34	Ma & Takahashi (2002)	
Erigeron crispus	0.11	Cowgill (1989)	
Eugenia uniflora	0.18	Pereire & Felcman (1998)	
Euonymus japonicus	0.39	Nakanishi et al. (2003)	

Eupatorium fortunei	0.79	Ma & Takahashi (2002)
Eurotia lanata	0.60	Bezeau et al. (1996)
Eurya japonica	0.33	Nakanishi et al. (2003)
Fagus sylvatica	16.05	Bartoli & Beaucire (1976)
Fagus sylvatica	23.53	Bartoli & Beaucire (1976)
Fagus sylvatica	1.35	Genßler (unpub)
Fallugia paradoxa	0.40	Bilbro *et al.* (1991)
Ferula varia	2.85	Kolesnikov & Gins (2001)
Festuca arundinacea	7.10	Butler & Hodges (1967)
Festuca arundinacea	0.48	Cowgill (1989)
Festuca arundinacea	3.54	Pahkala & Pihala (2000)
Festuca arundinacea	2.25	Saijonkari-Pahkala (2001)
Festuca arundinacea	2.39	Street (1974)
Festuca arundinacea	4.14	Street (1974)
Festuca arundinacea	0.23	Street (1974)
Festuca arundinacea	2.39	Street (1974)
Festuca arundinacea	4.14	Street (1974)
Festuca arundinacea	2.29	Tirtapradja (1971)
Festuca arundinacea	1.86	Tirtapradja (1971)
Festuca gigantea	5.13	Höhne (1963)
Festuca gigantea	5.71	Höhne (1963)
Festuca gigantea	3.96	Höhne (1963)
Festuca halleri	3.30	Carnelli et al. (2001)
Festuca idahoensis	3.59	Bezeau et al. (1996)
Festuca ovina	0.98	Cornelissen & Thompson (1997)
Festuca ovina	1.11	Cornelissen & Thompson (1997)
Festuca pratensis	2.63	Pahkala & Pihala (2000)
Festuca pratensis	2.04	Saijonkari-Pahkala (2001)
Festuca pratensis	1.59	Tirtapradja (1971)
Festuca pratensis	1.88	Tirtapradja (1971)
Festuca puccinellii	2.96	Carnelli et al. (2001)
Festuca rubra	1.54	Barbehenn (1993)
Festuca rubra	2.82	Bezeau et al. (1996)

Festuca rubra	3.17	Butler & Hodges (1967)	
Festuca rubra	1.63	de Bakker et al. (1999)	
Festuca rubra	7.06	Gadallah & Jefferies (1995)	
Festuca scabrella	3.15	Bezeau et al. (1996)	
Festuca scabrella	1.05	Johnston et al. (1967)	
Festuca scabrella	3.04	Johnston et al. (1967)	
Festuca scabriculmis	2.52	Carnelli et al. (2001)	
Festuca sylvatica	8.02	Bartoli & Beaucire (1976)	
Ficus lyrata	3.04	Lanning (1966)	
Fimbristylis spadicea	1.39	Lanning & Eleuterius (1983)	
Fimbristylis spadicea	3.27	Lanning & Eleuterius (1983)	
Foeniculum vulgare	0.14	Cowgill (1989)	
Fraxinus americana	0.42	Geis (1973)	
Fraxinus americana	0.90	Wilding & Drees (1971)	
Fraxinus oxyphylla	0.27	Cowgill (1989)	
Galega orientalis	0.34	Pahkala & Pihala (2000)	
Galega orientalis	0.27	Saijonkari-Pahkala (2001)	
Galium elongatum	0.79	Cowgill (1989)	
Galium mollugo	1.30	Hogenbirk & Sarrazin-Delay (1995)	
Gardenia jasminoides	0.33	Nakanishi et al. (2003)	
Gentiana decumbens	1.69	Kolesnikov & Gins (2001)	
Ginkgo biloba	0.11	Ma & Takahashi (2002)	
Ginkgo biloba	0.16	This Study	Pasadena, CA
Ginkgo biloba	0.39	This Study	Pasadena, CA
Ginkgo biloba	0.42	This Study	Pasadena, CA
Glaux maritima	3.00	Tyler (1971)	
Gleditsia triacanthos	0.01	Geis (1973)	
Gleichenia glauca	1.69	Ma & Takahashi (2002)	
Glinus lotoides	0.16	Cowgill (1989)	
Glycine max	1.43	Ellis et al. (1995)	
Glycine max	2.95	Ellis et al. (1995)	
Glycine max	0.02	Van der Vorm (1980)	
Glycine max	0.04	Van der Vorm (1980)	

Glycine wightii	0.20	McManus et al. (1977)	
Gnetum gnemon	0.23	This Study	Huntington
Gnetum gnemon	0.43	This Study	Huntington
Gnetum gnemon	0.64	This Study	Huntington
Gnetum gnemon	1.83	This Study	Huntington
Gossypium hirsutum	0.18	Cooper *et al.* (1948)	
Grindelia squarrosa	0.45	Lanning & Eleuterius (1989)	
Guarea macrophylla	0.20	Pereire & Felcman (1998)	
Gymnocarpium dryopteris	4.84	Höhne & Richter (1981)	
Gymnocladus dioicus	0.22	Geis (1973)	
Halodule beaudettei	1.53	Lanning & Eleuterius (1985)	
Hedysarum americanum	0.43	Bezeau et al. (1996)	
Helianthus angustifolius	1.32	Lanning & Eleuterius (1989)	
Helianthus annuus	0.05	Van der Vorm (1980)	
Helianthus annuus	0.08	Van der Vorm (1980)	
Helianthus atrorubens	1.15	Lanning & Eleuterius (1989)	
Helianthus maximilianii	5.00	Bilbro et al. (1991)	
Helianthus tuberosus	4.80	Lanning & Eleuterius (1989)	
Helictotrichon pratense	1.48	Cornelissen & Thompson (1997)	
Helictotrichon pratense	1.75	Cornelissen & Thompson (1997)	
Heliotropium supinum	0.44	Cowgill (1989)	
Heloniopsis orientalis	0.24	Ma & Takahashi (2002)	
Hemerocallis fulva	0.39	Ma & Takahashi (2002)	
Hibiscus cannabinus	0.30	Bilbro et al. (1991)	
Hibiscus moscheutos	1.61	Lanning & Eleuterius (1985)	
Hibiscus sabdariffa	0.70	Bilbro et al. (1991)	
Hibiscus syriacus	0.39	Ma & Takahashi (2002)	
Hilaria jamesii	5.18	Smith et al. (1971)	
Hilaria rigida	4.43	Wallace et al. (1976)	
Hilaria rigida	1.99	Wallace et al. (1976)	
Hippophae rhamnoides	5.01	Kolesnikov & Gins (2001)	
Hirschfeldia incana	0.39	Cowgill (1989)	
Holcus lanatus	1.75	Cornelissen & Thompson (1997)	

Hordeum vulgare	13.30	Bilbro et al. (1991)
Hordeum vulgare	0.67	Grosse-Brauckmann (1953)
Hordeum vulgare	1.95	Grosse-Brauckmann (1953)
Hordeum vulgare	3.40	McManus et al. (1977)
Hordeum vulgare	6.13	Saijonkari-Pahkala (2001)
Hordeum vulgare	3.10	Wallace (1989)
Hordeum vulgare	2.37	Wallace (1989)
Hordeum vulgare	2.65	Wallace (1989)
Hordeum vulgare	2.27	Wallace (1989)
Hordeum vulgare	3.08	Wallace (1989)
Hordeum vulgare	6.97	Wallace (1989)
Hordeum vulgare	4.28	Wallace (1989)
Hordeum vulgare	3.81	Wallace (1989)
Hordeum vulgare	8.02	Wallace (1989)
Hordeum vulgare	2.65	Wallace (1989)
Hordeum vulgare	2.78	Wallace (1989)
Hordeum vulgare	2.48	Wallace (1989)
Hordeum vulgare	3.85	Wallace et al. (1976)
Hordeum vulgare	1.78	Wallace et al. (1976)
Hosta longissima	0.19	Ma & Takahashi (2002)
Houttuynia cordata	2.46	Ma & Takahashi (2002)
Hybanthus glutinosus	0.58	Ma & Takahashi (2002)
Hydrangea macrophylla	0.73	Ma & Takahashi (2002)
Hydrangea macrophylla	0.97	Nakanishi et al. (2003)
Hydrocotyle bonariensis	0.08	Lanning & Eleuterius (1983)
Idesia polycarpa	0.54	Nakanishi et al. (2003)
Ilex aquifolium	0.21	Ma & Takahashi (2002)
Ilex integra	0.45	Nakanishi et al. (2003)
Ilex latifolia	0.32	Nakanishi et al. (2003)
Imperata cylindrica	1.34	Lanning & Eleuterius (1989)
Inula graveolens	0.46	Cowgill (1989)
Inula helenium	2.20	Kolesnikov & Gins (2001)
Inula viscosa	0.26	Cowgill (1989)

Ipomoea sagittata	0.47	Lanning & Eleuterius (1983)	
Iris ensata	0.32	Ma & Takahashi (2002)	
Iris florentina	0.17	Ma & Takahashi (2002)	
Iris setosa	0.39	Ma & Takahashi (2002)	
Iva frutescens	1.85	Lanning & Eleuterius (1983)	
Juglans cinerea	0.48	Geis (1973)	
Juglans nigra	0.28	Geis (1973)	
Juncus effusus	1.28	Ma & Takahashi (2002)	
Juncus gerardii	3.00	Tyler (1971)	
Juncus polycephalus	0.19	Lanning & Eleuterius (1985)	
Juncus roemerianus	0.45	Lanning & Eleuterius (1983)	
Juncus roemerianus	0.34	Lanning & Eleuterius (1983)	
Juniperus communis	0.04	Hodson et al. (1997)	
Juniperus nana	0.08	Carnelli et al. (2001)	
Juniperus sp.	0.25	This Study	Caltech
Juniperus sp.	0.42	This Study	Pasadena, CA
Juniperus virginiana	0.10	Lanning & Eleuterius (1983)	
Kalanchoe braziliensis	0.07	Pereire & Felcman (1998)	
Kerria japonica	0.71	Ma & Takahashi (2002)	
Kickxia spuria	0.51	Cowgill (1989)	
Kochia scoparia	1.50	Bilbro et al. (1991)	
Koeleria cristata	2.04	Bezeau et al. (1996)	
Kosteletzkya virginica	1.33	Lanning & Eleuterius (1983)	
Lactuca serriola	0.50	Cowgill (1989)	
Larix decidua	1.09	Carnelli et al. (2001)	
Larix decidua	2.21	Hodson et al. (1997)	
Larix decidua	1.37	Klein & Geis (1978)	
Larix laricina	0.24	Klein & Geis (1978)	
Lastrea limbosperma	1.84	Höhne & Richter (1981)	
Lastrea limbosperma	2.74	Höhne & Richter (1981)	
Lastrea oligophlebia	2.61	Ma & Takahashi (2002)	
Lathyrus ochroleucus	0.77	Bezeau et al. (1996)	
Lavandula angustifolia	1.16	Ma & Takahashi (2002)	

Leontodon hispidus	0.41	Cornelissen & Thompson (1997)
Leptogramma mollissima	6.42	Ma & Takahashi (2002)
Ligustrum japonicum	0.48	Nakanishi et al. (2003)
Ligustrum lucidum	0.43	Nakanishi et al. (2003)
Lilaeopsis chinensis	4.12	Lanning & Eleuterius (1983)
Lilium leichtlinii	0.24	Ma & Takahashi (2002)
Limonium carolinianum	0.44	Lanning & Eleuterius (1981)
Limonium vulgare	0.00	de Bakker et al. (1999)
Lindera benzoin	0.08	Geis (1973)
Lindera strychnifolia	0.19	Ma & Takahashi (2002)
Linum usitatissimum	0.10	Saijonkari-Pahkala (2001)
Liriodendron tulipifera	1.53	Nakanishi *et al.* (2003)
Lobelia cardinalis	0.04	Lanning & Eleuterius (1983)
Loiseleuria procumbens	0.32	Carnelli et al. (2001)
Lolium perenne	13.20	Butler & Hodges (1967)
Lolium perenne	0.93	Reay & Bennett (1987)
Lolium rigidum	1.58	Grosse-Brauckmann (1953)
Lolium rigidum	1.23	Grosse-Brauckmann (1953)
Lolium rigidum	1.40	Grosse-Brauckmann (1953)
Lolium rigidum	2.34	Jones & Handreck (1967)
Lotus corniculatus	0.11	Cornelissen & Thompson (1997)
Lotus corniculatus	0.60	Hogenbirk & Sarrazin-Delay (1995)
Loxogramme saziran	0.26	Ma & Takahashi (2002)
Ludwigia stolonifera	0.13	Cowgill (1989)
Luffa acutangula	1.11	Ma & Takahashi (2002)
Lupinus argenteus	0.86	Bezeau et al. (1996)
Lupinus nanus	0.24	Grosse-Brauckmann (1953)
Lupinus nanus	0.16	Grosse-Brauckmann (1953)
Luzula luzuloides	0.94	Höhne (1963)
Luzula luzuloides	0.43	Höhne (1963)
Luzula luzuloides	1.16	Höhne (1963)
Lycopersicon esculentum	0.17	Wallace (1989)
Lycopersicon esculentum	0.15	Wallace (1989)

Lycopodium carolinianum	1.20	Lanning & Eleuterius (1983)	
Lycopodium clavatum	1.37	Ma & Takahashi (2002)	
Lycopodium lucidulum	0.81	This Study	Commercial
Lycopodium lucidulum	0.91	This Study	Commercial
Lycopus europaeus	0.22	Cowgill (1989)	
Lycoris radiata	0.02	Ma & Takahashi (2002)	
Lycurus phleoides	4.16	Smith et al. (1971)	
Lygodium japonicum	2.57	Ma & Takahashi (2002)	
Lyonia ferruginea	0.29	Kalisz & Stone (1984)	
Lythrum lineare	0.69	Lanning & Eleuterius (1983)	
Lythrum salicaria	0.33	Cowgill (1989)	
Maclura pomifera	1.20	Geis (1973)	
Magnolia grandiflora	1.24	Lanning & Eleuterius (1983)	
Magnolia grandiflora	1.43	Ma & Takahashi (2002)	
Magnolia hypoleuca	1.38	Nakanishi et al. (2003)	
Magnolia kobus	0.95	Nakanishi et al. (2003)	
Mallotus japonicus	1.03	Ma & Takahashi (2002)	
Manisuris rugosa	6.75	Lanning & Eleuterius (1983)	
Marchantia polymorpha	11.87	Ma & Takahashi (2002)	
Matteuccia struthiopteris	4.69	Höhne & Richter (1981)	
Matteuccia struthiopteris	4.02	Höhne & Richter (1981)	
Medicago sativa	0.08	Bertrand & Ghitescu (1934)	
Medicago sativa	0.20	McManus et al. (1977)	
Medicago sativa	0.38	Saijonkari-Pahkala (2001)	
Megalodonta tripartita	0.35	Cowgill (1989)	
Melampyrum pratense	1.28	Höhne (1963)	
Melampyrum pratense	1.35	Höhne (1963)	
Melampyrum pratense	1.39	Höhne (1963)	
Melastoma candidum	0.39	Ma & Takahashi (2002)	
Melia azedarach	0.58	Ma & Takahashi (2002)	
Melica uniflora	4.04	Höhne (1963)	
Melilotus albus	0.31	Cowgill (1989)	
Melinis minutiflora	2.20	McManus *et al.* (1977)	

Melissa officinalis	0.68	Cowgill (1989)	
Melissa officinalis	3.44	Kolesnikov & Gins (2001)	
Mentha longifolia	0.36	Cowgill (1989)	
Mentha piperita	3.29	Kolesnikov & Gins (2001)	
Mercurialis perennis	0.24	Höhne (1963)	
Mercurialis perennis	0.24	Höhne (1963)	
Metasequoia glyptostroboides	0.09	This Study	Huntington
Metasequoia glyptostroboides	0.14	This Study	Huntington
Miscanthus sinensis	6.33	Ma & Takahashi (2002)	
Molinia caerulea	2.74	Höhne (1963)	
Molinia caerulea	3.08	Höhne (1963)	
Molinia caerulea	1.01	Höhne (1963)	
Morus alba	1.39	Ma & Takahashi (2002)	
Morus rubra	3.79	Geis (1973)	
Morus rubra	3.12	Lanning & Eleuterius (1985)	
Muhlenbergia richardsonis	7.34	Smith et al. (1971)	
Musa basjoo	2.31	Ma & Takahashi (2002)	
Myrica cerifera	0.03	Lanning & Eleuterius (1983)	
Myrica cerifera	0.09	Lanning & Eleuterius (1983)	
Nandina domestica	0.41	Ma & Takahashi (2002)	
Nandina domestica	0.15	Nakanishi et al. (2003)	
Nardus stricta	2.67	Carnelli et al. (2001)	
Nasturtium officinale	2.46	Kolesnikov & Gins (2001)	
Nephrolepis cordifolia	0.58	Ma & Takahashi (2002)	
Nephrolepis cordifolia	0.21	This Study	Commercial
Nephrolepis cordifolia	1.64	This Study	Commercial
Nephrolepis exaltata	1.59	This Study	Commercial
Nephrolepsis exaltata	0.11	This Study	Commercial
Nephrolepsis exaltata	0.11	This Study	Commercial
Nephrolepsis exaltata	0.32	This Study	Commercial
Nerium oleander	0.41	Cowgill (1989)	
Nerium oleander	0.39	Ma & Takahashi (2002)	
Nuphar lutea	0.27	Cowgill (1989)	

Oenothera lamarckiana	0.17	Ma & Takahashi (2002)
Olea europaea	0.28	Ma & Takahashi (2002)
Onoclea sensibilis	3.29	Ma & Takahashi (2002)
Origanum vulgare	1.58	Kolesnikov & Gins (2001)
Oryza sativa	13.48	Ma & Takahashi (2002)
Oryza sativa	16.50	McManus et al. (1977)
Oryza sativa	0.95	Van der Vorm (1980)
Oryza sativa	1.24	Van der Vorm (1980)
Oryza sativa	4.84	Wallace (1989)
Oryza sativa	3.74	Wallace (1989)
Oryzopsis asperifolia	3.50	Hogenbirk & Sarrazin-Delay (1995)
Osmanthus fragrans	0.46	Nakanishi et al. (2003)
Osmunda cinnamomea	2.25	Höhne & Richter (1981)
Osmunda gracilis	2.55	Ma & Takahashi (2002)
Osmunda japonica	11.96	Ma & Takahashi (2002)
Osmunda lancea	5.18	Lanning & Eleuterius (1983)
Osmunda regalis	1.87	Höhne & Richter (1981)
Ostrya virginiana	0.31	Geis (1973)
Panax ginseng	0.43	Ma & Takahashi (2002)
Panicum amarum	0.76	Lanning & Eleuterius (1983)
Panicum commutatum	7.95	Lanning & Eleuterius (1989)
Panicum maximum	2.40	McManus et al. (1977)
Panicum obtusum	4.41	Smith et al. (1971)
Panicum repens	1.17	Lanning & Eleuterius (1983)
Panicum repens	1.14	Lanning & Eleuterius (1989)
Panicum texanum	16.60	Bilbro et al. (1991)
Panicum virgatum	2.70	Bilbro et al. (1991)
Panicum virgatum	9.44	Geis (1978)
Panicum virgatum	2.43	Lanning & Eleuterius (1983)
Panicum virgatum	5.04	Lanning & Eleuterius (1987)
Papaver bracteatum	1.60	Ma & Takahashi (2002)
Papaver rhoeas	1.24	Ma & Takahashi (2002)
Paspalum dilatatum	3.68	Barbehenn (1993)

Paspalum urvillei	4.84	Lanning & Eleuterius (1987)	
Paspalum vaginatum	0.28	Cowgill (1989)	
Paspalum wettsteinii	2.50	McManus et al. (1977)	
Paulownia tomentosa	0.98	Nakanishi et al. (2003)	
Pelargonium graveolens	0.64	Ma & Takahashi (2002)	
Pellaea rotundifolia	1.65	This Study	Huntington
Pellaea rotundifolia	12.34	This Study	Huntington
Pennisetum clandestinum	2.12	Barbehenn (1993)	
Pennisetum clandestinum	2.09	Pereire & Felcman (1998)	
Persea palustris	3.40	Lanning & Eleuterius (1983)	
Phalaris arundinacea	5.73	Pahkala & Pihala (2000)	
Phalaris tuberosa	4.80	McManus et al. (1977)	
Phaseolus atropurpureus	0.30	McManus et al. (1977)	
Phaseolus vulgaris	1.82	Wallace et al. (1976)	
Phaseolus vulgaris	0.92	Wallace et al. (1976)	
Phegopteris connectilis	5.88	Höhne & Richter (1981)	
Phellodendron amurense	0.86	Ma & Takahashi (2002)	
Philadelphus satsumi	0.11	Ma & Takahashi (2002)	
Phleum pratense	1.59	Bezeau et al. (1996)	
Phleum pratense	4.00	Hogenbirk & Sarrazin-Delay (1995)	
Phleum pratense	1.60	Saijonkari-Pahkala (2001)	
Phlox subulata	2.37	Ma & Takahashi (2002)	
Phoenix dactylifera	0.58	Ma & Takahashi (2002)	
Phoenix roebelenii	0.64	Ma & Takahashi (2002)	
Phragmites australis	0.90	Cowgill (1989)	
Phragmites communis	5.26	Lanning & Eleuterius (1985)	
Physalis alkekengi	0.21	Ma & Takahashi (2002)	
Picea abies	0.85	Carnelli et al. (2001)	
Picea abies	1.11	Genßler (unpub.)	
Picea abies	2.57	Hodson et al. (1997)	
Picea glauca	0.76	Hodson & Sangster (1998, 2002)	
Picea glauca	1.31	Hodson & Sangster (1998, 2002)	
Picea glauca	1.05	Klein & Geis (1978)	

Picea mariana	0.17	Klein & Geis (1978)
Picea orientalis	2.16	Hodson et al. (1997)
Picea rubens	0.43	Klein & Geis (1978)
Picris echioides	0.32	Cowgill (1989)
Pieris japonica	0.20	Nakanishi et al. (2003)
Pinus armandii	0.41	Hodson et al. (1997)
Pinus banksiana	0.18	Klein & Geis (1978)
Pinus cembra	0.13	Carnelli et al. (2001)
Pinus clausa	0.43	Kalisz & Stone (1984)
Pinus contorta	0.11	Hodson et al. (1997)
Pinus cooperi	0.73	Hodson et al. (1997)
Pinus flexilis	0.13	Hodson et al. (1997)
Pinus jeffreyi	0.03	Hodson et al. (1997)
Pinus koraiensis	0.42	Hodson et al. (1997)
Pinus luchuensis	0.17	Ma & Takahashi (2002)
Pinus mugo	0.10	Carnelli et al. (2001)
Pinus palustris	1.09	Kalisz & Stone (1984)
Pinus palustris	0.71	Ma & Takahashi (2002)
Pinus parviflora	0.24	Hodson et al. (1997)
Pinus peuce	0.09	Hodson et al. (1997)
Pinus pinea	0.01	Hodson et al. (1997)
Pinus resinosa	0.08	Klein & Geis (1978)
Pinus strobiformis	0.06	Hodson et al. (1997)
Pinus strobus	0.17	Hodson & Sangster (1998)
Pinus strobus	0.40	Hodson & Sangster (1998)
Pinus strobus	0.05	Hodson et al. (1997)
Pinus strobus	0.09	Klein & Geis (1978)
Pinus sylvestris	1.07	Bartoli & Beaucire (1976)
Pinus sylvestris	0.49	Bartoli & Beaucire (1976)
Pinus sylvestris	0.17	Genßler (unpub.)
Pinus sylvestris	0.11	Hodson et al. (1997)
Pinus sylvestris	0.18	Klein & Geis (1978)
Pistia stratiotes	0.36	Ma & Takahashi (2002)

Pisum sativum	0.25	Jones & Handreck (1967)	
Pittosporum tobira	0.15	Nakanishi et al. (2003)	
Plantago lagopus	0.56	Cowgill (1989)	
Plantago lanceolata	0.13	Cornelissen & Thompson (1997)	
Plantago maritima	0.26	de Bakker et al. (1999)	
Platanus occidentalis	0.42	Geis (1973)	
Plectranthus japonicus	0.15	Ma & Takahashi (2002)	
Pleioblastus chino	11.06	Ma & Takahashi (2002)	
Pluchea purpurascens	0.41	Lanning & Eleuterius (1983)	
Poa chaixii	0.75	Höhne (1963)	
Poa chaixii	0.94	Höhne (1963)	
Poa compressa	2.50	Hogenbirk & Sarrazin-Delay (1995)	
Poa pratensis	1.98	Butler & Hodges (1967)	
Poa pratensis	0.28	Street (1974)	
Poa pratensis	3.63	Street (1974)	
Poa pratensis	3.66	Street (1974)	
Poa pratensis	6.10	Street (1974)	
Poa pratensis	6.10	Street (1974)	
Poa pratensis	3.48	Taber et al. (2002)	
Poa secunda	2.63	Bezeau et al. (1996)	
Podocarpus gracilior	0.10	This Study	Huntington
Podocarpus gracilior	0.34	This Study	Huntington
Podocarpus gracilior	0.44	This Study	Huntington
Podocarpus nagi	0.10	This Study	Huntington
Podocarpus nagi	0.15	This Study	Huntington
Podocarpus nagi	0.17	This Study	Huntington
Podocarpus neriifolius	0.12	Lanning (1966)	
Polygonatum odoratum	0.19	Ma & Takahashi (2002)	
Polygonum acuminatum	0.39	Cowgill (1989)	
Polygonum arenastrum	0.12	Cowgill (1989)	
Polygonum aviculare	4.21	Kolesnikov & Gins (2001)	
Polygonum fagopyrum	0.04	Bertrand & Ghitescu (1934)	
Polygonum hydropiper	0.32	Ma & Takahashi (2002)	

Polygonum lapathifolium	0.20	Cowgill (1989)	
Polygonum patulum	0.12	Cowgill (1989)	
Polygonum punctatum	2.15	Lanning & Eleuterius (1985)	
Polygonum salicifolium	0.10	Cowgill (1989)	
Polygonum senegalense	0.13	Cowgill (1989)	
Polymnia uvedalia	0.98	Lanning & Eleuterius (1983)	
Polymnia uvedalia	1.48	Lanning & Eleuterius (1987)	
Polypodium vulgare	0.26	Höhne & Richter (1981)	
Polypodium vulgare	0.83	Höhne & Richter (1981)	
Polyscias filicifolia	0.11	Lanning (1966)	
Polystichopsis amabilis	0.71	Ma & Takahashi (2002)	
Polystichopsis pseudo-aristata	0.62	Ma & Takahashi (2002)	
Polystichopsis standishii	0.45	Ma & Takahashi (2002)	
Polystichum lepidocaulon	0.13	Ma & Takahashi (2002)	
Polystichum polyblepharum	0.15	Ma & Takahashi (2002)	
Polystichum pseudo-makinoi	0.79	Ma & Takahashi (2002)	
Polystichum tripteron	0.41	Ma & Takahashi (2002)	
Polystichum tsussimense	2.48	This Study	Huntington
Polystichum tsussimense	11.96	This Study	Huntington
Poncirus trifoliata	1.05	Ma & Takahashi (2002)	
Populus deltoides	0.94	Geis (1973)	
Populus euphratica	0.34	Cowgill (1989)	
Populus sieboldii	0.55	Fu et al. (2001)	
Populus tremuloides	0.11	Bezeau et al. (1996)	
Potentilla erecta	2.50	Kolesnikov & Gins (2001)	
Potentilla fruticosa	0.15	Bezeau et al. (1996)	
Primula veris	2.08	Kolesnikov & Gins (2001)	
Prunus serotina	0.04	Geis (1973)	
Prunus virginiana	0.52	Geis (1973)	
Pseudolarix amabilis	0.06	Hodson et al. (1997)	
Pseudotsuga flauhauti	0.38	Hodson et al. (1997)	
Pseudotsuga macrolepis	1.93	Hodson et al. (1997)	
Pseudotsuga menziesii	0.74	Hodson et al. (1997)	

Pseudotsuga menziesii	0.37	Klein & Geis (1978)	
Psilotum nudum	0.19	This Study	Huntington
Psilotum nudum	0.19	This Study	Huntington
Pteridium aquilinum	1.24	Höhne (1963)	
Pteridium aquilinum	1.11	Höhne (1963)	
Pteridium aquilinum	1.67	Höhne (1963)	
Pteridium aquilinum	1.95	Höhne (1963)	
Pteridium aquilinum	2.44	Höhne & Richter (1981)	
Pteridium aquilinum	3.89	Höhne & Richter (1981)	
Pteridium aquilinum	10.61	Ma & Takahashi (2002)	
Pteris ensiformis	3.49	Ma & Takahashi (2002)	
Puccinellia maritima	0.53	de Bakker et al. (1999)	
Puccinellia phryganodes	0.74	Gadallah & Jefferies (1995)	
Pulicaria dysenterica	0.13	Cowgill (1989)	
Pulsatilla multifida	2.01	Kolesnikov & Gins (2001)	
Pyracantha crenulata	0.19	Ma & Takahashi (2002)	
Pyrrosia lingua	0.11	Ma & Takahashi (2002)	
Pyrrosia lingua	0.09	Ma & Takahashi (2002)	
Quercus alba	0.90	Geis (1973)	
Quercus chapmanii	0.38	Kalisz & Stone (1984)	
Quercus geminata	1.33	Kalisz & Stone (1984)	
Quercus imbricaria	0.38	Geis (1973)	
Quercus laevis	0.39	Kalisz & Stone (1984)	
Quercus macrocarpa	0.44	Geis (1973)	
Quercus muehlenbergii	0.58	Geis (1973)	
Quercus myrtifolia	0.44	Kalisz & Stone (1984)	
Quercus petraea	0.63	Genßler (unpub.)	
Quercus robur	1.84	Bartoli & Beaucire (1976)	
Quercus robur	0.62	Genßler (unpub.)	
Quercus rubra	0.15	Geis (1973)	
Quercus suber	0.73	Ma & Takahashi (2002)	
Quercus velutina	0.13	Geis (1973)	
Ranunculus japonicus	0.77	Ma & Takahashi (2002)	

Renealmia petasites	0.91	Pereire & Felcman (1998)
Rhapis humilis	1.11	Ma & Takahashi (2002)
Rhodiola linearifolia	6.29	Kolesnikov & Gins (2001)
Rhododendron ferrugineum	0.04	Carnelli et al. (2001)
Rhododendron japonicum	0.90	Ma & Takahashi (2002)
Rhododendron pulchrum	0.39	Nakanishi et al. (2003)
Rhynchospora plumosa	3.85	Lanning & Eleuterius (1989)
Rohdea japonica	0.53	Ma & Takahashi (2002)
Rosa woodsii	0.51	Bezeau et al. (1996)
Rubia tinctorum	1.07	Ma & Takahashi (2002)
Rubus idaeus	0.11	Höhne (1963)
Rubus idaeus	0.13	Höhne (1963)
Rubus idaeus	0.02	Höhne (1963)
Rubus idaeus	0.02	Höhne (1963)
Rubus idaeus	0.26	Höhne (1963)
Rumex dentatus	0.10	Cowgill (1989)
Ruppia maritima	3.20	Lanning & Eleuterius (1985)
Sabal etonia	3.14	Kalisz & Stone (1984)
Sabal minor	2.11	Lanning & Eleuterius (1983)
Saccharum officinarum	1.75	Lanning & Eleuterius (1985)
Saccharum officinarum	1.65	Ma & Takahashi (2002)
Saccharum officinarum	0.32	Van der Vorm (1980)
Saccharum officinarum	0.90	Van der Vorm (1980)
Sagittaria lancifolia	0.27	Lanning & Eleuterius (1983)
Sagittaria trifolia	0.83	Ma & Takahashi (2002)
Salicornia bigelovii	0.74	Lanning & Eleuterius (1985)
Salicornia europaea	0.00	de Bakker et al. (1999)
Salicornia virginica	0.02	Lanning & Eleuterius (1985)
Salix acmophylla	0.20	Cowgill (1989)
Salix matsudana	0.30	Ma & Takahashi (2002)
Salsola kali	1.60	Bilbro et al. (1991)
Salvia officinalis	1.41	Ma & Takahashi (2002)
Sansevieria trifasciata	0.02	Ma & Takahashi (2002)

Saponaria officinalis	0.58	Ma & Takahashi (2002)	
Sasa nipponica	1.54	Fu *et al.* (2001)	
Sasa nipponica	9.24	Ma & Takahashi (2002)	
Sassafras albidum	0.07	Geis (1973)	
Saururus chinensis	0.60	Ma & Takahashi (2002)	
Schisandra chinensis	3.15	Kolesnikov & Gins (2001)	
Sciadopitys verticillata	0.13	This Study	Huntington
Scirpus americanus	2.84	Lanning & Eleuterius (1985)	
Scirpus cyperinus	3.31	Lanning & Eleuterius (1989)	
Scirpus olneyi	2.55	Lanning & Eleuterius (1983)	
Scirpus robustus	5.23	Lanning & Eleuterius (1983)	
Scirpus tabernaemontani	0.45	Ma & Takahashi (2002)	
Scirpus validus	2.37	Lanning & Eleuterius (1981)	
Secale cereale	0.49	Grosse-Brauckmann (1953)	
Secale cereale	2.41	Jones & Handreck (1967)	
Secale cereale	2.23	Ma & Takahashi (2002)	
Secale cereale	1.40	Robbins et al. (1987)	
Secale cereale	3.61	Saijonkari-Pahkala (2001)	
Securinega suffruticosa	0.41	Ma & Takahashi (2002)	
Sedum hybridum	7.68	Kolesnikov & Gins (2001)	
Selaginella caulescens	1.95	Ma & Takahashi (2002)	
Selaginella caulescens	11.32	Ma & Takahashi (2002)	
Selaginella involvens	5.54	Ma & Takahashi (2002)	
Selaginella involvens	8.37	Ma & Takahashi (2002)	
Selaginella kraussiana	2.57	This Study	Commercial
Selaginella sp.	0.69	This Study	Huntington
Selaginella sp.	1.06	This Study	Huntington
Selaginella sp.	4.14	This Study	Huntington
Selaginella uncinata	3.94	Ma & Takahashi (2002)	
Selaginella uncinata	4.08	This Study	Commercial
Senecio fuchsii	1.24	Höhne (1963)	
Senecio fuchsii	0.98	Höhne (1963)	
Senecio fuchsii	0.53	Höhne (1963)	

Senecio fuchsii	1.52	Höhne (1963)	
Sequoia sempervirens	0.51	Ma & Takahashi (2002)	
Sequoia sp.	0.15	This Study	Huntington
Sequoiadendron giganteum	0.39	Hodson et al. (1997)	
Sequoiadendron sp.	0.27	This Study	Huntington
Serenoa repens	3.67	Kalisz & Stone (1984)	
Serenoa repens	5.24	Lanning & Eleuterius (1985)	
Seriphidium maritimum	0.66	de Bakker et al. (1999)	
Setaria geniculata	3.96	Lanning & Eleuterius (1987)	
Setaria italica	0.55	Bilbro et al. (1991)	
Setaria italica	0.30	Bilbro et al. (1991)	
Setaria magna	6.06	Lanning & Eleuterius (1989)	
Setaria sphacelata	1.00	McManus et al. (1977)	
Smilacina japonica	1.16	Ma & Takahashi (2002)	
Solanum americanum	0.36	Pereire & Felcman (1998)	
Solanum nigrum	0.17	Cowgill (1989)	
Solidago sempervirens	0.22	Lanning & Eleuterius (1983)	
Sonchus oleraceus	0.28	Cowgill (1989)	
Sophora flavescens	0.11	Ma & Takahashi (2002)	
Sophora japonica	0.15	Ma & Takahashi (2002)	
Sorghastrum nutans	5.00	Geis (1978)	
Sorghastrum nutans	7.18	Lanning & Eleuterius (1987)	
Sorghum bicolor	0.57	Bilbro *et al.* (1991)	
Sorghum bicolor	0.98	Bilbro *et al.* (1991)	
Sorghum bicolor	4.26	Ellis et al. (1995)	
Sorghum bicolor	4.69	Ellis et al. (1995)	
Sorghum halepense	1.55	Lanning & Eleuterius (1985)	
Spartina alterniflora	0.83	Lanning & Eleuterius (1981)	
Spartina alterniflora	2.28	Lanning & Eleuterius (1983)	
Spartina anglica	2.61	de Bakker et al. (1999)	
Spartina cynosuroides	2.52	Lanning & Eleuterius (1983)	
Spartina cynosuroides	0.56	Lanning & Eleuterius (1989)	
Spartina patens	0.89	Lanning & Eleuterius (1983)	

Spartina patens	2.19	Lanning & Eleuterius (1983)
Spergularia media	0.00	de Bakker et al. (1999)
Sphagnum cymbifolium	2.93	Ma & Takahashi (2002)
Spiraea thunbergii	0.34	Ma & Takahashi (2002)
Spiranthes sinensis	0.13	Ma & Takahashi (2002)
Sporobolus cryptandrus	3.50	Smith *et al.* (1971)
Staphylea trifolia	0.30	Geis (1973)
Stemona japonica	0.49	Ma & Takahashi (2002)
Stipa comata	0.94	Bezeau et al. (1996)
Stipa comata	0.88	Johnston et al. (1967)
Stipa comata	1.17	Johnston et al. (1967)
Stipa comata	0.89	Johnston et al. (1967)
Stipa comata	1.50	Johnston et al. (1967)
Stipa comata	1.55	Johnston et al. (1967)
Stipa comata	1.08	Johnston et al. (1967)
Stipa comata	2.62	Johnston et al. (1967)
Stipa comata	2.54	Johnston et al. (1967)
Stipa comata	1.13	Johnston et al. (1967)
Stipa comata	1.83	Johnston et al. (1967)
Stipa comata	2.76	Lanning & Eleuterius (1987)
Stipa richardsonii	2.64	Bezeau et al. (1996)
Stipa spartea	2.86	Bezeau et al. (1996)
Stipa spartea	3.67	Lanning & Eleuterius (1987)
Stipa viridula	3.40	Bezeau et al. (1996)
Struthanthus marginatus	0.27	Pereire & Felcman (1998)
Struthiopteris niponica	6.78	Ma & Takahashi (2002)
Styrax japonicus	0.34	Nakanishi et al. (2003)
Suaeda maritima	0.26	de Bakker et al. (1999)
Symphoricarpos occidentalis	1.05	Bezeau et al. (1996)
Tamarix chinensis	1.16	Ma & Takahashi (2002)
Tamarix jordanis	0.19	Cowgill (1989)
Taxodium distichum	0.08	Hodson et al. (1997)
Taxodium japonicum	0.07	Fu et al. (2001)

Taxus baccata	0.16	Hodson et al. (1997)
Taxus cuspidata	0.51	Hodson et al. (1997)
Ternstroemia japonica	0.32	Nakanishi et al. (2003)
Thea sinensis	0.12	Fu et al. (2001)
Thea sinensis	0.09	Ma & Takahashi (2002)
Thuja orientalis	0.06	Hodson et al. (1997)
Thuja orientalis	0.43	Ma & Takahashi (2002)
Thymus marschallianus	4.73	Kolesnikov & Gins (2001)
Tibouchina pulchra	0.03	Pereire & Felcman (1998)
Tilia americana	0.49	Geis (1973)
Tillandsia usneoides	0.44	Lanning & Eleuterius (1983)
Torreya nucifera	0.24	Ma & Takahashi (2002)
Trachycarpus fortunei	3.02	Ma & Takahashi (2002)
Tradescantia ohiensis	0.83	Ma & Takahashi (2002)
Trichachne californica	2.70	Bilbro et al. (1991)
Tricyrtis hirta	0.51	Ma & Takahashi (2002)
Trifolium fragiferum	0.14	Cowgill (1989)
Trifolium hybridum	0.80	Hogenbirk & Sarrazin-Delay (1995)
Trifolium incarnatum	0.12	Jones & Handreck (1967)
Trifolium pratense	0.12	Grosse-Brauckmann (1953)
Trifolium pratense	0.10	Grosse-Brauckmann (1953)
Trifolium pratense	0.09	Grosse-Brauckmann (1953)
Trifolium pratense	0.70	Hogenbirk & Sarrazin-Delay (1995)
Trifolium pratense	0.30	McManus et al. (1977)
Trifolium pratense	0.31	Saijonkari-Pahkala (2001)
Trifolium repens	1.00	Hogenbirk & Sarrazin-Delay (1995)
Trifolium repens	0.11	Reay & Bennett (1987)
Trifolium subterraneum	3.10	McManus et al. (1977)
Triglochin maritima	0.00	de Bakker et al. (1999)
Triglochin maritima	2.40	Tyler (1971)
Triglochin striata	1.20	Lanning & Eleuterius (1983)
Trilisa odoratissima	0.77	Lanning & Eleuterius (1985)
Tripsacum dactyloides	0.93	Lanning & Eleuterius (1983)

Triticosecale spp.	5.90	Bilbro et al. (1991)	
Triticum aestivum	10.74	Bilbro *et al.* (1991)	
Triticum aestivum	2.16	Cooper et al. (1948)	
Triticum aestivum	0.96	Grosse-Brauckmann (1953)	
Triticum aestivum	3.08	Ma & Takahashi (2002)	
Triticum aestivum	5.60	McManus et al. (1977)	
Triticum aestivum	2.48	Reay & Bennett (1987)	
Triticum aestivum	2.70	Robbins et al. (1987)	
Triticum aestivum	3.52	Saijonkari-Pahkala (2001)	
Triticum aestivum	5.99	Schnug & v. Franck (1985)	
Triticum aestivum	16.05	Schnug & v. Franck (1985)	
Triticum aestivum	24.18	Schnug & v. Franck (1985)	
Triticum aestivum	0.18	Van der Vorm (1980)	
Triticum aestivum	0.41	Van der Vorm (1980)	
Triticum aestivum	2.20	Wallace (1989)	
Triticum aestivum	0.90	Wallace (1989)	
Triticum boeoticum	5.58	Ma & Takahashi (2002)	
Triticum dicoccoides	2.85	Ma & Takahashi (2002)	
Triticum percicumxAegilopssquarrosa	3.64	Ma & Takahashi (2002)	
Tropaeolum majus	0.04	Ma & Takahashi (2002)	
Tsuga canadensis	0.26	Hodson et al. (1997)	
Tsuga canadensis	0.16	Klein & Geis (1978)	
Tsuga caroliniana	0.14	Klein & Geis (1978)	
Tsuga diversifolia	0.81	Hodson et al. (1997)	
Tsuga heterophylla	0.23	Hodson et al. (1997)	
Typha angustata	0.20	Cowgill (1989)	
Typha angustifolia	0.04	Lanning & Eleuterius (1983)	
Ulmus americana	3.30	Geis (1973)	
Ulmus americana	5.10	Lanning (1966)	
Ulmus rubra	2.14	Geis (1973)	
Uniola paniculata	0.38	Lanning & Eleuterius (1983)	
Uniola paniculata	1.58	Lanning & Eleuterius (1985)	
unknown fern	2.68	This Study	Huntington

unknown fern	7.33	This Study	Huntington
Urtica dioica	1.60	Cornelissen & Thompson (1997)	
Urtica dioica	1.13	Höhne (1963)	
Urtica dioica	4.75	Höhne (1963)	
Urtica dioica	3.74	Höhne (1963)	
Urtica dioica	2.82	Höhne (1963)	
Urtica hulensis	0.28	Cowgill (1989)	
Vaccinium myrtillus	2.14	Bartoli & Beaucire (1976)	
Vaccinium myrtillus	0.04	Carnelli et al. (2001)	
Vaccinium myrtillus	0.13	Höhne (1963)	
Vaccinium myrtillus	0.11	Höhne (1963)	
Vaccinium myrtillus	0.15	Höhne (1963)	
Vaccinium myrtillus	0.19	Höhne (1963)	
Vaccinium myrtillus	0.24	Höhne (1963)	
Vaccinium myrtillus	0.30	Höhne (1963)	
Vaccinium uliginosum	0.10	Carnelli et al. (2001)	
Vaccinium vitis-idaea	0.04	Carnelli et al. (2001)	
Valeriana officinalis	2.08	Kolesnikov & Gins (2001)	
Verbena officinalis	0.46	Cowgill (1989)	
Verbena officinalis	1.01	Ma & Takahashi (2002)	
Vicia americana	0.41	Bezeau et al. (1996)	
Vicia villosa	0.02	Fu *et al.* (2001)	
Viola tricolor	0.28	Ma & Takahashi (2002)	
Vitex agnus-castus	0.56	Cowgill (1989)	
Vitis aestivalis	0.34	Lanning & Eleuterius (1983)	
Wasabia japonica	0.30	Ma & Takahashi (2002)	
Washingtonia filifera	3.05	Lanning (1966)	
Welwitschia mirabilis	0.33	This Study	Huntington
Wisteria brachybotrys	0.51	Ma & Takahashi (2002)	
Wollemia nobilis	0.20	This Study	Huntington
Wollemia nobilis	0.36	This Study	Huntington
Woodwardia orientalis	5.03	Ma & Takahashi (2002)	
Xanthium strumarium	0.32	Cowgill (1989)	

Yucca aloifolia	0.03	Lanning & Eleuterius (1983)
Yucca filamentosa	0.17	Ma & Takahashi (2002)
Zamioculcas zamiifolia	0.46	Lanning (1966)
Zanthoxylum americanum	0.44	Geis (1973)
Zanthoxylum piperitum	0.77	Ma & Takahashi (2002)
Zea mays	1.71	Ellis et al. (1995)
Zea mays	2.44	Ellis et al. (1995)
Zea mays	1.82	Wallace (1989)
Zea mays	1.77	Wallace (1989)
Zea mays	1.95	Wallace (1989)
Zea mays	2.17	Wallace (1989)
Zea mays	1.54	Wallace (1989)
Zea mays	1.54	Wallace (1989)
Zea mays	1.27	Wallace (1989)
Zea mays	1.43	Wallace (1989)
Zea mays	1.68	Wallace (1989)
Zea mays	1.30	Wallace (1989)
Zea mays	1.39	Wallace (1989)
Zea mays	1.56	Wallace (1989)
Zephyranthes candida	0.21	Ma & Takahashi (2002)
Zingiber mioga	0.47	Ma & Takahashi (2002)
Zizania aquatica	6.00	Lanning & Eleuterius (1981)
Zizania aquatica	4.15	Lanning & Eleuterius (1983)
Zizaniopsis miliacea	7.37	Lanning & Eleuterius (1983)
Zoysia japonica	2.85	Butler & Hodges (1967)

Supplemental Table 2: Top z-score matched protein structure for all I-TASSER NIP structural models.

NIP Type	Modeled Protein	Top Z-score Match
NIP I	At_NP_198598_1.pdb	1J4N
NIP I	Hv_BAJ96213.pdb	2W1P
NIP I	Os_EEC78689.pdb	2W1P
NIP I	Ps_CAB45652_1.pdb	1J4N
NIP II	Ea_CCI55658_1.pdb	4NEF
NIP II	Os_NP_001041813_1.pdb	2W1P
NIP II	Pp_XP_001779449_1.pdb	1J4N
NIP II	Sm_XP_002963220_1.pdb	2D57
NIP III	Ca_CAG34223.pdb	1J4N
NIP III	Cm_BAK09176.pdb	3GD8
NIP III	Hv_BAH24163_1.pdb	2B6P
NIP III	OS_NP_001048108_1.pdb	2B6P

Appendix B

GENE SEQUENCING-BASED ANALYSIS OF MICROBIAL MAT MORPHOTYPES, CAICOS PLATFORM, BRITISH WEST INDIES

Abstract

Active carbonate platforms provide modern analogs to study microbial-mat development and taphonomy in the sedimentary record. Microbial-mat descriptions and classifications for tropical tidal-flat environments have focused predominantly on morphological observations. This is exemplified by flat and biscuit-shaped mats, where the mat morphotypes are postulated to reflect different Cyanobacteria communities as the main mat-building taxa. To compare the total microbial communities of these two mat types and test this Cyanobacteria hypothesis, we applied optical microscopy and gene sequencing methods using samples from a tidal algal marsh on Little Ambergris Cay, Turks and Caicos, B.W.I. With gene sequencing we find that total diversity and community composition differs significantly between morphotypes; the biscuit mat is more diverse than the flat mat. Microscopy results support that Cyanobacteria populations colonizing the surface layer of these two mat types are responsible for much of the mat's structural elements; however, genetic data find the Cyanobacteria population is indistinguishable between the two mat types. The recovered Cyanobacteria populations fall predominantly into three taxa: *Scytonema*, *Halomicronema,* and *Crinalium*. We propose that the morphology of these two mat types is not controlled by the Cyanobacteria, but instead reflects a time-integrated microbial response to environmental factors, where the microbial community becomes more diverse with time since environmental disturbance.

Introduction

In carbonate strata of all ages, stromatolites — attached, lithified sedimentary growth structures, accretionary away from a point or limited surface of initiation (Grotzinger & Knoll 1999) — are widely interpreted as a record of the interaction of microbial communities (particularly Cyanobacteria) with carbonate sediments and cements (e.g., Frantz et al. 2015). Although it is unclear what processes determine the morphology of microbial mats, hypotheses include microbial community composition (Dupraz & Visscher 2005; Gerdes et al. 2000; Golubic et al. 2000; Noffke 2010; Shepard & Sumner 2010), metazoan and protistan grazing (Bernhard et al. 2013; Garrett 1970), hydrodynamics and sedimentation (Andres & Reid 2006; Gebelein 1969; Mariotti et al. 2014; Martin et al. 1993), and other environmental factors (Gerdes et al. 2000; Petroff et al. 2010; Wharton et al. 1983). In modern environments, microbial mats in tidal flats, sabkhas, and shallow subtidal zones are commonly considered analogs and possible precursors to stromatolites (e.g., Browne et al. 2000; Dupraz et al. 2009).

Modern microbial mats develop a variety of morphologies ranging from flat, laminar forms, to cone-like pinnacles, to small domal biscuit structures (Browne et al. 2000). While naming conventions and classification of microbial mats in active carbonate tidal environments vary, general descriptions have focused on distinctions between a flat, laminar mat type (Fig. 1D) and a raised, biscuit-mat type (Fig. 1C). In their description of storm-disturbed West Caicos microbial mats Wanless et al. (1988) suggested that the two morphologies were a result of different "algae" with different colonization strategies: the flat laminar mats were made by Schizothrix and the biscuit-type mats by Scytonema. Schizothrix was described as a rapid colonizer, forming a new surface mat layer within weeks of the storm, where sediment cover was millimeters thick. Scytonema was described as a slower-growing mat type that eventually colonizes areas previously colonized by Schizothrix on the order of months, in the absence of smothering sediment flux.

In contrast, Gebelein (1969) described what were termed Schizothrix mats composed of the same organism with different surface expressions based on sedimentation rates and water velocity. Additional observations of open marine microbial-mat structures from the Bahamas suggested that accommodation space (water depth) is another important factor in controlling growth morphology (Andres & Reid 2006). Finally, Golubic (1991) described mat types similar to the Bahamian mats

in the sabhkas of Abu Dhabi as “gelatinous laminated biscuits” and “low flat mats,” in the subtidal and mid-intertidal zones, respectively, distinguished both by their environmental context and by different Cyanobacteria communities, as determined by morphology. Therefore these forms, and the factors controlling them, may not pertain only to the Bahamas.

It is critical to note that for the prior studies cited above, and others (Freytet & Verrecchia 1999; Paerl et al. 2001), the names “Schizothrix” and “Scytonema” were used to define shapes of microorganisms found in the mats rather than the genetic identity associated with those classifications. This led to the description of entire mats by the names Schizothrix or Scytonema (e.g., “Schizothrix mats” and “Scytonema mats”; Wanless et al. 1988) based on microscopic morphological observation. While microscopy still holds substantial value for many aspects of microbial ecology, genetic identification provides an objective comparative-biology framework and is the current gold standard for taxonomic classification of microorganisms (Woese 1987). This is particularly important for Cyanobacteria, for which morphology may appear diagnostic but is homoplasic, particularly baeocystous and filamentous cells types (Shih et al. 2013). Formally the terms *Schizothrix* and *Scytonema* define different genera of Cyanobacteria; here we reserve the use of these terms solely to denote the genetic clades, not morphological attributions at either microscopic or macroscopic length scales. Consequently we use the terms biscuit mat and flat mat to describe the two most common morphotypes on the Caicos platform (Wanless et al. 1988) and test the implicit assumption that these different morphotypes reflect the mat-building activities of different Cyanobacteria by mapping between observed structures and the phylogenetic identity of the taxa within them. We labor under current Cyanobacteria nomenclature accepted by the 16S rRNA gene *SILVA* database maintained by the Microbial Genomics and Bioinformatics Research Group in Bremen, Germany (Quast et al. 2013) as the classification scheme for all microorganisms discussed in this work.

To compare the microbial populations of the flat and biscuit mat morphologies, we collected microbial-mat samples of both morphotypes from tidal flats of the Caicos platform, Turks and Caicos, B.W.I. (Fig. 2). We examined the mats using microscopy and NGS iTag technologies. iTag sequencing is particularly valuable for profiling and comparing microbial diversity in complex samples because it focuses on a short, hypervariable region of the 16S rRNA gene — a classic marker used in phylogenetic studies (Caporaso et al. 2012). This technology is able to

produce hundreds of thousands of sequencing reads per sample, and is therefore the currently preferred technique to access and compare the microbial diversity of a wide range of environmental samples. Our results showed high similarity in Cyanobacteria populations between morphotypes. Holistic community analysis showed differences between morphotypes and suggested that other factors have greater influence on determining mat morphology on the Caicos platform than the Cyanobacteria.

Methods

The mats studied occur in a tidal marsh in the center of Little Ambergris Cay, West Caicos, B.W.I., visited in February, 2014 (Fig. 2). Both mat samples were collected during midday. The flat-mat sample was collected near the main tidal channel connecting the lagoon to the Caicos platform interior (Fig. 1A), and the biscuit mat sample was collected in the more interior part of the lagoon. The portion of the lagoon surveyed by foot and unmanned aerial vehicle (UAV) contained large regions of biscuit mats, intermixed with areas of flat laminar mats, and dynamic sediment-filled channels with no mat development (Fig. 1A, B). Individual biscuits ranged up to 20 cm wide and 10 cm high (Fig. 1C, E). Samples were collected from two representative locations on each mat morphotype (Fig. 1C, D), but importantly none of the microscopic and macroscopic visualization of the two mat types (n = 10) gave any indication of compositional differences between the Cyanobacteria observed in each morphotype.

Sampling was accomplished by aseptic coring (upper ~ 3 cm of mat) with sterile 50 ml polypropylene conical centrifuge tubes. Samples were kept at 4^{o}C until processed. A subset of the two mats was sectioned visually by pigment layer under a dissection microscope and then preserved in paraformaldehyde. These samples were washed and stored in ethanol at -20 ^{o}C. Preserved samples were vortexed to disaggregate the mat layers before pipetting onto slides used for microscopy and micrographs.

DNA was extracted from a thin (~ 3 mm^2 cross-sectional area), vertical section of each mat (~ 1 g total biomass) removed by sterile razor. Samples were mechanically lysed in a bead beater (FastPrepFP120, ThermoElectronCorp.) for 45 s at setting 5.5. DNA was extracted using the Power Soil DNA extraction kit (Mo Bio Laboratories, Inc.). iTag samples were prepared with

Earth Microbiome Project primers (515f and 806r) and recommended reagents 5 Prime Master Mix; (Caporaso et al. 2012). An initial amplification of 30 cycles with primers lacking the barcode, linker, pad, and adapter was performed for all samples, in duplicate. All samples yielded PCR amplicons when viewed on a gel after initial pre-barcoding PCR (30 cycles). Duplicate PCR reactions were pooled and reconditioned for five cycles with barcoded primers. PCR negative controls, substituting PCR water for DNA template, were amplified for 40 cycles total and also sequenced.

Resulting iTag sequences were processed using the *mothur* (Schloss et al. 2009) Standard Operating Procedure (SOP) for Illumina MiSeq sequencing of the 16S rRNA gene V4 region (accessed online May 2015). A concatenated file of the *mothur* version of separate archaeal and bacterial SILVA version 119 databases was used for alignment and taxonomic classification of sequence reads (Quast et al. 2013; Schloss & Westcott 2011; Schloss et al. 2009). Any taxa in the PCR negative control sample were removed from the resulting microbial-mat taxon database. iTag sequences have been submitted to the SRA under Bioproject: PRJNA316900.

Assessment of sampling depth was made with Good's Coverage — a common ecological approach that estimates the percent of the total species in an environment that were recovered in the sampling of that environment, equal to 1-[number of operational taxonomic units (OTUs) that have been sampled once, divided by the total number of all individuals sampled] multiplied by 100 (Good 1953). Alpha diversity was estimated using the Inverse Simpson metric (1/D) where D is a measure of the number of times an OTU is observed (species richness) divided by the total number of individuals in a community (species evenness) (Hill 1973; Simpson 1949). We used the UniFrac distance metric (Lozupone & Knight 2005) to assess the microbial community phylogenetic similarity. This method determines phylogenetic trees from the sequences in each sample and computes the branch length that is unshared between the each sample's tree, effectively quantifying how dissimilar the two communities are. All statistics were calculated using scripts in *mothur* and are reported at the unique sequence, 99%, and 97% OTU similarity levels.

Figure 1: A) UAV photo mosaic from north shore of Little Ambergris Cays over tidal marsh. Red star indicates the sampling location of flat mats and the yellow star in darker region marks the sampling location of biscuit mats. Orange star indicates sediment-filled channel with no mat growth, and red arrow highlights a person for scale. White dots numbered 1 - 4 orient the mosaic with the following GPS coordinates: 1) 21.306231° N, 71.675926° W; 2) 21.301820° N, 71.686693° W; 3) 21.305593° N, 71.691211° W; 4) 21.297430° N, 71.725451° W. B) Close-up of contact between and examples of flat mats (lower half of image) and biscuit mats (upper half). Black bar is approximately 0.5 m. C) Close-up of biscuit mats in the sampled region with hand for scale. D) Close-up of the flat mats in the sampled region with hand for scale. E) Vertical cross-section through a biscuit mat showing shape and internal structure with hand for scale. F) Vertical crosssection through a biscuit mat showing annotated pigmentation layers (G = green, P = purple and pink, B = brown). Black bar is approximately 1 cm. G) Vertical crosssection through flat mat showing annotated pigmentation layers (G = green, P = purple and pink, B = brown). Hand is for scale.

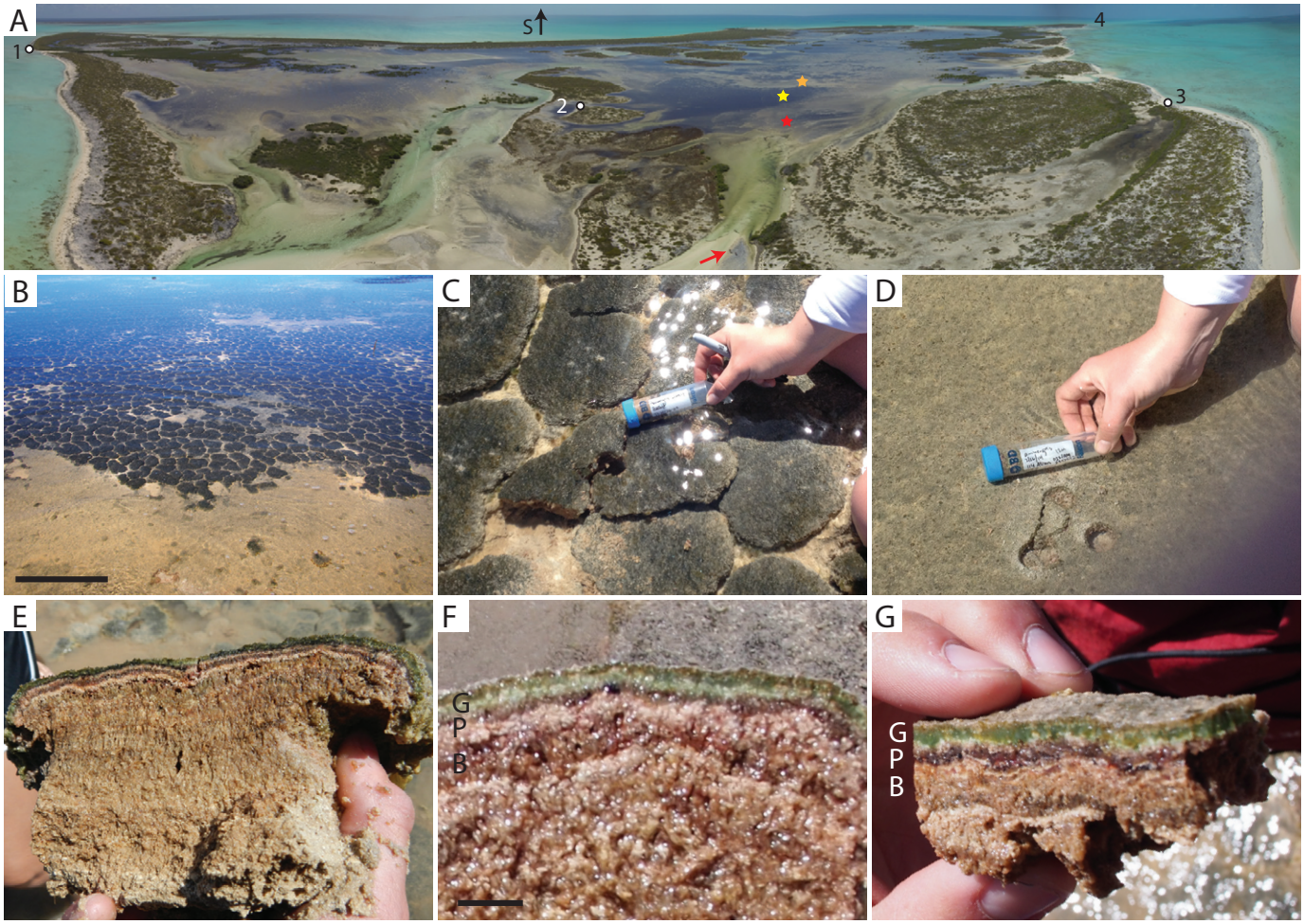


Figure 2: Location of study on Little Ambergris Cay within the Caicos platform. Inset shows location of Turks and Caicos with respect to the Bahamas, and neighboring Caribbean countries.

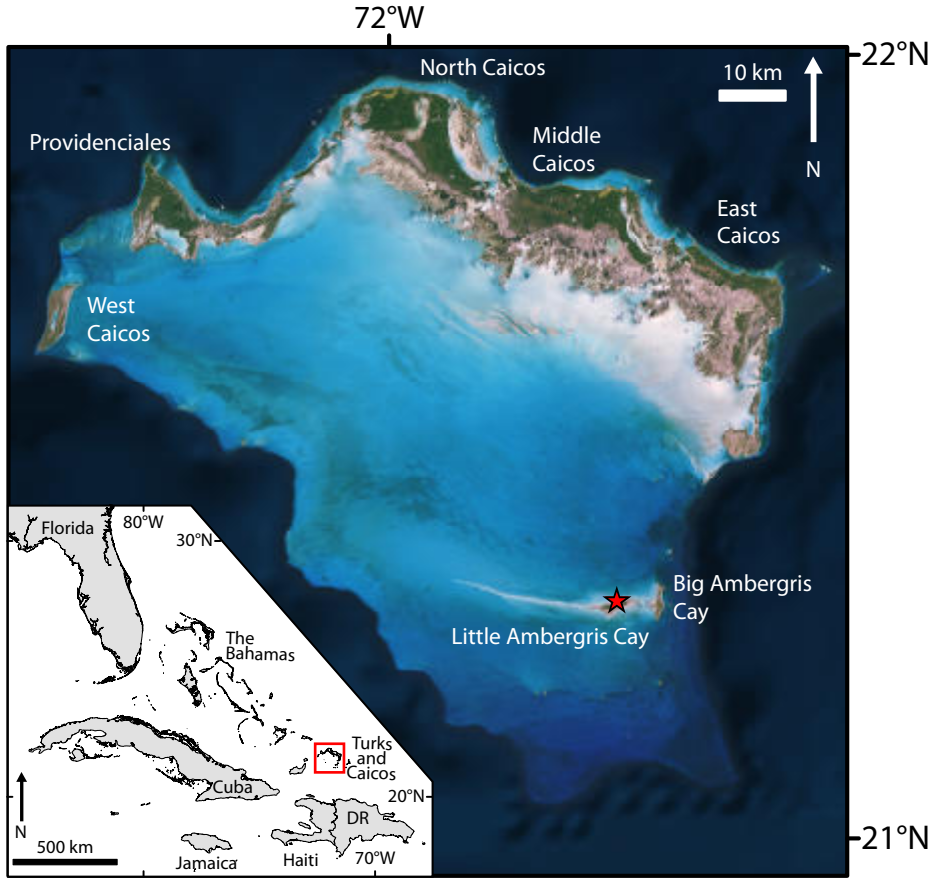
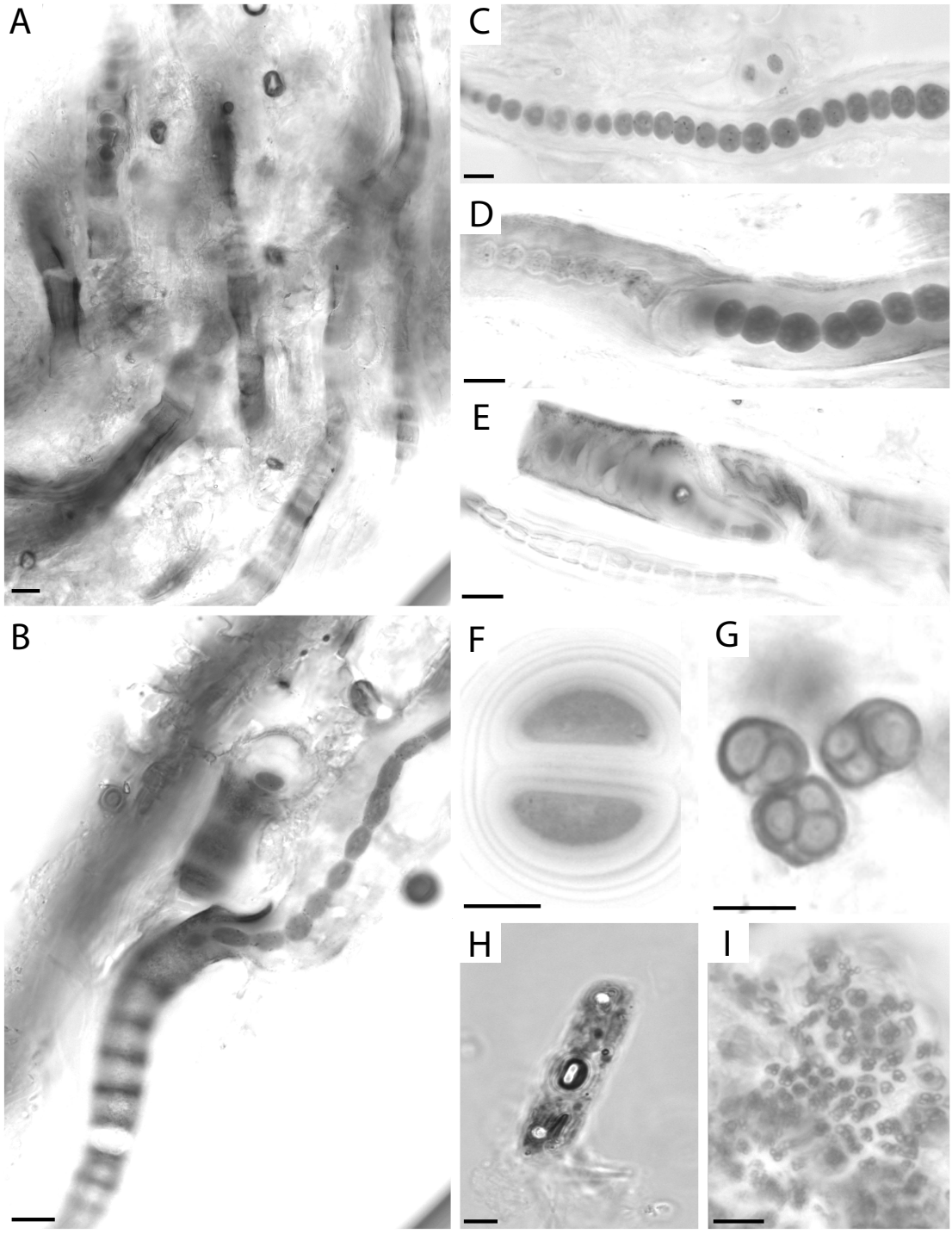


Figure 3: High-contrast black-and-white photomicrographs of the most dominant microbial morphologies found in both flat and biscuit mats from Ambergris Cay, West Caicos, BWI. A) Abandoned light brown sheaths dominant in the lower layers of the biscuit mat. B) Cyanobacteria with brown sheaths can also be found in flat mat. C - E) Images of bright green filamentous Cyanobacteria enclosed in sheaths. Sheaths grade from clear in C to light brown in D. Light and dark granules or occlusions are visible in single Cyanobacteria cells. F) Green and G) brown coccoid cells. H) Micrograph of putative single-celled green algae. I) Purple colonial microbes. Images A, C, D, E, and I are from biscuit mat. Images B, F, G, and H are from flat mat. Black bar is 10 μm.



Results

Mat Description

Flat mats (Fig. 1D) are characterized by lateral continuity of surface mat laminae (Fig. 1G), whereas biscuit mats (Fig. 1E) are present as discrete quasi-polygons with mat layering oriented normal to the convex-up surface of the quasi-polygon, curving away from the horizontal on quasi-polygon margins. We observed no clear differences in the relative abundances of grazers (only rare millimeter-scale cerithid gastropods) in the region surveyed. The general compositions of both flat and biscuit mat types follow an expected vertical progression of mm-scale pigmented zonation, from a dark surface layer to lighter green layers to purple and pink layers near the bottom of the pigmented mat section (Fig. 1F, G). These visible color changes in pigments are indicative of canonical photosynthetic microbial-mat communities stratified by light penetration and anaerobic biogeochemical processes (Stal 2012; Stolz 2000; Van Gemerden 1993). Below the bright, pigmented portion of the mat are several centimeters of brown organic material and carbonate sediment (ooids and mud), at times creating visibly inter-bedded laminae (Fig. 1E). Overprinting these layers is a palisade-type texture consisting of the empty sheaths of Cyanobacteria (Fig. 3A) that no longer contain the bright, green living Cyanobacteria cells seen in high contrast in Fig. 3B-E. These sheaths are the dominant structural component of the biscuit-type morphology. Brown (Fig. 3G) and green (Fig. 3G) coccoid bacteria, putative single-celled green algae (Fig. 3H), and maroon coccoids (Fig. 3I) were also observed in the pigmented regions of both mat types.

Sequencing

Cyanobacteria.--- Twenty-two Cyanobacteria taxa were recovered between the two mats (Table 1). The same three taxa predominate in both mat types: *Scytonema* (0.94% relative abundance in flat mat, 0.89% relative abundance in biscuit mat), *Halomicronema* (0.42% relative abundance in flat mat, 0.57% relative abundance in biscuit mat), and *Crinalium* (0.06% relative abundance in flat mat, 0.15% relative abundance in biscuit mat) (Table 1, gray). While sequences belonging to *Schizothrix* were not observed in our samples, *Halomicronema* and *Crinalium* were classified in the same morphological Subsection and Family (Subsection III; Family I), and share similar

morphologies. While the dominant Cyanobacteria are effectively the same between the two mat types, the biscuit mat has higher alpha diversity than the flat mat. In 1,000 random *in silico* subsamplings to equal depth of the Cyanobacteria populations in each sample, the same number of unique operational taxonomic units (OTUs) were observed.

Table 1. Relative abundances of the complete cyanobacterial taxonomic diversity of flat and biscuit microbial-mat samples. The only differences in Cyanobacteria diversity between mat types occur in rare taxa that are all less than 0.15% of the total relative abundance. Seven taxa were observed only in the flat-type mat [underlined], and six taxa were found only in the biscuit-type mat [italics]. Average and standard deviation for Number of OTUs and Inverse Simpson metric calculated for 1000 random subsamplings to equal depth of the Cyanobacteria population of each mat.

Cyanobacterial Taxonomy	Flat Mat Rel. Abund.	Biscuit Rel. Abund.
SubsectionIV;FamilyI;Scytonema	0.924%	0.893%
SubsectionIII;FamilyI;Halomicronema	0.424%	0.572%
SubsectionIII;FamilyI;Crinalium	0.062%	0.146%
SubsectionI;FamilyI;Cyanothece	0.036%	0.012%
SubsectionV;FamilyI;Hapalosiphon	0.030%	0.004%
SubsectionIII;FamilyI;Tychonema	0.018%	0.110%
SubsectionI;FamilyI;Chroococcus	0.015%	0.001%
SubsectionIII;FamilyI;Phormidium	0.013%	0.000%
SubsectionIII;FamilyI;Euhalothece	0.012%	0.000%
SubsectionIII;FamilyI;Geitlerinema	0.009%	0.004%
SubsectionII;FamilyI;Xenococcus	0.006%	0.000%
SubsectionIII;FamilyI;Rubidibacter	0.004%	0.000%
SubsectionIII;FamilyI;Prochlorothrix	0.002%	0.000%
SubsectionIII;FamilyI;Arthrospira	0.001%	0.000%
SubsectionIII;FamilyI;Spirulina	0.001%	0.027%
SubsectionII;FamilyI;Stanieria	0.000%	0.003%
SubsectionII;FamilyII;Pleurocapsa	0.000%	0.006%
SubsectionIII;FamilyI;Aerosakkonema	0.000%	0.001%
SubsectionIII;FamilyI;Haloleptolyngbya	0.000%	0.027%
SubsectionIII;FamilyI;Leptolyngbya	0.000%	0.126%
SubsectionIII;FamilyI;Trichocoleus	0.000%	0.002%
SubsectionIV;FamilyI;Cylindrospermum	0.000%	0.031%
Total Cyanobacterial Rel. Abund.	1.6%	2.0%
Top 3 Species Rel. Abund.	1.4%	1.6%
Top 3/Total Rel. Abund.	0.91	0.82
Avg. Number of OTUs	62.00	67.00
Std. Number of OTUs	0.00	3.79
Avg. Inverse Simpson	2.61	4.16
Std. Inverse Simpson	0.00	3.82-4.55

Total Diversity.--- In aggregate, we recovered 85,319 sequences for the flat-type mat and 101,610 sequences for the biscuit-type mat. More OTUs were also observed for the biscuit morphology (Table 2). We captured 98% of the microbial community for both samples based on the Good's Coverage statistic at the unique and 99% OTU level, and 100% at the 97% OTU level (Table 2), demonstrating that these differences are not due to differential community recovery during sequencing. Finally, the flat-mat morphology has half the Inverse Simpson diversity of the biscuit mat across all OTU levels within the 95% confidence interval (Table 2). The biscuit mat has more sequences, OTUs, and higher diversity than the flat mat.

Table 2. Number of observed OTUs, sequencing coverage (Goods Coverage), and species richness (alpha diversity - Inverse Simpson) for rarefied dataset.

OTU Clustering	Sample	OTUs Observed	Goods Coverage	Inv. Simpson	Inv. Simpson 95% Conf.
unique	Flat	3137	0.98	86	83-88
unique	Biscuit	3518	0.98	173	170-177
0.01	Flat	3092	0.98	85	83-88
0.01	Biscuit	3460	0.98	173	170-176
0.03	Flat	1472	1.00	68	66-70
0.03	Biscuit	1562	1.00	139	136-141

Based on UniFrac analysis, 31% of the phylogenetic diversity is unshared between the two microbial-mat samples. Seven of the top 10 taxa of both samples are not found in the other sample, most of which are from the phylum Proteobacteria (Table 3). *Gammaproteobacteria*; *Vibrionales*, *Holophagae*, *Alphaproteobacteria*; *Rhodospirllales*, *Deltaproteobacteria*; Sh765B-Tzt-29, *Alphaproteobacterales*; *Rhodobacterales* were all observed in the flat mat but not the biscuit mat. Conversely, *Alphaproteobacteria*; *Rhizobiales*, *Deltaproteobacteria*; *Desulfovibrionales*, *Gammaproteobacteria*; *Chromatiales*, *Planctomycetes*, *Bacteroidetes*, and *Deltaproteobacteria*; *Syntrophobacterales* were found in the biscuit mat but not in the flat mat.

Table 3. Top ten most abundant sequences in both flat and biscuit mat types with SILVA taxonomy. Highlighted taxa appear in both mat sample types.

Top 10 SILVA Taxa of Flat Mat	Rel. Abund.	Top 10 SILVA Taxa of Biscuit Mat	Rel. Abund.
Proteobacteria;GammaproteoVibrionales; Vibrionaceae;Vibrio	10.19%	Chloroflexi;Anaerolineae;Anaerolineales; Anaerolineaceae	14.81%
Chloroflexi;Anaerolineae;Anaerolineales; Anaerolineaceae	5.23%	Proteobacteria;Alphaproteobacteria; Rhizobiales; Hyphomicrobiaceae;Rhodomicrobium	4.86%
Acidobacteria;Holophagae;B276-D12	4.15%	Spirochaetae;Spirochaetes;Spirochaetales; Spirochaetaceae;Spirochaeta	3.69%
Proteobacteria;Alphaproteobacteria; ss1-B-07-44	3.57%	Proteobacteria;Deltaproteobacteria; Desulfovibrionales;Desulfovibrionaceae; Desulfocurvus	3.37%
Proteobacteria;Alphaproteobacteria; Rhodospirillales;MSB-1E8	2.44%	Proteobacteria;Gammaproteobacteria; Chromatiales;Chromatiaceae;Thiococcus	3.13%
Proteobacteria;Deltaproteobacteria; Sh765B-TzT-29	2.37%	Planctomycetes;Phycisphaerae;mle1-8	2.63%
Spirochaetae;Spirochaetes;Spirochaetales; Spirochaetaceae;Spirochaeta	2.23%	Bacteroidetes;SB-5	2.08%
Proteobacteria;Alphaproteobacteria; Rhodospirillales;Rhodospirillales_ Incertae_Sedis; Candidatus_Alysiosphaera	2.13%	Bacteroidetes;Sphingobacteriia; Sphingobacteriales; Saprospiraceae;Lewinella	2.05%
Bacteroidetes;Sphingobacteriia; Sphingobacteriales;Saprospiraceae; Lewinella	2.02%	Proteobacteria;Alphaproteobacteria; ss1-B-07-44	1.92%
Proteobacteria;Gammaproteobacteria; Pseudomonadales;Pseudomonadaceae; Pseudomonas	1.59%	Proteobacteria;Deltaproteobacteria; Syntrophobacterales; Syntrophobacteraceae;Desulfacinum	1.88%
Proteobacteria;Alphaproteobacteria; Rhodobacterales;Rhodobacteraceae; Tropicimonas	1.53%	TA06	1.70%
Four shared species	13.04%	Four shared species	22.47%
All top ten species	37.44%	All top ten species	42.11%

Discussion

Visual observation confirmed that the Cyanobacteria construct the main structural components of both microbial-mat types in the form of discarded sheaths. The biomass also contained interbedded layers of sediments indicative of past episodes of sedimentation, followed by recolonization of the substrate by the microbial community. Sequence analysis shows that the phylogenetic identity of the Cyanobacteria populations is extremely similar between the two mat types, where greater than 82% of the Cyanobacteria observed in both mat types belong to the same

three taxa. We therefore fail to reject the null hypothesis that the two morphotypes have the same Cyanobacteria population.

From sequence data, Cyanobacteria make up less than 2% of the total relative abundance of either mat community. However, it is important to note that sequence abundance cannot be directly correlated with population size, as iTag data can have biases in amplification efficiencies between different types of microorganisms (Parada et al. 2015). Our microscopic evaluation shows that Cyanobacteria are more than 2% of the microbial population by number, and certainly by biovolume, and therefore remain relevant, structure-building members of the mat community.

Though our sequencing efforts were limited to two representative samples, we also completed microscopic evaluations from a larger sample size (n = 10) collected at the same locations and times as the sequencing samples. These microscopic evaluations did not show any clear differences between the two morphotypes. As was observed in stromatolites at Highborne Cay (Foster et al. 2009), morphological observations did not completely capture the diverse and complex Cyanobacteria community diversity in these mat morphotypes.

The dominant Cyanobacteria present in these mat samples are members of the genera *Scytonema, Halomicronema,* and *Crinalium*. Members of *Scytonema* were also found in the thrombolite metagenome from Highborne Cay, Bahamas (Mobberley et al. 2013). Statistical analyses of the two microbial-mat samples show that we recovered the majority of the microbial community in our sampling (Good's Coverage 98% or greater) and that the alpha diversity of the biscuit mat is twice that of the flat mat. Whole-community diversity analysis (UniFrac) shows that about one third of the diversity in each mat sample is unshared.

The difference in diversity between the two microbial-mat morphotypes supports the hypothesis that mat morphology is defined mainly by time since mat colonization. Observations of our sampling site over time also corroborate this hypothesis, where flat mats were later found growing in locations previously containing biscuit mats after storm events (S. Bachtel, 2015, personal communication). In this scenario, faster-growing populations initially colonize the microbial mat. Then, with time, the more established mat would accumulate a more complex and diverse microbial population (Reid et al. 2000; Stal et al. 1985), concurrent with a development in mat

morphology from flat to biscuit. This is analogous to plant diversity in the development of a forest, where initial colonization is performed by a few rapidly growing or stress-tolerant species that are later joined or replaced by a more complex community (Sigler & Zeyer 2004). Successional diversity mechanics have also been suggested by other studies looking at a different set of Bahamas mats in Highborne Cay (Baumgartner et al. 2009).

When examining the top ten most abundant taxa of each mat morphotype, most taxa that differ between the two occur in the bacterial phyla Proteobacteria. Deltaproteobacteria make up 2 - 3% of the microbial population in the two mats — these organisms are commonly capable of sulfate reduction, which is consistent with the view that sulfur cycling is important in these ecosystems (Baumgartner et al. 2006; Visscher et al. 2000; Visscher et al. 1998; Visscher & Stolz 2005). OTUs belonging to the anoxygenic phototrophic sulfide-oxidizing Gammaproteobacteria genus, *Thiococcus,* were present at about 3.1% abundance in the biscuit-type mat, likely using the photosynthetic oxidation of sulfide produced by sulfate reducers to drive carbon fixation. This type of anaerobic closed internal sulfur cycle has been demonstrated in siliclastic tidal marshes at mid-latitudes (Wilbanks et al. 2014).

Chloroflexi have not been the focus of previous Bahamas microbial mat studies, though it is possible that their filamentous morphology may have been mistaken for Cyanobacteria. Sequences belonging to the Chloroflexi are among the most abundant taxa in both mat types (Table 3; biscuit-type mat 14.81%, flat-type mat 5.23%). The Chloroflexi present in the Ambergris Cay mats fall into the class Anaerolineae, a group which is typically characterized by anaerobic, nonphototrophic heterotrophs (Yamada et al. 2006), though many appear to be capable of aerobic respiration (Hemp et al. 2015a; Hemp et al. 2015b; Pace et al. 2015; Ward et al. 2015). Thus we suggest that the Chloroflexi observed in these mats play an important role in both aerobic and anaerobic carbon cycling, breaking down biomass produced by Cyanobacteria and other photo- and chemoautotrophs. We also note that it has also recently been shown that Cyanobacteria themselves may play important roles as heterotrophs respiring organic matter in these systems (Stuart et al. 2015).

We observed members of the Rhodospiralles (typically facultative photoheterotrophs) present in the flat-type mat at 2.4% and 2.1% abundance, as well as *Tropicomonas* at 1.5%, a member of the

Rhodobacteraceae — a metabolically versatile group that includes aerobic and anaerobic heterotrophs as well as facultative photoheterotrophs. The biscuit-type mat had 4.9% of sequences corresponding to *Rhodomicrobium*, a photoheterotrophic Alphaproteobacteria. These anoxygenic photoheterotrophic organisms likely inhabit the base of the photic zone in the mat, below the Cyanobacteria, where they can utilize organic compounds from the breakdown of biomass from primary producers while also making use of light energy to generate ATP (Imhoff 1995; Overmann & Garcia-Pichel 2013). A high proportion of Rhodobacteriales Alphaproteobacteria were also found in thrombolitic microbial mats in Highborne Cay, Bahamas, by genetic sequencing (Myshrall et al. 2010) and lipid analysis (Edgcomb et al. 2013).

Both samples also contain high abundances of sequences corresponding to various aerobic and anaerobic heterotrophs. Included in this grouping are Spirochaeta, common saccharolytic bacteria likely breaking down algal or Cyanobacteria extracellular polysaccharides in the mat (Leschine et al. 2006), as well as the Bacteroidetes genus *Lewinella*, a group known to be capable of protein and polysaccharide breakdown (Khan et al. 2007) and likely responsible for degrading the organic polymers common to these microbial mats. The flat mat also contains high concentrations of widespread Gammaprotobacteria aerobic heterotrophs *Vibrio* (~ 10%) and *Pseudomonas* (2%). These organisms likely occur in the upper, aerobic layers of the mat, where they aerobically respire organic compounds produced by Cyanobacteria and other autotrophs.

Conclusions

Microscopic and genomic data reveal that the flat and biscuit microbial-mat types present on the tidal flats of Little Ambergris Cay are not distinguished by their Cyanobacteria communities. They contain Cyanobacteria of similar morphology (based on microscopy) and similar phylogenetic diversity (based on gene sequence identity), at similar relative abundances. Additionally, there is no evidence for differences in the relative abundances of metazoan grazers between them. The two microbial-mat types do, however, contain differences in their non-Cyanobacteria populations, and the biscuit mat has a more diverse microbial community than the flat mat.

If we assume that the biscuit-mat morphology developed from an initial flat-mat architecture, as repeated field observations have also suggested, this diversity difference could be explained by mat

communities becoming more diverse the longer they remain undisturbed by changes in environmental conditions, with the frequency of sedimentation or erosion events due to storms or proximity to tidal channels probably the most important among these. Thus the results of this study support the hypothesis postulated early by Gebelein (1969), and expanded on by others (Andres & Reid 2006; Mariotti et al. 2014; Martin et al. 1993), that environmental factors play a more fundamental role in microbial-mat morphology than the Cyanobacteria communities concentrated within their upper layers. If the results from the Caicos mats are more broadly applicable to mat morphologies observed elsewhere, within the limited degree to which these mat morphologies may display differential textural expressions in the rock record, morphological interpretations might more profitably focus on paleoenvironmental information rather than the signatures of different microorganisms (e.g., (e.g., Grotzinger & Knoll 1999).

Acknowledgments

We thank Jena Johnson, Daven Quinn, Alison Piasecki, Mathieu Lapotre, Jennifer Buz, Hayden Miller, Ted Present, Kirsten Siebach, and Brooke Dallas for assistance in the field, and Chris Zahm for UAV photography. We thank Victoria J. Orphan for sequencing support. We thank Kathleen Wood (DEMA) for assistance with our study. Partial support for the fieldwork was graciously provided by the Agouron Institute.

References

Andres MS, and Reid RP. 2006. Growth morphologies of modern marine stromatolites: a case study from Highborne Cay, Bahamas. *Sedimentary Geology* 185:319-328.

Baumgartner LK, Reid RP, Dupraz C, Decho AW, Buckley D, Spear J, Przekop KM, and Visscher PT. 2006. Sulfate reducing bacteria in microbial mats: changing paradigms, new discoveries. *Sedimentary Geology* 185:131-145.

Baumgartner LK, Spear JR, Buckley DH, Pace NR, Reid RP, Dupraz C, and Visscher PT. 2009. Microbial diversity in modern marine stromatolites, Highborne Cay, Bahamas. *Environmental microbiology* 11:2710-2719. 10.1111/j.1462-2920.2009.01998.x

Bernhard JM, Edgcomb VP, Visscher PT, McIntyre-Wressnig A, Summons RE, Bouxsein ML, Louis L, and Jeglinski M. 2013. Insights into foraminiferal influences on microfabrics of microbialites at Highborne Cay, Bahamas. *National Academy of Sciences Proceedings (USA)* 110:9830-9834.

Browne KM, Golubic S, and Lee S-J. 2000. Shallow marine microbial carbonate deposits. In: Riding RE, and Awramik SM, editors. Microbial Sediments. Berlin: Springer-Verlag. p 233-249.

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM, Betley J, Fraser L, and Bauer M. 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *The International Society for Microbial Ecology (ISME) Journal* 6:1621-1624.

Dupraz C, Reid RP, Braissant O, Decho AW, Norman RS, and Visscher PT. 2009. Processes of carbonate precipitation in modern microbial mats. *Earth-Science Reviews* 96:141-162.

Dupraz C, and Visscher PT. 2005. Microbial lithification in marine stromatolites and hypersaline mats. *Trends in Microbiology* 13:429-438. http://dx.doi.org/10.1016/j.tim.2005.07.008

Edgcomb V, Bernhard J, Beaudoin D, Pruss S, Welander P, Schubotz F, Mehay S, Gillespie A, and Summons R. 2013. Molecular indicators of microbial diversity in oolitic sands of Highborne Cay, Bahamas. *Geobiology* 11:234-251.

Foster JS, Green SJ, Ahrendt SR, Golubic S, Reid RP, Hetherington KL, and Bebout L. 2009. Molecular and morphological characterization of cyanobacterial diversity in the stromatolites of Highborne Cay, Bahamas. *The International Society for Microbial Ecology (ISME) Journal* 3:573-587.

Frantz C, Petryshyn V, and Corsetti F. 2015. Grain trapping by filamentous cyanobacterial and algal mats: implications for stromatolite microfabrics through time. *Geobiology* 13:409-423.

Freytet P, and Verrecchia E. 1999. Calcitic radial palisadic fabric in freshwater stromatolites: diagenetic and recrystallized feature or physicochemical sinter crust? *Sedimentary Geology* 126:97-102.

Garrett P. 1970. Phanerozoic stromatolites: noncompetitive ecologic restriction by grazing and burrowing animals. *Science* 169:171-173.

Gebelein CD. 1969. Distribution, morphology, and accretion rate of recent subtidal algal stromatolites, Bermuda. *Journal of Sedimentary Petrology* 39:49-69.

Gerdes G, Klenke T, and Noffke N. 2000. Microbial signatures in peritidal siliciclastic sediments: a catalogue. *Sedimentology* 47:279-308. 10.1046/j.1365-3091.2000.00284.x

Golubic S. 1991. Microbial mats of Abu Dhabi. In: Margulis L, and Oledzenski L, eds. *Environmental evolution, Effects of the Origin and Evolution of Life on Planet Earth*. Cambridge: MIT Press, 103-130.

Golubic S, Lee S-J, and Browne KM. 2000. Cyanobacteria: Architects of sedimentary structures. In: Riding RE, and Awramik SM, eds. *Microbial Sediments*. Berlin: Springer-Verlag, 57-67.

Good IJ. 1953. The population frequencies of species and the estimation of population parameters. *Biometrika* 40:237-264.

Grotzinger JP, and Knoll AH. 1999. Stromatolites in Precambrian carbonates: evolutionary mileposts or environmental dipsticks? *Annual review of earth and planetary sciences* 27:313-358.

Hemp J, Ward LM, Pace LA, and Fischer WW. 2015a. Draft Genome Sequence of Ardenticatena maritima 110S, a Thermophilic Nitrate- and Iron-Reducing Member of the Chloroflexi Class Ardenticatenia. *Genome Announcements* 3:e01347-01315. 10.1128/genomeA.01347-15

Hemp J, Ward LM, Pace LA, and Fischer WW. 2015b. Draft Genome Sequence of Levilinea saccharolytica KIBI-1, a Member of the Chloroflexi Class Anaerolineae. *Genome Announcements* 3:e01357-01315. 10.1128/genomeA.01357-15

Hill MO. 1973. Diversity and evenness: A unifying notation and its consequences. *Ecology* 54:427-432. 10.2307/1934352

Imhoff JF. 1995. Taxonomy and physiology of phototrophic purple bacteria and green sulfur bacteria. In: Blankenship RE, Madigan MT, and Bauer CE, eds. *Anoxygenic Photosynthetic Bacteria*. Amsterdam: Springer Netherlands, 1-15.

Khan ST, Fukunaga Y, Nakagawa Y, and Harayama S. 2007. Emended descriptions of the genus *Lewinella* and of *Lewinella cohaerens*, *Lewinella nigricans* and *Lewinella persica*, and description of *Lewinella lutea* sp. nov. and *Lewinella marina* sp. nov. *International journal of systematic and evolutionary microbiology* 57:2946-2951.

Leschine S, Paster BJ, and Canale-Parola E. 2006. Free-living Saccharolytic Spirochetes: The Genus *Spirochaeta*. In: Balows A, Trüper HG, Dworkin M, Harder W, and Schleifer K-H, eds. *The Prokaryotes*. New York: Springer, 195-210.

Lozupone C, and Knight R. 2005. UniFrac: a new phylogenetic method for comparing microbial communities. *Applied and environmental microbiology* 71:8228-8235. 10.1128/aem.71.12.8228-8235.2005

Mariotti G, Pruss S, Perron J, and Bosak T. 2014. Microbial shaping of sedimentary wrinkle structures. *Nature Geoscience* 7:736-740.

Martin JM, Braga JC, and Riding R. 1993. Siliciclastic stromatolites and thrombolites, late Miocene, SE Spain. *Journal of Sedimentary Petrology* 63:131-139.

Mobberley J, Khodadad CM, and Foster J. 2013. Metabolic potential of lithifying cyanobacteria-dominated thrombolitic mats. *Photosynthesis Research* 118:125-140. 10.1007/s11120-013-9890-6

Myshrall KL, Mobberley JM, Green SJ, Visscher PT, Havemann SA, Reid RP, and Foster JS. 2010. Biogeochemical cycling and microbial diversity in the thrombolitic microbialites of Highborne Cay, Bahamas. *Geobiology* 8:337-354. 10.1111/j.1472-4669.2010.00245.x

Noffke N. 2010. *Geobiology: Microbial Mats in Sandy Deposits from the Archean Era to Today*. Berlin Heidelberg: Springer-Verlag.

Overmann J, and Garcia-Pichel F. 2013. The phototrophic way of life. In: Rosenberg E, DeLong EF, Lory S, Stackebrandt E, and Thompson F, eds. *The Prokaryotes*. Berlin: Springer, 203-257.

Pace LA, Hemp J, Ward LM, and Fischer WW. 2015. Draft Genome of Thermanaerothrix daxensis GNS-1, a Thermophilic Facultative Anaerobe from the Chloroflexi Class Anaerolineae. *Genome Announcements* 3:e01354-01315. 10.1128/genomeA.01354-15

Paerl HW, Steppe TF, and Reid RP. 2001. Bacterially mediated precipitation in marine stromatolites. *Environmental microbiology* 3:123-130.

Parada A, Needham DM, and Fuhrman JA. 2015. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time-series and global field samples. *Environmental microbiology*. 10.1111/1462-2920.13023

Petroff AP, Sim MS, Maslov A, Krupenin M, Rothman DH, and Bosak T. 2010. Biophysical basis for the geometry of conical stromatolites. *National Academy of Sciences Proceedings (USA)* 107:9956-9961.

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

Reid RP, Visscher PT, Decho AW, Stolz JF, Bebout B, Dupraz C, Macintyre I, Paerl H, Pinckney J, and Prufert-Bebout L. 2000. The role of microbes in accretion, lamination and early lithification of modern marine stromatolites. *Nature* 406:989-992.

Schloss PD, and Westcott SL. 2011. Assessing and improving methods used in operational taxonomic unit-based approaches for 16S rRNA gene sequence analysis. *Applied and environmental microbiology* 77:3219-3226. 10.1128/aem.02810-10

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, and Robinson CJ. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and environmental microbiology* 75:7537-7541.

Shepard RN, and Sumner DY. 2010. Undirected motility of filamentous cyanobacteria produces reticulate mats. *Geobiology* 8:179-190. 10.1111/j.1472-4669.2010.00235.x

Shih PM, Wu D, Latifi A, Axen SD, Fewer DP, Talla E, Calteau A, Cai F, de Marsac NT, and Rippka R. 2013. Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *National Academy of Sciences Proceedings (USA)* 110:1053-1058.

Sigler W, and Zeyer J. 2004. Colony-forming analysis of bacterial community succession in deglaciated soils indicates pioneer stress-tolerant opportunists. *Microbial ecology* 48:316-323.

Simpson EH. 1949. Measurement of diversity. *Nature* 163:688. 10.1038/163688a0

Stal LJ. 2012. Cyanobacterial mats and stromatolites. In: Whitton BA, ed. *Ecology of Cyanobacteria II*. Amsterdam: Springer, 65-125.

Stal LJ, van Gemerden H, and Krumbein WE. 1985. Structure and development of a benthic marine microbial mat. *Federation of European Microbiological Societies (FEMS) Microbiology Ecology* 31:111-125.

Stolz JF. 2000. Structure of microbial mats and biofilms. In: Riding RE, and Awramik SM, eds. *Microbial Sediments*. Berlin: Springer, 1-8.

Stuart RK, Mayali X, Lee JZ, Craig Everroad R, Hwang M, Bebout BM, Weber PK, Pett-Ridge J, and Thelen MP. 2015. Cyanobacterial reuse of extracellular organic carbon in microbial mats. *The International Society for Microbial Ecology (ISME) Journal*. 10.1038/ismej.2015.180

Van Gemerden H. 1993. Microbial mats: a joint venture. *Marine Geology* 113:3-25.

Visscher PT, Reid RP, and Bebout BM. 2000. Microscale observations of sulfate reduction: correlation of microbial activity with lithified micritic laminae in modern marine stromatolites. *Geology* 28:919-922.

Visscher PT, Reid RP, Bebout BM, Hoeft SE, Macintyre IG, and Thompson JA. 1998. Formation of lithified micritic laminae in modern marine stromatolites (Bahamas): the role of sulfur cycling. *American Mineralogist* 83:1482-1493.

Visscher PT, and Stolz JF. 2005. Microbial mats as bioreactors: populations, processes, and products. *Palaeogeography, Palaeoclimatology, Palaeoecology* 219:87-100. http://dx.doi.org/10.1016/j.palaeo.2004.10.016

Wanless HR, Tyrrell KM, Tedesco LP, and Dravis JJ. 1988. Tidal-flat sedimentation from Hurricane Kate, Caicos Platform, British West Indies. *Journal of Sedimentary Petrology* 58:724-738.

Ward LM, Hemp J, Pace LA, and Fischer WW. 2015. Draft Genome Sequence of Herpetosiphon geysericola GC-42, a Nonphototrophic Member of the Chloroflexi Class Chloroflexia. *Genome Announcements* 3:e01352-01315. 10.1128/genomeA.01352-15

Wharton RA, Parker BC, and Simmons GM. 1983. Distribution, species composition and morphology of algal mats in Antarctic dry valley lakes. *Phycologia* 22:355-365. doi:10.2216/i0031-8884-22-4-355.1

Wilbanks EG, Jaekel U, Salman V, Humphrey PT, Eisen JA, Facciotti MT, Buckley DH, Zinder SH, Druschel GK, and Fike DA. 2014. Microscale sulfur cycling in the phototrophic pink berry consortia of the Sippewissett Salt Marsh. *Environmental microbiology* 16:3398-3415.

Woese CR. 1987. Bacterial evolution. *Microbiological Reviews* 51:221-271.

Yamada T, Sekiguchi Y, Hanada S, Imachi H, Ohashi A, Harada H, and Kamagata Y. 2006. *Anaerolinea thermolimosa* sp. nov., *Levilinea saccharolytica* gen. nov., sp. nov. and *Leptolinea tardivitalis* gen. nov., sp. nov., novel filamentous anaerobes, and description of the new classes Anaerolineae classis nov. and Caldilineae classis nov. in the bacterial phylum Chloroflexi. *International journal of systematic and evolutionary microbiology* 56:1331-1340.