UNIVERSITY OF CALIFORNIA SAN DIEGO

Chemical ecology of marine microbial communities: An assessment of bacterial diversity and dynamics in tropical marine sediments

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Marine Biology

by

Alyssa Marie Demko

Committee in charge:

Professor Paul R. Jensen, Chair
Professor Eric E. Allen
Professor Jeff S. Bowman
Professor Ronald S. Burton
Professor Kaustuv Roy

2021

The Dissertation of Alyssa Marie Demko is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2021

DEDICATION

To my family, thank you for always believing in me.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 2

Chapter 3

Chapter 4

LIST OF TABLES

Chapter 2

Chapter 3

Chapter 4

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the help and support of many people. First and foremost, I want to thank my advisor, Dr. Paul Jensen for accepting me into his lab and for many years of great discussions. I would also like to thank Paul for taking me on multiple research trips; experiencing fieldwork in remote parts of the world has left me with some of my fondest memories.

Thank you to my committee: Eric, Ron, Jeff and Kaustuv. Eric, thank you for always making time to talk when I had questions or needed advice and thank you for supporting me in my teaching endeavors. Ron, I am so grateful that you were assigned to be on my first-year committee. You provided instrumental insight and I could not envision a thesis committee without you. Thank you for all of your advice on science and teaching the past few years. Jeff, thank you for letting me be your dive buddy, I greatly enjoyed chatting with you about research while pushing dive gear down the pier and am thankful for your guidance and help with bioinformatics. Kaustuv, thank you for talking with me about science and career trajectories, I have found your advice and feedback to be extremely helpful over the years.

I would also like to thank the SIO Grad Office. Thank you to Gilbert Bertado, Maureen McGreevy, Denise Darling and Shelley Weisel for your assistance in answering questions, facilitating a smooth graduate experience and hosting fun and informative events. Thank you to Tim DeBold for help navigating the logistics of TAing at SIO and thanks to Peter Franks for being such an approachable and understanding professor and department chair.

I am forever grateful to the various members of the Jensen lab, both past and present. In particular I want to acknowledge Krystle Chavarría, Dulce Guillén Matus, Natalie Millán-Aguiñaga, Kaitlin Creamer, Leesa Klau, Tracey Somera McDole, Anne-Catrin Letzel, Hans Singh,

Henrique Machado, Gregory Amos, Gabriel Castro-Falcón, Julia Busch and Nastassia Patin. Thanks to you all for your helpful guidance and for making time in and out of the lab so enjoyable.

I would also like to thank everyone else in the SIO community who has positively contributed to my experience at SIO. Thank you to Rich Walsh, Ashleigh Palinkas and Christian McDonald for keeping me up to date on scientific diving. Thanks to Amro Hamdoun, I am so glad I could TA for such an incredible professor. Thanks to the members of the Center for Marine Biotechnology and Biomedicine and those in regular attendance to 262, in particular Sheila Podell, Reiko Cullum, Bill Fenical, Lena Gerwick, Bill Gerwick and Brad Moore. I would also like to thank my CMBB peers, especially Chris Leber, Michelle Schorn, Trevor Purdy and Kayla Wilson.

I would not have pursued graduate school without the advice and mentorship of so many amazing teachers, professors and mentors along the way. Special thanks to Carla Bobko, Lisa Milke, Andy Rhyne, Brad Bourque, Paul Webb, Brian Wysor, Erik Sotka, Allan Strand and Valerie Paul.

Finally, I would like to acknowledge my family for their endless support and encouragement over the years. Thank you to my parents for being such great role models, instilling me with a strong work ethic and for fostering my love of the ocean. And a particular thanks to my mom, who is always willing to listen to me ramble about science and provide reassuring words when I am stressed or down. Thank you to my siblings, grandparents and Estelle for all the text messages, phone calls and video chats that make it seem like I am not so far from home. Thanks to Wendy, Aubrey, Megan and Nicole, words cannot express how grateful I am to have such wonderful friends. And thanks to Rick, for being an amazingly supportive partner who keeps me smiling and laughing in spite of the stresses of graduate school during a global pandemic.

Chapter 2 is coauthored with Nastassia V. Patin and Paul R. Jensen and was recently submitted for publication. The dissertation author was the primary investigator and author of this chapter.

Chapter 3 is coauthored with Leesa Klau, Alexander Bogdanov, Monica Cisneros and Paul R. Jensen. The dissertation author was the primary investigator and author of this chapter.

Chapter 4 is coauthored with Leesa Klau, Alexander Bogdanov, Alexander Chase and Paul R. Jensen. The dissertation author was the primary investigator and author of this chapter.

VITA

2008 – 2012	Bachelor of Science in Marine Biology and Chemistry Roger Williams University. Bristol, RI
2012 – 2015	Master of Science in Marine Biology College of Charleston. Charleston, SC
2015 – 2021	Doctor of Philosophy in Marine Biology Scripps Institution of Oceanography University of California San Diego. La Jolla, CA

PUBLICATIONS

Demko, AM, Amsler, CD, Hay, ME, Long, JD, McClintock, JB, Paul, VJ and Sotka, EE. 2017. Declines in plant palatability from polar to tropical latitudes depend on herbivore and plant identity. *Ecology* 98(9): 2312-2321.

Cavole, LM, Demko, AM, Diner, RE, Giddings, A, Koester, I, Pagniello, CMLS, Paulson, M, Ramirez-Valdez, A, Schwenck, SM, Yen, NK, Zill, ME and Franks, PJS. 2016. Biological impacts of the 2013-2015 warm water anomaly in the northeast pacific: winners, losers and the future. *Oceanography Magazine,* 29(2):273–285.

Tuttle, RN, Demko, AM, Patin, N, Kapono, C, Dorrestein, P and Jensen, PR. 2019. The detection of specialized metabolites and their producers in ocean sediments. *Applied Environmental Microbiology* 85(8): e02830-18.

Gauglitz, JM et al. 2020. Untargeted Mass Spectrometry-Based Metabolomics Tracks Molecular Changes in Raw and Processed Foods and Beverages. *Food Chemistry* 302: 125290.

Schorn, M et al. 2021. A community resource for paired genomic and metabolomic data mining. *Nature Chemical Biology. In Press.*

Demko, AM, Patin, NV and Jensen, PR. Microbial diversity in tropical marine sediments assessed using culture-dependent and culture-independent techniques. *In Review.*

Habener, LA, Podell, S, Creamer, KE, Demko, AM, Allen, E, Moore, BS, Ziemert, N, Letzel, AC and Jensen PR. The Natural Product Domain Seeker (NaPDoS) version 2: Relating ketosynthase phylogeny to biosynthetic function. (in preparation)

ABSTRACT OF THE DISSERTATION

Chemical Ecology of Marine Microbial Communities:

An Assessment of Bacterial Diversity and Dynamics in Tropical Marine Sediments

by

Alyssa Marie Demko

Doctor of Philosophy in Marine Biology

University of California San Diego, 2021

Professor Paul R. Jensen, Chair

Marine sediments cover ~70% of the earth and host rich and diverse microbial communities. These microbial communities play an integral role in global nutrient cycling and the food web. They can be both a source of disease and/or an agent of mitigation through the natural products they produce, which can have cascading impacts on community structure and ecosystem function. Despite their importance, marine sediment microbes remain woefully understudied. The goal of this dissertation was to use next generation sequencing technology and newly developed bioinformatic pipelines to gain insight into these complex communities. First, I sought to

reevaluate the 1% culturability paradigm by comparing sediment microbial communities using culture-dependent and culture-independent techniques. This comparative approach not only highlighted that >1% of sediment bacteria could be cultured, but also revealed the biases associated with culture-independent methods. Thirty-nine genera were identified in culture that were not detected with culture-independent methods, including some taxa that were fairly divergent from known cultured representatives. Next, I wanted to assess connections between sediment microbial communities, the sediment metabolome and sediment characteristics across varying spatial scales. To do this, microbial communities were sampled at three spatial scales, 1 m^2 quadrats, 10 m transects, and sites across a 12 km^2 area. Additionally, a small molecule *in situ* resin capture (SMIRC) method was employed to capture the metabolome present in sediments. The results from this study indicate that microbial diversity significantly increases with spatial scale and that sediment characteristics, such as grain size and nitrate concentrations, are significantly correlated with microbial communities. The SMIRC method was able to capture natural products and revealed the vast chemical landscape of marine sediments, much of which remains unexplored. Finally, I sought to evaluate how microbial communities in marine sediments vary in relation to the surrounding benthic environment by comparing fringing and back reefs of Mo'orea, French Polynesia. Even within a small area, ~1 km^2, fringing and back reef sediment communities were distinct from each other. Back reefs exhibited greater richness and diversity in the microbial communities while fringing reefs had greater metabolomic richness. Supervised correlative analyses identified connections between microbes, metabolites and environmental characteristics such as nutrient concentration. Many of the taxa identified in the network analyses belong to relatively unknown lineages, providing important insight into the role these lineages may be playing in their communities. In conclusion, the results of this dissertation provide fundamental

baseline information about the microbial communities and metabolites associated with marine sediments.

Chapter 1

Introduction

1.1 Microbial Diversity

While the field of microbiology dates back to the pioneering work of Antonie van Leeuwenhoek in the 1600s, environmental microbiology or microbial ecology was not an area of study until the late 1800s with the research of Martinus Beijerinck and Sergei Winogradsky. Their development of culture and enrichment techniques facilitated the ability to study specific groups of microbes and gain insight into their physiology and functional roles in the environment. While the first attempts to classify microbes relied solely on morphology, the use of culturing techniques and the ability to achieve pure cultures allowed for further phenotypic characterization and thus, improved microbial classification (Rosselló-Mora and Amann, 2001).

In the last century, the use of DNA in culture-independent methods has transformed our understanding of microbial taxonomy and microbial community ecology. DNA-DNA hybridization techniques became the standard practice for identifying new bacterial species which was then followed by the use of 16S rRNA gene sequencing. Whereas hybridization techniques require repetition with each potentially new species, sequencing of genes such as 16S result in a definitive characteristic that can be archived and repeatedly used (Staley, 2006). One major limitation however with the 16S rRNA gene is its accuracy in identifying bacteria down to the species level. Since 16S rRNA can be almost identical while DNA-DNA hybridization methods are indicative of separate species (Fox *et al.*, 1992), a threshold needed to be established for when

16S rRNA was indicative of a delineation at the species level (Staley, 2006). It was from comparative work with DNA-DNA hybridization and 16S rRNA sequencing, that a threshold of 97% became the marker: bacteria that shared <97% similarity with known species could therefore be classified as novel species (Stackebrandt and Goebel, 1994; Staley, 2006).

As sequencing technology continued to improve, whole microbial communities could be sequenced and identified based on the 16S rRNA gene, revolutionizing our understanding of microbial diversity (Rosselló-Mora and Amann, 2001; Madigan *et al.*, 2003). However, since next-generation sequencing methods frequently use short regions of the 16S rRNA gene and the 97% similarity threshold is now known to be more variable in its applicability to species, bioinformatic methods and classification standards continue to change. Whole genome sequencing is now frequently used for species level identification and for greater accuracy in building phylogenies crossing the domains of life. Not only do we now know that bacteria are everywhere, but the sheer diversity of bacteria continues to astound, with significantly more bacterial diversity on the tree of life relative to Archaea and Eukaryotes (Hug *et al.*, 2016). The development of genomic and metagenomic sequencing has not eliminated the use of 16S rRNA studies. Rather, given the efficiency of use and low cost of 16S rRNA sequencing, 16S rRNA is still widely used for high-throughput studies on microbial communities. However, in place of 97% operational taxonomic unit (OTU) clustering, these methods now use denoising pipelines to identify unique amplicon sequence variants (ASVs, also sometimes referred to as ESVs or zOTUs) which are then taxonomically classified (Callahan *et al.*, 2016; Amir *et al.*, 2017; Nearing *et al.*, 2018).

1.2 Microbial Communities and Marine Sediments

Additional improvements in the technology and bioinformatic pipelines associated with 16S rRNA have facilitated studying microbial communities (or microbiomes) in virtually all environments, unlocking insight into everything from the role of the gut microbiome in human disease (Shreiner *et al.*, 2015; Halfvarson *et al.*, 2017; Armour *et al.*, 2019; Tang *et al.*, 2019) to the diversity and function of microbial communities in remote and extreme environments (Zak and Kling, 2006; Achberger *et al.*, 2016; Lang *et al.*, 2017; Peoples *et al.*, 2019). Standardized methodology across studies, such as that employed by the Earth Microbiome Project (Thompson *et al.*, 2017), has further facilitated comparative insights into microbial communities. For example, Thompson *et al.* (2010) showed that free-living microbial communities exhibit significantly greater species richness than most host-associated communities, saline and non-saline communities are distinct from each other, and environment type has a strong influence on the alpha and beta diversity of microbial communities.

The use of culture-independent methods makes it possible to capture not only a broad picture of the microbial community present in different environments but it can also provide insight into community function through extrapolated connections to cultured representatives via tools like PICRUSt (Langille *et al.*, 2013) and PAPRICA (Bowman and Ducklow, 2015) or more directly through shotgun metagenomic analyses (e.g. Haas *et al.*, 2016; Dai *et al.*, 2018; Hillmann *et al.*, 2018). While advancements have also been made to culture microbes more effectively through methods such as the ichip (Berdy *et al.*, 2017) and adapting culturing techniques based on metabolic information gained from metagenomics (e.g. Tripp *et al.*, 2008), culture-independent studies continue to highlight the large diversity of microbes present on Earth, many of which remain uncultured to date (Lloyd *et al.*, 2018).

Not surprisingly, both culture-dependent and culture-independent efforts in environmental microbiology have predominantly focused on microbial communities associated with human interests. In particular, many human microbiome studies have yielded valuable insights into human health and disease (Gilbert *et al.*, 2018). For example patients suffering from recurrent *Clostridium difficile* infections can now be effectively treated using fecal microbiota transplantation procedures which cause patient fecal microbiomes to shift and resemble healthy donor communities (Weingarden *et al.*, 2015). Environments such as soils are also frequently studied due to the importance of agriculture. For example, novel biotechnological approaches are being proposed based on microbiome studies to optimize crop yield (Qiu *et al.*, 2019).

In contrast, marine sediments, which cover more than 2/3 of Earth, remain relatively understudied. The Earth Microbiome Project showed marine sediments to be among the most species rich environments (Thompson *et al.*, 2017) and studies on microbial abundances estimate sediments to contain the most phylogenetically novel microbes (Lloyd *et al.*, 2018). However, it was only recently that a study focused specifically on the diversity present globally in marine sediments (Hoshino *et al.*, 2020). Hoshino *et al.* (2020) compared sediment microbial communities from across 299 globally distributed locations and found significant correlations between the taxonomic composition of communities, the presence or absence of oxygen, and the organic carbon concentration of sediments. Additionally, taxonomic richness was predicted to be 7.85×10^3 - 6.10×10^5 ASVs for archaea and 3.28×10^4 - 2.46×10^6 ASVs for bacteria globally, an estimate comparable to seawater and topsoil environments (Hoshino *et al.*, 2020).

Previous research has also demonstrated the microbial richness and diversity on a fine scale. Probandt *et al.* (2018) examined individual grains of sand and visualize the attached microbial cells through CARD-FISH and microscopy techniques. Sequencing of sand grain

communities further demonstrated that thousands of unique taxa were associated with a single grain of sand (Probandt *et al.*, 2018). However, despite the ability to generate large, culture-independent datasets and associated functional information through metagenomics, sediments remain relatively understudied. Given the lack of study, much remains unknown about even common and globally distributed sediment lineages, adding to the difficulty in understanding these complex communities (Baker *et al.*, 2021).

1.3 Ecological Importance of Microbial Communities

Microbes act as the foundation for all ecosystems, providing essential functions for life on earth. The photosynthetic marine cyanobacterial genus *Prochlorococcus* alone, which was not discovered until the 1980s (Chisholm *et al.*, 1988), is both the smallest known photosynthetic organism and the most abundant in the ocean, making it a critical contributor to global oxygen supplies (Partensky *et al.*, 1999). And in ecosystems devoid of light, chemosynthetic microbes act as the primary producers in the system (Madigan *et al.*, 2003). Furthermore, microbes are a food source for other microscopic organisms at the base of the food web and are responsible for the decomposition of organic matter like decaying plants and animals, thus cycling nutrients back into the ecosystem (Snelgrove, 1997). Macro-organisms are also associated with a community of microbes, internally and externally, through their specific microbiomes. These remain poorly studied but can be important for the well-being of the host.

While microbial communities can exhibit natural fluctuations, instability in the community (or dysbiosis), can have drastic impacts on the host and/or system. For instance, dysbiosis in the human microbiome had been linked to health conditions such as Crohn's, obesity and cystic fibrosis (Cho and Blaser, 2012; Lynch and Bruce, 2013). Similarly, in marine host-systems,

dysbiosis in the coral microbiome contributes to health state and disease progression (Bourne *et al.*, 2009; Glasl *et al.*, 2019; Meyer *et al.*, 2019). In free-living microbial communities, changes to the community composition can have drastic impacts on community function. For example, coral reefs with greater algal cover host more copiotrophic bacteria in the surrounding seawater compared to reefs with greater coral cover (Haas *et al.*, 2016). Shifts in the metabolism of the microbial communities are also observed, with higher algal cover associated with a switch from the Embden-Meyerhof pathway of glycolysis to the less energetically efficient Entner-Doudoroff (ED) and pentose phosphate (PP) pathways (Haas *et al.*, 2016).

It is important to note that changes to the sediment microbial community can also have important ramifications for overall ecosystem structure and function. For instance, microbial communities associated with ecologically important *Zostera* seagrass beds are significantly different in their composition when compared to the sediment communities of *Caulerpa* beds (Gribben *et al.*, 2017). Laboratory experiments have demonstrated that the microbial communities from *Zostera* habitats mitigate the growth of invasive *Caulerpa* fragments while the communities from *Caulerpa* beds facilitate growth of *Caulerpa* fragments (Gribben *et al.*, 2017). Therefore, *Zostera* beds with dysbiotic sediment communities are at higher risk for succumbing to the invasive seaweed *Caulerpa* (Gribben *et al.*, 2017). Thus, it is valuable to have baseline information about microbial communities associated with different benthic environments and understand how alterations to those communities may impact the ecosystem as a whole.

1.4 Microbial Natural Products and Marine Chemical Ecology

In addition to gaining further insight into microbial community members, another important avenue of research investigates the ecological roles of bacterial natural products, also

known as secondary or specialized metabolites. These chemical compounds, which include potent antibiotics and cytotoxins (Smith and Nicolaou, 1996), are well known from laboratory cultures, yet their production and function in nature remain poorly understood. Some secondary metabolites have been linked to nutrient acquisition (siderophores- Butler, 2005), communication (quorum sensing-Waters and Bassler, 2005), and defense (feeding deterrence- Burkepile *et al.*, 2006), but the vast majority of work has been focused on the extraction and elucidation of novel, bioactive compounds for drug discovery. In fact, relatively few marine natural products have been directly linked to a defined ecological role (e.g. Gil-Turnes *et al.*, 1989; Wietz *et al.*, 2013).

The great microbial richness observed in marine sediments is thought to host untapped bioactive compound diversity. The obligate marine actinomycete *Salinispora*, which has been isolated from tropical and sub-tropical sediments (Jensen and Mafnas, 2006), dedicates ~10% of its genome to the production of secondary metabolites (Penn *et al.*, 2009). Interestingly, two *Salinispora* species have been shown to exhibit an ecological trade-off in regard to the production of secondary metabolites. *Salinispora arenicola* contains almost twice the number of secondary metabolite biosynthetic gene clusters compared to *S. tropica* and exhibits inhibitory activity linked to secondary metabolite production. In contrast, *S. tropica* grows faster and inhibits growth of competitors via nutrient depletion (Patin *et al.*, 2015). Chemical extracts from *Salinispora* strains have also been shown to alter sediment microbial communities in laboratory mesocosms, indicative of the important role these small molecules may be playing in sediment communities (Patin *et al.*, 2017).

Recent culture-independent work has demonstrated that marine sediments exhibit high biosynthetic potential and are unique in their diversity of operational biosynthetic units when compared to seawater and soil (Bech *et al.*, 2020). Interestingly, results from Bech *et al.* (2020)

suggest that low abundance organisms may be disproportionate in their biosynthetic potential. Rare taxa, also known as the rare biosphere, are thought to fulfill important ecological roles and disproportionately influence biogeochemical cycles, thus impacting overall ecosystem functioning (Lynch and Neufeld, 2015; Jousset *et al.*, 2017). However, despite the importance of the rare biosphere, many culture-independent studies focus solely on the dominant taxa present in the environment. Thus, it would be valuable to consider the entire microbial community when evaluating the role secondary metabolites may be playing in structuring these complex communities.

1.5 Environmental Metabolomics

Along with improvements in sequencing techniques, recent improvements in chemical instrumentation and associated bioinformatic tools has led to the emergence of environmental metabolomics as a method to gain insight into the potential ecological roles of secondary metabolites. With environmental metabolomics, chemical extracts of environmental samples can be used to assess the suite of chemical compounds present within a sample (the metabolome). These can be visualized using web tools like the Global Natural Products Social Molecular Networking (GNPS) platform (Wang *et al.*, 2016) in conjunction with Cytoscape (Shannon *et al.*, 2003). Programs have also been developed for statistical comparisons of metabolomics datasets such as MetaboAnalyst (Chong *et al.*, 2019). Comparative metabolomic studies have provided insight into a variety of systems including the metabolic response of *Staphylococccus aureus* to different antibiotic treatments (Schelli *et al.*, 2017), the role of bacteria in phytoplankton metabolism (Paul *et al.*, 2013), how variation in allelopathic potency of dinoflagellates may impact bloom dynamics (Poulin *et al.*, 2018) and how chemical diversity and distribution of cyanobacteria

can be used to target sampling sites for natural product discovery (Luzzatto-Knaan *et al.*, 2017). Recent work has also highlighted the use of non-targeted metabolomics in environmental systems to track shifts in organic matter chemotypes of seawater and the introduction of anthropogenic compounds with rain events (Petras *et al.*, 2021).

1.6 Linking Microbes and Metabolites

Improvements in sequencing and metabolomic techniques have facilitated paired studies into microbe-metabolite interactions, primarily in human systems. For example, multi-omic studies indicate that dietary compounds and phytochemicals can affect the abundance of bacteria in the human gut, potentially altering host metabolism (Tang *et al.*, 2019) and the fecal metabolome has been shown to explain almost 70% of the gut microbial composition, making it a potential tool to functionally profile the human gut (Zierer *et al.*, 2018). As of yet, less work has been done outside of humans, but the research that has been done indicates growing insight into connections between microbes and metabolites. For instance, a longitudinal survey of wine fermentations revealed that both the microbes associated with grapes and the final wine metabolome can be indicative of the viticultural area and individual vineyard and further suggests that grape microbiota at harvest may be used to predict final wine composition (Bokulich *et al.*, 2016). Additionally, a paired microbiome-metabolome supervised analysis was used to correlate the production of specific peptides with a Fusobacteriaceae strain of bacteria associated with *Chaetodon lunulatus* butterfly fish that are able to avoid gill ectoparasites (Reverter *et al.*, 2020). For marine sediments, the construction of a causeway coincided with changes to both the sediment microbiome and metabolome (Soliman *et al.*, 2017) while the abundance of *Salinispora* in

Belizean sediments was correlated with staurosporine, a known metabolite produced by *Salinispora* (Tuttle *et al.*, 2019).

However, one of the common caveats with microbial community datasets is that taxonomic composition is relative within a sample. Additionally, both microbial community and metabolomic datasets frequently require normalization techniques, such as the removal of zeros, for statistical evaluation (Li, 2015). Microbiome and metabolome datasets can encompass thousands of microbes and metabolites per sample, which results in computationally intensive methods to extract meaningful information (Li, 2015). Consequently, analysis methods for omics datasets are constantly being updated and improved. Methods used to tackle paired omic questions include multivariate tests such as PERMANOVA and Mantel tests (Xia and Sun, 2017), neural network approaches including mmvec (Morton *et al.*, 2019), and matrix integration techniques like mixOmics (Rohart *et al.*, 2017).

While multi-omic approaches are still relatively new, they can be harnessed to ask interesting questions about how microbes and metabolites are correlated in different systems and in response to changing variables. And as the technologies continue to improve, environmental datasets can be revisited to extract further meaning and generate hypotheses for subsequent work. Given the ecological importance of marine sediments (Snelgrove *et al.*, 1997; Baker *et al.*, 2021) and the lack of fundamental knowledge on how marine sediment communities and their metabolites interact, I sought to evaluate these communities and assess the dynamics at play in this dissertation.

1.7 Overview of the Dissertation

The overarching goal of this thesis is to assess bacterial diversity in tropical marine sediments and evaluate linkages between the bacterial community and the metabolome present in these environments. This was done through the use of 16S rRNA community sequencing, metagenomics, metabolomics and a variety of bioinformatic techniques. The dissertation is divided into three main research chapters as follows:

In Chapter 2, I assessed the diversity of bacterial communities in marine sediments around Carrie Bow Cay, Belize using both culture-dependent and culture-independent methods. To evaluate culture-dependent diversity, two media types were employed to culture bacteria from sediment samples associated with five different sites. Rather than hand-picking only visible bacterial colonies, I extracted culture plates in their entirety and used next-generation sequencing methods to evaluate the culture-dependent community. Sediments were also directly extracted and sequenced to evaluate bacterial diversity in a culture-independent manner. By using both culture-dependent and culture-independent techniques, I was able to determine culturing efficiency and assess the strengths and limitations to both methods. Chapter 2 is currently in review at *Environmental Microbiology* and thus remains largely unchanged from its manuscript form in this dissertation.

Chapter 3 seeks to explore not only the bacterial diversity present in tropical marine sediments, but to also assess how metabolomic profiles vary across five sites in Fiji and establish connections between the communities, metabolites and sediment characteristics. For this chapter, I employed a 1 m^2 quadrat that was subdivided into sixteen sections at each site in addition to 10 m transects at two selected sites. By collecting sediment samples using this spatial scheme, I was able to probe spatial differences among marine bacterial communities. Chapter 3 also

introduces the small molecule *in situ* resin capture (SMIRC) method designed in the Jensen Laboratory to capture the secreted sediment metabolome at each site. Additionally, a suite of sediment characteristics was assessed including sediment grain size and organic content. Bioinformatic techniques were then used to determine connections between the microbial communities, the detected metabolites, and the sediment characteristics. This chapter is currently being written into a manuscript for publication.

Chapter 4 builds on the methodology and research from Chapter 3 to gain further insight into how bacterial communities and their metabolites vary across space and in relation to the benthic environment. The SMIRC method was modified to enhance spatial resolution in the metabolomics data making for a valuable paired dataset to explore connections between sediment bacterial communities and metabolomes. Eight sites were assessed across a 1 m^2 quadrat and surrounding 4 m transects for both the bacterial community and the metabolome. Additionally, a sample subset was analyzed with a metagenomic pipeline to gain insight into the ecological functioning of the sediment community. This chapter is currently being written into a manuscript for publication.

Finally, chapter 5 discusses the significant findings and general conclusions brought forth through this dissertation. Future directions related to this research are discussed.

Chapter 2

Microbial diversity in tropical marine sediments assessed using culture-dependent and culture-independent techniques

2.1 Abstract

The microbial communities associated with marine sediments are critical for ecosystem function yet remain poorly characterized. While culture-independent (CI) approaches capture the broadest perspective on community composition, culture-dependent (CD) methods can capture low abundance taxa that are missed using CI approaches. The aim of this study was to assess microbial diversity in tropical marine sediments collected from five shallow water sites in Belize using both CD and CI approaches. CD methods captured approximately 3% of the >800 genera detected across the five sites. Additionally, 39 genera were only detected using CD approaches revealing rare taxa that were missed with the CI approach. Significantly different communities were detected across sites, with rare taxa playing an important role in the delineation of sediment communities. This study provides important baseline data describing shallow water sediment microbial communities and evidence that standard cultivation techniques may be more effective than previously recognized.

2.2 Introduction

Advances in sequencing technologies and bioinformatics have led to major improvements in our ability to assess the diversity and distributions of environmental microbes (bacteria and archaea) (Lynch and Neufeld, 2015; Hug *et al.*, 2016; Thompson *et al.*, 2017). The application of

culture-independent (CI) methods has transformed our understanding of microbial diversity while metagenome assembled genomes and single cell genomics have provided insight relevant to functional traits in yet to be cultured organisms (Kalisky and Quake, 2011; Evans *et al.*, 2015; Parks *et al.*, 2017). Despite these advances, the microbial diversity associated with marine sediments remains poorly characterized relative to other major biomes such as soil and seawater (Lloyd *et al.*, 2018; Martiny, 2019; Baker *et al.*, 2021). Sediment microbial communities are diverse (Thompson *et al.*, 2017), densely populated, (Dale, 1974; Musat *et al.*, 2006), play integral roles in fundamental ecosystem processes (Snelgrove *et al.*, 1997; Baker *et al.*, 2021), and can exhibit extraordinary levels of fine-scale spatial structure (Probandt *et al.*, 2018). While the inaccessibility of deep-sea sediments (>200 m depth) may contribute to the lack of data, the communities associated with shallow euphotic (≤200 m depth) water sediments also remain poorly described (Baker *et al.*, 2021). Given that marine sediments cover ~70% of the earth's surface (Parks and Sass, 2009), baseline information describing sediment microbial diversity provides an important mechanism to understand community structure over time and across environmental gradients.

The gains afforded by CI diversity estimates can overshadow the intrinsic value of microbial cultivation. Culture-dependent (CD) methods provide opportunities to assess microbial metabolism and contributions to ecosystem function in ways that cannot be achieved using CI approaches. For instance, cultivation of *Nitrospira* provided critical insight into the first bacterium known to perform complete nitrification (Daims *et al.*, 2015) while culture-dependent research with *Thermosulfidibacter takaii* ABI70S6^{T} resulted in the discovery of a reversible TCA cycle that was not detected with metagenomics (Nunoura *et al.*, 2018). Additionally, testing for inhibition among marine *Vibrio* strains revealed that competition is greater between than within ecologically

cohesive populations (Cordero *et al.*, 2012) while CD work on two closely related species of *Salinispora* demonstrated ecological trade-offs in competitive strategies (Patin *et al.*, 2015). Culturing techniques have been developed based on metabolic requirements inferred using CI techniques (Tripp *et al.*, 2008). Thus, these approaches can offer complementary insights into microbial ecology, with the general observation that CI techniques provide more comprehensive taxonomic coverage while CD methods can provide clearer taxonomic resolution (Orphan *et al.*, 2000; Chen *et al.*, 2008; Shivaji *et al.*, 2011; Vaz-Moreira *et al.*, 2011; Dickson *et al.*, 2014). While both techniques have been used in tandem, surprisingly few studies have focused on major environmental biomes such as marine sediments. Of note, the importance of media specificity and rare taxa when comparing CI and CD approaches was recently reported (Pédron *et al.*, 2020).

Recent bioinformatic comparisons of CD versus CI bacterial diversity have revived discussion of the "great plate count anomaly" and the canonical theory that fewer than one percent of bacterial taxa have been cultured (Martiny, 2019; Steen *et al.*, 2019). These studies provide contrasting views on the improvements that have been made in culturing relative to the proportions of bacteria that remain uncultured (Martiny, 2019; Steen *et al.*, 2019). Advances such as these highlight the need to reassess frequently cited paradigms describing bacterial culturability, including the great plate count anomaly. While the majority of bacterial taxa have yet to be cultured (Lloyd *et al.*, 2018), the development of innovative techniques in combination with persistent use of traditional methods has led to the successful cultivation of notable microbes including bacterioplankton in the SAR11 clade (Rappé *et al.*, 2002; Henson *et al.*, 2018), the first Asgard archaea representative '*Candidatus* Prometheoarchaeum syntrophicum' (Imachi *et al.*, 2020) and three Saccharibacteria (TM7) species with their Actinobacteria host (Cross *et al.*, 2019),

suggesting that many if not most microbes can ultimately be brought into the laboratory (Lewis *et al.*, 2020).

The aim of this study was to assess microbial diversity in marine sediments using both CD and CI techniques. CD sample communities were determined by sequencing environmental plates inoculated from sediments while CI sample communities were determined by direct sequencing of sediments. Through the use of next-generation sequencing, we determined that culturing efficiency was 1-2% based on the number of 16S rRNA amplicon sequence variants (ASVs) detected and 3-4% in terms of the number of genera detected. Additionally, the CD method detected 39 genera that were not detected using the CI approach, highlighting the importance of culturing for capturing rare members of the community. These results emphasize the importance of both CD and CI methods for assessments of microbial diversity in marine sediments.

2.3 Methods

2.3.1 Sample Collection and Processing

In September 2015, divers collected marine sediment samples from five locations around the Smithsonian Field Station at Carrie Bow Cay, Belize (Table 2.1). At each site, five replicate Whirl-Pak® (Nasco) bags were filled with sediment from a 3 m^2 area. Upon return to the field station, 20 ml of wet sediment from each Whirl-Pak® was transferred into 50 ml falcon tubes with 20 ml of RNAlater® and stored at 4°C before transport on dry ice to Scripps Institution of Oceanography (SIO) where they were stored at -40°C prior to DNA extraction. Two types of media prepared at SIO were used for on-site, culture-dependent sample processing: seawater agar (SWA) comprised of 16 g agar and 1 liter natural seawater, and 50% marine agar (MA) comprised of 0.5 g yeast extract, 2.5 g peptone, 16 g agar, and 1 liter natural seawater. Both media contained the

antifungal agent cyclohexamide at a final concentration of 200 μg/ml. Freshly collected sediment samples were diluted 1:2 with autoclaved seawater in 4 ml vials, vigorously shaken, and further diluted 1:10 and 1:100 after which 50 μl of each dilution (1:2 1:10, 1:100) was inoculated onto each medium, spread with a sterile glass rod, and the plates allowed to dry in a laminar flow hood then sealed with parafilm. This resulted in a total of 150 plates (3 dilutions x 5 replicate sediments x 2 media x 5 locations), which were transported back to SIO. In an effort to facilitate colony growth but prevent one or two bacteria from swamping the plates, we kept the plates at room temperature for eight days post inoculation and then stored them at -40°C prior to DNA extraction.

Two sets of control samples for the CD method were generated in January 2020. To control for DNA contamination present in the media or reagents, plates made with both of the experimental medium types were inoculated with a known amount (~$3x10^7$ cells) of *Vibrio coralliilyticus,* then immediately parafilmed and stored at -40°C. To assess the approximate read counts that might be expected from the initial inoculation without subsequent colony growth, dilution series of local sediment (San Diego, CA) were plated following the experimental plate methods. Three replicate dilution series were plated for each medium, each was spiked with a known amount (~$3x10^7$ cells) of *Vibrio coralliilyticus* to ensure adequate DNA concentrations for sequencing, parafilmed immediately to prevent colony growth, and stored at -40°C.

Table 2.1: Site information for samples collected around Carrie Bow Cay, Belize.

Site	Latitude (N)	Longitude (W)	Depth (m)	Water Temperature (°C)	Habitat Description
1	16° 47' 54.06"	88° 04' 59.7"	8	27	Spur and groove. Area with patches of sand, low relief reef and a patch a seagrass.
2	16° 48' 11.46"	88° 04' 41.46"	20	27	Reef slope. Large sand area with scattered reef.
3	16° 49' 32.76"	88° 06' 24.48"	1	25	Entrance to Twin Cays mangroves. Samples collected in a sandy patch surrounded by a seagrass bed.
4	16° 47' 43.56"	88° 07' 20.94"	6	30	Sea mound in the lagoon between fringing reef and mainland.
5	16° 48' 10.44"	88° 04' 55.5"	1	Not collected	Near field station dock. Sand/rubble.

2.3.2 DNA Extraction

Environmental DNA (eDNA) was extracted from approximately 1 g of freshly thawed sediment per sample following physical (bead beating) and chemical (phenol-chloroform) DNA extraction methods (Patin *et al.*, 2013). One replicate from site 5 was lost resulting in a total of 24 sediment samples processed for CI analysis. DNA extractions were performed in duplicate for each sediment sample (2 g of sediment in total extracted per sample) and the extracts combined prior to purification. For the CD analyses, agar plates were left to thaw at room temperature for 30 min and 3 ml molecular grade water was added to the surface. A heat sterilized metal loop was used to scrape the surface of each plate and the resulting suspension pipetted into a 15 ml falcon tube. The three dilutions plated for each sediment were combined into a single falcon tube and centrifuged at 8,000 RPM (9,803 RCF) and 4°C for 5 minutes generating 50 samples (5 replicate sediments x 2 media x 5 locations). The supernatant was removed and the bottom 2 ml including

the cell pellet were distributed into two ceramic bead-beating tubes prior to DNA extraction following the protocol applied to the sediments (Patin *et al.*, 2013). All control samples were extracted using the same protocol described above, however due to samples being lost in transit to the sequencing facility, minimal DNA remained for subsequent sequencing. As a result, replicates for each control type were pooled, thus producing four control samples: a SWA blank control spiked with *V. coralliilyticus,* a SWA inoculum control spiked with *V. coralliilyticus,* a MA blank control spiked with *V. coralliilyticus,* and a MA inoculum control spiked with *V. coralliilyticus.*

2.3.3 PCR and Sequencing

The v4 region of the 16S rRNA gene was PCR amplified using the primers 515F (TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG GTGYCAGCMGCCGCGGTAA) and 806Rb (GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGACTACNVGGGTWTC TAAT) (Caporaso *et al.*, 2012). PCR was performed following the Phusion Hot Start Flex 2x Master Mix protocol with an annealing temperature of 60°C and 1 µl of 5 ng/µl DNA. Products were cleaned using ExoSap-IT® before adding Nextera XT (Illumina) indices and sequencing adapters with the following PCR program: 98°C for 1 min followed by five cycles of 98°C for 10 sec, 65°C for 20 sec and 72°C for 30 sec with a final extension at 72°C for 2 min. Gel electrophoresis was used to confirm the presence of a PCR product of the predicted size. Sequences were normalized based on DNA concentration, pooled and cleaned with AMPure XP beads. The purified library was then sent to the Institute for Genomic Medicine (IGM), University of California, San Diego (UCSD) for sequencing on an Illumina MiSeq v2 500 cycle at a depth of 130k reads per sample. For controls, DNA was sent to Novogene (South Plainfield, NJ) for library

preparation following their proprietary methods and sequencing with the above described primers (515F & 806Rb) on an Illumina NovasSeq.

2.3.4 Analysis

Raw sequences were imported into QIIME2-2020.2 (Bolyen *et al.*, 2018) and denoised using the DADA2 (Callahan *et al.*, 2016) denoise-paired pipeline with an input of p-trim-left-f of 19 and p-trim-left-r of 20 to remove primers. Based on the raw files, p-trunc-len-f and p-trunc-len-r were set to 250 and 155 base pairs respectively and chimeras were removed with the default consensus method. In an effort to control for background inoculum that remained present but was not actively growing, all amplicon sequence variants (ASVs) associated with the controls were quantified to their relative read abundances (Appendix A). After accounting for ASVs associated with *Vibrio*, the most abundant ASV present was identified as an unknown bacterium with 67 reads in one of the control inoculum samples (Appendix A). Since all other ASVs associated with the inoculum controls, outside of *Vibrio* which was used as a spike-in, contained less than 67 reads, we opted to set a minimum read threshold of 70 reads per feature for each CD sample before proceeding with subsequent analyses. This threshold removed approximately 50% of ASVs associated with CD samples (Table 2.2). Additionally, we checked each ASV associated with the four control samples and determined that none of the ASVs remained in the experimental samples after applying the filtration step. Taxonomy was then assigned using the SILVA v132 database (Quast *et al.*, 2013) and samples were subsequently filtered to remove chloroplast and mitochondria sequences (Table S3) in QIIME2-2020.2.

Table 2.2: Number of amplicon sequence variants (ASVs) remaining after filtration steps were applied.

	Sample Type				
Filtration Step	Control Blank (n=2)	Control with Inoculum (n=2)	Culture-Dependent MA (n=25)	Culture-Dependent SWA (n=25)	Culture-Independent (n=24)
Unfiltered	14	38	562	709	27342
ASVs removed with inoculum control threshold	14	38	284	346	NA
Chloroplast & mitochondria ASVs	0	0	0	1	1,091
Final Number of ASVs	0	0	278	345	26,251

In order to assess how similar ASV sequences were to cultured representatives, an approach based on Steen *et al.* (2019) was employed. Using the align.seqs command in Mothur (Schloss and Westcott, 2011), both CD ASVs and CI ASVs were aligned with sequences from the SILVA database. To compare to cultured strains, type strains [T] and cultured s[C] strains were searched for in SILVA and downloaded as an aligned fasta file including gaps. Since SILVA v138.1 was the database used for sequence extraction, the corresponding non-redundant full library was also downloaded (SILVA_138.1_SSURef_NR99_tax_silva_full_align_trunc.fasta). The align.report files were then filtered to remove any sequences with inadequate alignments (pairwise alignment lengths <250). Histograms based on similarity to the nearest cultured representative and

scatterplots with the nearest cultured relative compared to the most similar sequence in the full library were generated using ggplots2 (Wickham, 2016) in R. Some of the ASVs were then further interrogated by using the NCBI BLAST tool by selecting to exclude uncultured sequences.

QIIME2-2020.2 was used to perform alpha and beta diversity analyses at the ASV level after rarefying samples to a depth of 62,830 reads. Associated statistical analyses were also performed with QIIME2-2020.2 at the ASV level using the non-parametric Kruskal-Wallis test (Kruskal and Wallis, 1952) for method and site comparisons in relation to alpha diversity indices (Faith's Phylogenetic Diversity and Pielou's Evenness) and multivariate PERMANOVA tests with 999 permutations (Anderson, 2001) for beta diversity (both weighted and unweighted UniFrac (Lozupone and Knight, 2005)) comparisons across methods and sites. ANCOM analysis (Mandal *et al.*, 2015), which was specifically designed to address compositional microbial data, was performed at the genus level using QIIME2-2020.2 to determine genera that significantly differed across sites in culture-independent samples. Figures were generated using QIIME2-2020.2, RStudio version 3.6.2 (R Core Team, 2019), ggplot2 (Wickham, 2016), nVennR (Quesada, 2020), and Excel version 16.36.

2.3.5 Data Availability

All raw sequences files are available through NCBI's Sequence Read Archive (SRA). Accession numbers for CI files are SAMN08824420 – SAMN0882443, CD files are SAMN15932210 – SAMN15932259 and controls are SAMN15932260 – SAMN15932263.

2.4 Results

2.4.1 Culture-Dependent (CD) Sediment Microbial Diversity

Across all five sites, CD alpha diversity analyses revealed on average 54 ± 3 16S rRNA amplicon sequence variants (ASVs) for sediments plated on seawater agar (SWA) and 39 ± 2 ASVs for the same sediment samples plated on marine agar (MA) (Figure 2.1a). SWA yielded significantly higher phylogenetic richness (Figure 2.2a; Faith's PD Kruskal-Wallis H=8.870, p=0.003) and greater evenness (Figure 2.2b; Pielou's Kruskal-Wallis H=7.645, p=0.006) when compared to the ASVs detected on MA.

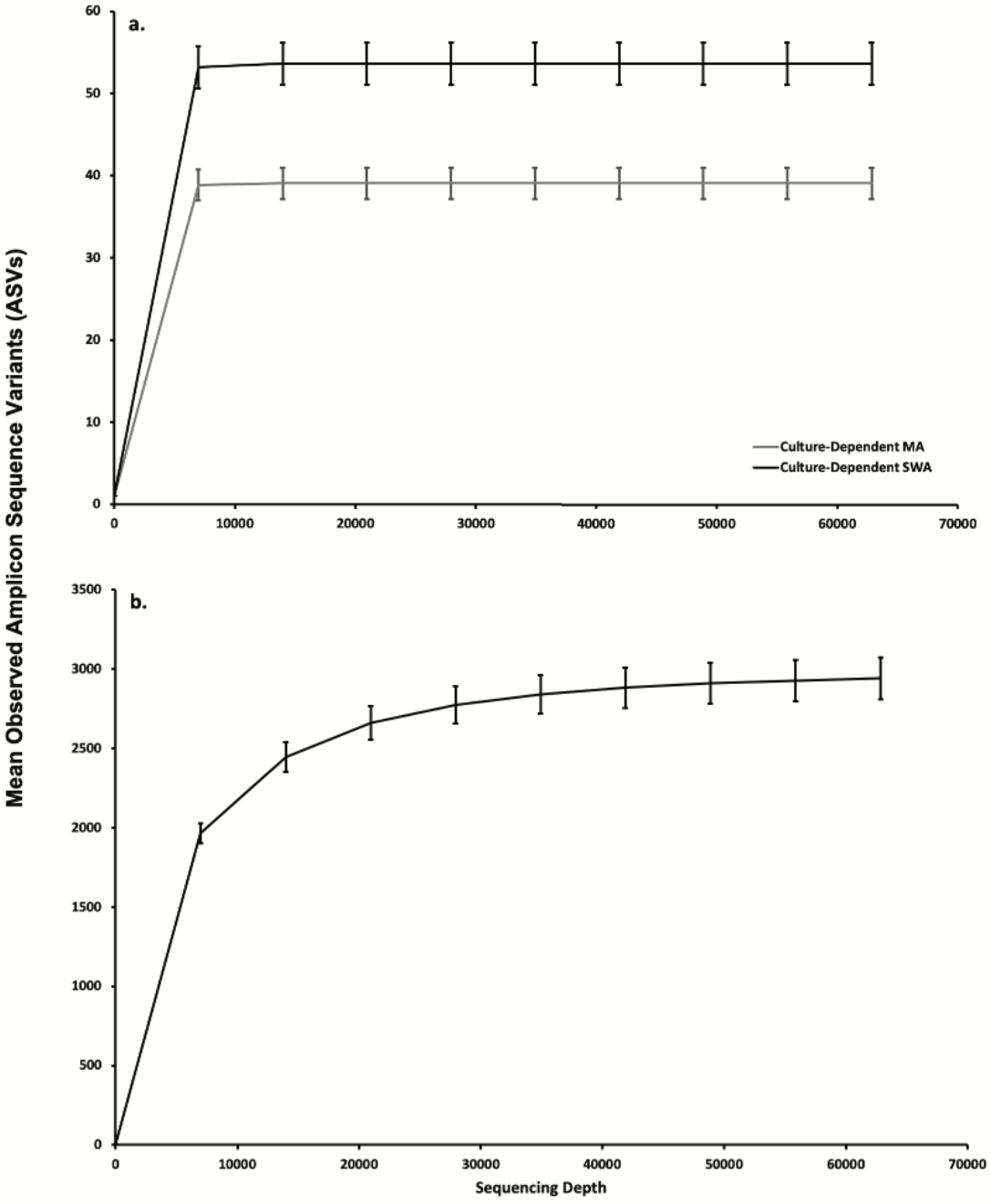

Figure 2.1: Mean alpha rarefaction curves across sediment samples from five sites in Belize. a) Culture-dependent results obtained using seawater agar (SWA) and marine agar (MA) media and b) Culture- independent results. Error bars represent standard error among replicates.

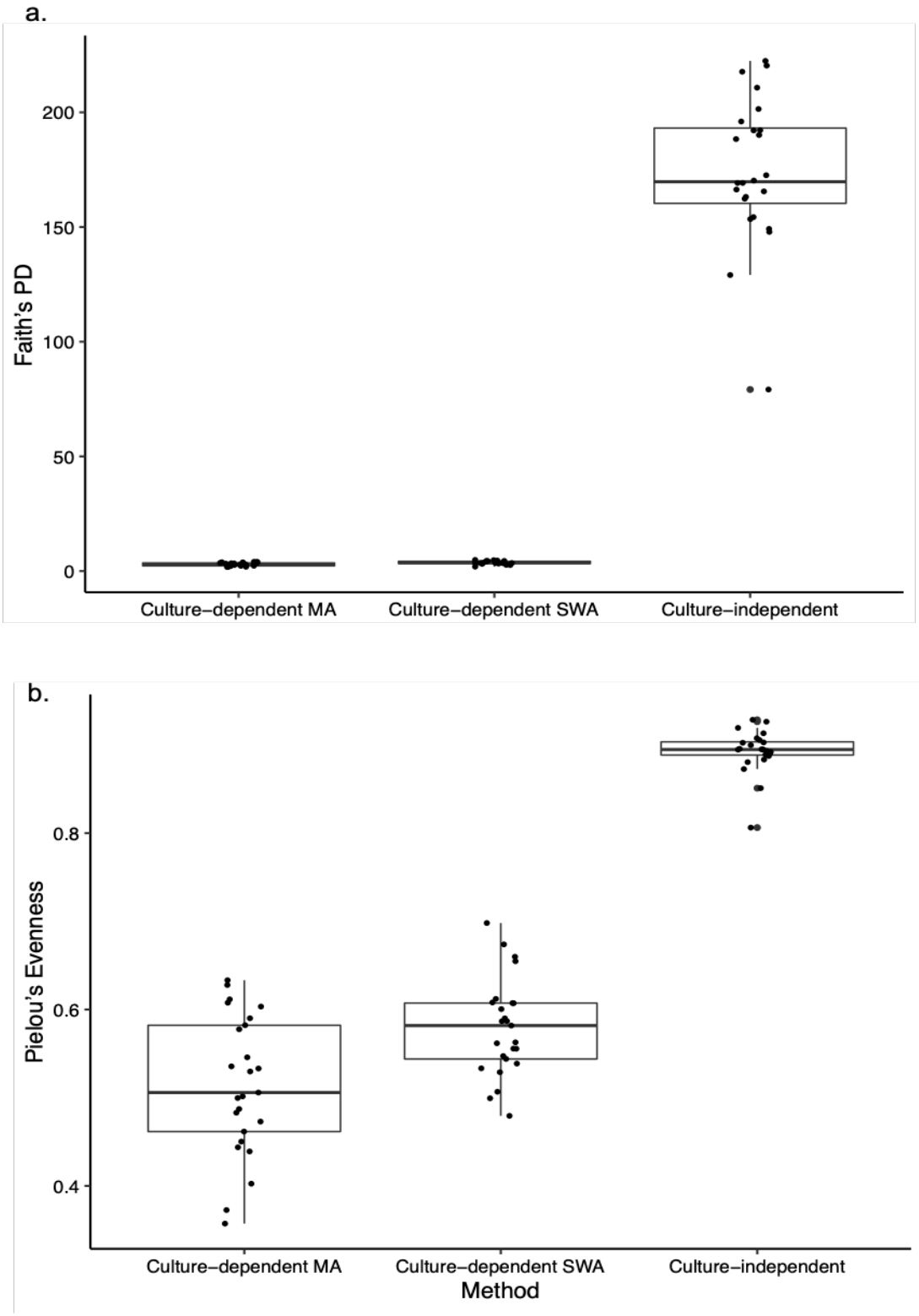

Figure 2.2: Alpha diversity boxplots of marine sediment microbial communities from Carrie Bow Cay, Belize in culture-dependent and culture-independent samples determined using a) Faith's Phylogenetic Diversity Index and b) Pielou's Evenness.

The microorganisms detected on the SWA and MA plates were classified into five bacterial phyla (Figure 2.3). The three most abundant phyla were similar for both media, with Proteobacteria being the most abundant, averaging 80% and 96% of the communities on SWA and MA plates, respectively. Bacteroidetes represented 20% of the SWA communities compared to ca. 1.6% for MA, indicating that the low nutrient medium was more selective for this phylum. On MA, Firmicutes was the second most abundant phylum at ca. 2.6%. In comparison, Firmicutes averaged ca. 0.13% on SWA, suggesting that the nutrient rich MA media better selects for this phylum. The other two phyla detected in culture were Epsilonbacteraeota (formerly Epsilonproteobacteria) and Actinobacteria, both of which averaged <1% of the community in SWA and MA (Figure 2.3).

At the genus level, the ASVs detected on SWA were assigned to 102 genera while those detected on MA were assigned to 76 genera. CD methods identified 128 different genera, of which 50 were detected using both SWA and MA while 52 were unique to SWA and 26 were unique to MA (Figure 2.4). *Vibrio* represented the most abundant genus observed on both media, accounting for 27% and 44% of the sequences detected on SWA and MA, respectively. Other relatively abundant genera cultured included *Ruegeria* (SWA 25.4% & MA 35.1%), *Persicobacter* (SWA 18.8% & MA 0.7%), *Microbulbifer* (SWA 6.0% & MA 3.5%), and *Alteromonas* (SWA 3.9% & MA 1.7%). Most taxa detected with CD methods averaged <1% of the community, including 92 SWA assigned genera and 62 MA assigned genera.

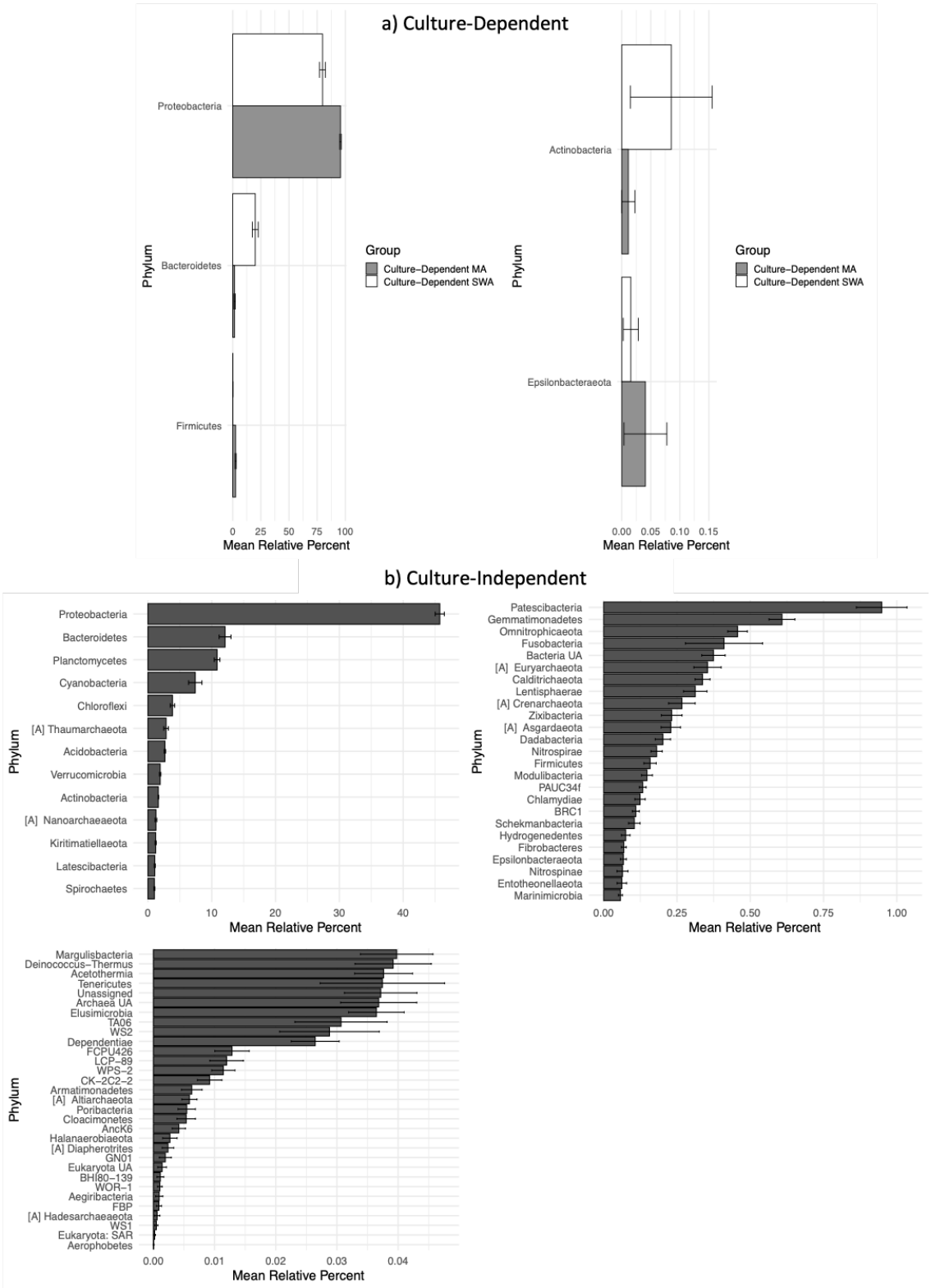

Figure 2.3: Rank abundance of phyla detected using a) culture-dependent methods and b) culture-independent methods. Light gray bars represent taxa associated with marine agar (MA) while white bare represent seawater agar (SWA) samples. Archaea are denoted as [A] and error bars indicate standard error.

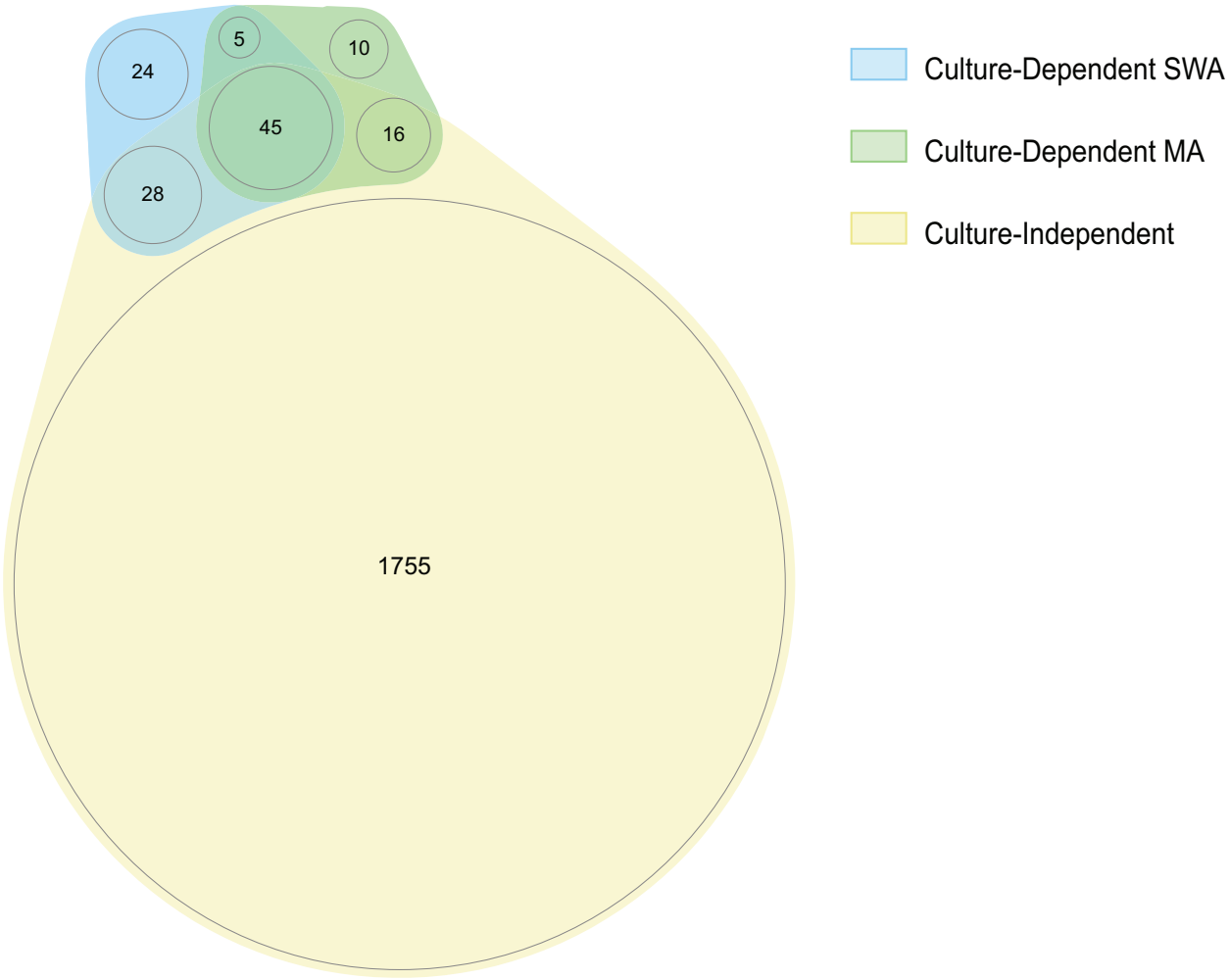

Figure 2.4: Proportionally scaled Venn diagram of microbial genera detected using culture-dependent (SWA and MA) and culture-independent methods.

2.4.2 Culture-Independent (CI) Microbial Diversity

CI diversity for the same five sites was considerably greater than what was detected using the CD techniques, averaging 2,942 ± 133 ASVs (Figure 2.1b). CI communities included 68 phyla (57 Bacteria, 9 Archaea and 2 Eukarya) (Figure 2.3b), with about half of the sequences assigned to Proteobacteria. After Proteobacteria, the relatively most abundant phyla were Bacteroidetes, Planctomycetes and Cyanobacteria. On average across all samples, 55 of the 68 phyla detected (81%) represented less than 1% of the relative community, and thus can be considered rare. When

combined, these rare taxa averaged ~7% of the total community, indicating their importance to community composition. More than 15 phyla represent either candidate phyla or unannotated taxa, indicating that microbial diversity at the phylum level remains poorly characterized in marine sediments (Figure 2.3b).

In total, 1,844 genera (1,728 Bacteria, 113 Archaea and 3 Eukarya) were observed using the CI technique (Figure 2.4), with the communities across all five sites displaying fairly rich (Figure 2.5a; Faith's PD range 79.15-222.88) and even distributions (Figure 2.5b; Pielou's evenness index range 0.81-0.93). The most commonly observed genus was *Woeseia*, which averaged 4.96% of the relative community across all five sites. Of the 1,844 genera identified, only 16 averaged ≥1% of the community. Many CI ASVs were not annotated at the genus level or were annotated as "uncultured," suggesting they belong to poorly described taxa. Those that averaged ≥1% of the community and could be identified at the genus level included *Woeseia* (4.96%), *Xenococcus* (2.49%), *Zeaxanthinibacter* (2.18%), Candidatus *Nitrosopumilus* (1.65%), *Pleurocapsa* (1.46%), *Chroococcidiopsis* (1.10%), and *Rhodopirellula* (1.07%). Minor (≤0.001 average relative percent), non-target amplification of eukaryotic sequences was observed. These were annotated as unassigned eukaryotic or ciliate associated.

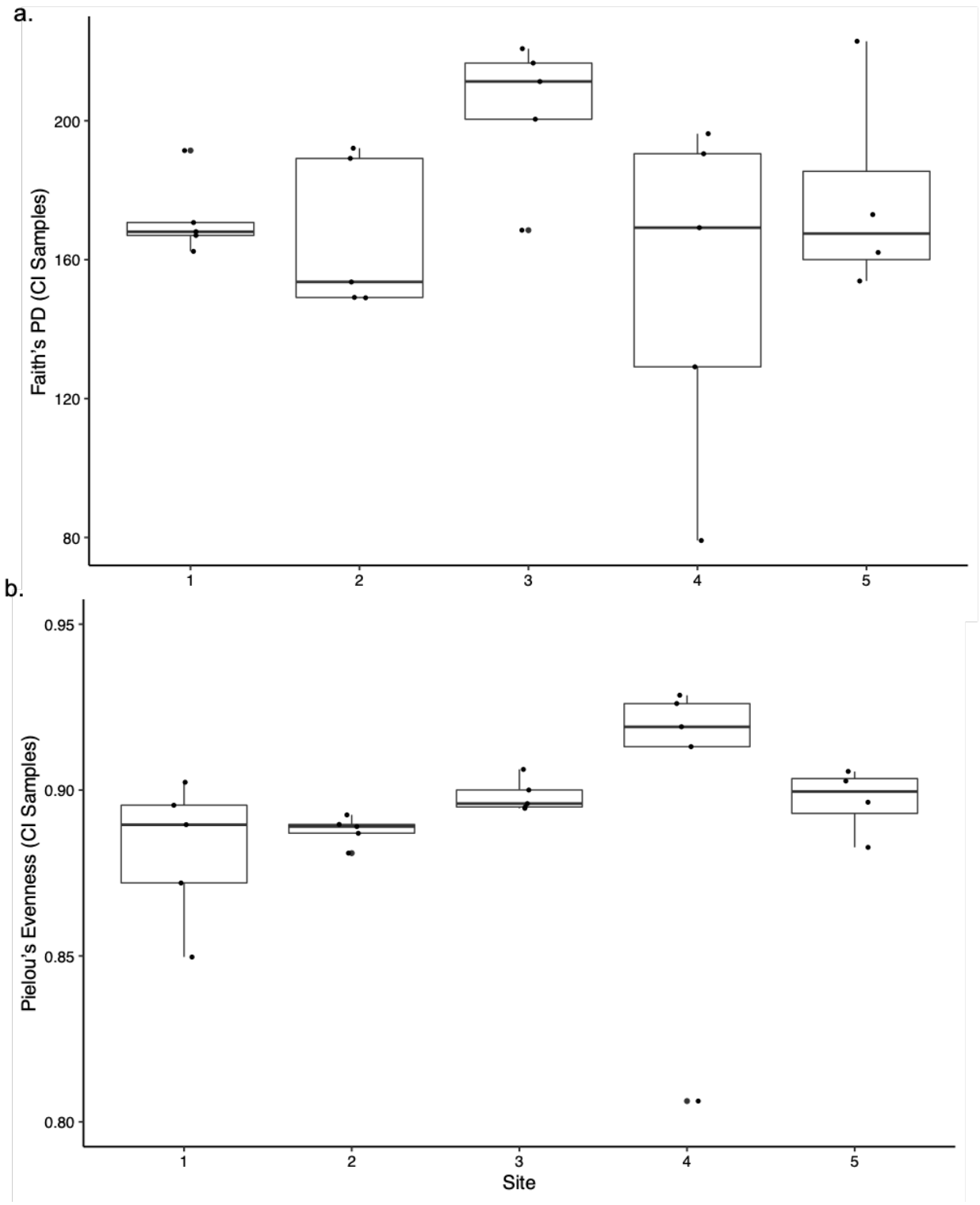

Figure 2.5: Culture-independent alpha diversity boxplots of marine sediment microbial communities from across five sites around Carrie Bow Cay, Belize determined using a) Faith's Phylogenetic Diversity Index and b) Pielou's Evenness.

2.4.3 Method Comparisons

In total, ASVs were annotated to 1,883 genera across all methods used in this study (Figure 2.4). As expected, the CI communities were significantly richer (Figure 2.2a; Faith's PD Kruskal-Wallis $H=52.076$ $p<0.001$) and included 1,755 genera that were only detected using this approach (Figure 2.4). An additional 89 genera were detected using both the CI and CD approaches. Of those 89 genera, 45 were detected on both media types, 28 only on SWA and 16 only on MA (Figure 2.4). Surprisingly, 39 genera detected in culture were not detected using the CI technique (Figure 2.4; Table 2.3).

Of the 39 genera that were only detected with CD methods, 24 of these genera were specific to SWA, ten were specific to MA, and five were detected on both SWA and MA (Figure 2.4). All of the genera uniquely detected using the CD technique were relatively rare, with 37 averaging <1% of the community (Table 2.3). The two most relatively abundant genera detected in culture but not using the CI technique were *Celeribacter*, which averaged 1.92% and 0.05% of the SWA and MA communities, respectively, and *Halomonas*, which averaged 0.71% and 1.14% of the SWA and MA communities, respectively. Taxa specific to the SWA method included Proteobacteria, Bacteroidetes and Actinobacteria such as *Marinomonas, Mesonia* and *Streptomyces*. Taxa only detected via MA cultures included members of the Firmicutes, Bacteroidetes and Proteobacteria such as *Fictibacillus, Taeseokella* and *Fangia* (Table 2.3). While the vast majority of taxa detected using CD methods have previously been isolated from marine sources, both the SWA and MA yielded genera that do not have named marine species. Interestingly, for all but one of these genera, closely related strains have been detected in marine samples (Table 2.3).

Table 2.3: Amplicon sequence variants (ASVs) annotated to the genus level detected using culture-dependent methods only (n= number of detected replicates). If the genus was not annotated (UA), the lowest taxonomic rank is indicated. NA = not applicable. If the genus has named species from the marine environment, at least one example is cited.

Genus	Site(s)	Medium (n)	Mean relative percent MA ± standard error	Mean relative percent SWA ± standard error	Named species reported from marine environments
Corynebacterium 1	2	MA (1)	1.14E-02 ± 1.14E-02	NA	Yes (Ben-Dov *et al.*, 2009)
Microbacteriaceae UA	1	SWA (1)	NA	6.14E-03 ± 6.14E-03	Yes (Lee, 2008)
Agrococcus	1	SWA (1)	NA	5.45E-03 ± 5.45E-03	Yes (Lee, 2008)
Kocuria	1	SWA (1)	NA	5.91E-02 ± 5.91E-02	Yes (Kim *et al.*, 2004)
Streptomyces	5	SWA (3)	NA	1.20E-02 ± 7.85E-03	Yes (Gallagher and Jensen, 2015)
Luteivirga	5	SWA (1)	NA	2.37E-03 ± 2.37E-03	Yes (Haber *et al.*, 2013)
Pontibacter	5	SWA (1)	NA	4.53E-03 ± 4.53E-03	Yes (Nedashkovskaya *et al.*, 2005)
Taeseokella	5	MA (1)	3.04E-02 ± 3.04E-02	NA	Yes (Li *et al.*, 2019)
Mesoflavibacter	3	SWA (1)	NA	3.42E-03 ± 3.42E-03	Yes (Asker *et al.*, 2007)
Mesonia	3	SWA (1)	NA	1.43E-02 ± 1.43E-02	Yes (Choi *et al.*, 2015)
Pseudofulvibacter	1	MA (1)	6.12E-03 ± 6.12E-03	NA	Yes (S. H. Yang *et al.*, 2016)
Robertkochia	4	SWA (1)	NA	2.17E-03 ± 2.17E-03	Yes (Hameed *et al.*, 2014)
Salegentibacter	5	SWA (1)	NA	1.84E-03 ± 1.84E-03	Yes (Nedashkovskaya *et al.*, 2006)
Zunongwangia	5	MA (1)	4.50E-03 ± 4.50E-03	NA	Yes (Shao *et al.*, 2014)
Fictibacillus	5	MA (1)	4.58E-02 ± 4.58E-02	NA	Yes (Dastager *et al.*, 2014)

Table 2.4: Amplicon sequence variants (ASVs) annotated to the genus level detected using culture-dependent methods only, Continued.

Genus	Site(s)	Medium (n)	Mean relative percent MA ± standard error	Mean relative percent SWA ± standard error	Named species reported from marine environments
Halobacillus	5	MA (1)	3.64E-03 ± 3.64E-03	NA	Yes (Teasdale *et al.*, 2009)
Paenibacillus	5	MA (1)	7.88E-03 ± 7.88E-03	NA	Yes (Lee *et al.*, 2013)
Staphylococcus	2	MA (1)	5.33E-03 ± 5.33E-03	NA	Yes (Arora, 2013)
Skermanella	1	SWA (1)	NA	1.50E-02 ± 1.50E-02	No marine isolates, but see *Maldonado et al 2009 (Maldonado *et al.*, 2009)
Brevundimonas	1	SWA (1)	NA	7.22E-03 ± 7.22E-03	Yes (Fritz *et al.*, 2005)
Aureimonas	1	SWA (1)	NA	6.46E-02 ± 6.46E-02	**No
Hoeflea	1	SWA (1)	NA	8.32E-03 ± 8.32E-03	Yes (Biebl *et al.*, 2006)
Nesiotobacter	3	MA (1) & SWA (1)	2.48E-02 ± 2.48E-02	1.62E-02 ± 1.62E-02	Yes (Kumar *et al.*, 2019)
Amaricoccus	5	SWA (1)	NA	2.61E-03 ± 2.61E-03	No marine isolates, but see ***Pohlner et al 2019 (Pohlner *et al.*, 2019)
Celeribacter	1 & 5	MA (2) & SWA (3)	4.90E-02 ± 3.57E-02	1.92E+00 ± 1.91E+00	Yes (Lee *et al.*, 2012)
Paracoccus	1	SWA (1)	NA	5.24E-03 ± 5.24E-03	Yes (Z. P. Liu *et al.*, 2008)
Sulfitobacter	3	SWA (1)	NA	3.53E-03 ± 3.53E-03	Yes (Park *et al.*, 2007)
Thalassobius	2 & 4	MA (1) & SWA (1)	2.74E-03 ± 2.74E-03	2.75E-03 ± 2.75E-03	Yes (Yi and Chun, 2006)

Table 2.5: Amplicon sequence variants (ASVs) annotated to the genus level detected using culture-dependent methods only, Continued.

Genus	Site(s)	Medium (n)	Mean relative percent MA ± standard error	Mean relative percent SWA ± standard error	Named species reported from marine environments
Blastomonas	3	SWA (1)	NA	2.73E-02 ± 2.73E-02	Yes (Meng *et al.*, 2017)
Sphingomonas	1	SWA (1)	NA	5.43E-03 ± 5.43E-03	Yes (Schut *et al.*, 1997)
Neiella	1, 2 & 4	SWA (3)	NA	1.28E-02 ± 7.43E-03	Yes (Du *et al.*, 2013)
Colwelliaceae uncultured	1, 2, 3, 4 & 5	MA (11) & SWA (16)	8.48E-02 ± 2.88E-02	1.74E-01 ± 8.17E-02	Yes (Jean *et al.*, 2006)
Psychromonadaceae UA	1 & 5	SWA (3)	NA	1.20E-02 ± 8.22E-03	Yes (Li *et al.*, 2013)
Fangia	3	MA (1)	2.90E-03 ± 2.90E-03	NA	Yes (Lau *et al.*, 2007)
Halomonas	1, 2, 3, 4 & 5	MA (15) & SWA (15)	1.14E+00 ± 3.93E-01	7.07E-01 ± 2.57E-01	Yes (L. A. Romanenko *et al.*, 2002)
Marinomonas	1	SWA (1)	NA	3.53E-02 ± 3.53E-02	Yes (Romanenko *et al.*, 2009)
Nitrincolaceae UA	2	SWA (1)	NA	2.77E-03 ± 2.77E-03	Yes (Arahal *et al.*, 2007)
Oleibacter	5	SWA (1)	NA	1.46E-03 ± 1.46E-03	Yes (Teramoto *et al.*, 2011)
Psychrobacter	2, 3 & 4	MA (3)	9.14E-03 ± 5.96E-03	NA	Yes (L. a Romanenko *et al.*, 2002)

*Maldonado et al. 2009 isolated a strain from the marine environment where the closest 16S hit was *Skermenella* (Maldonado *et al.*, 2009).
**_Aureimonas_ is the sister genus of *Aurantimonas* (Rathsack *et al.*, 2011). *Aurantimonas* that has been isolated from marine sources, however it has also been identified as a common contaminant (Rathsack *et al.*, 2011; Salter *et al.*, 2014).
***All *Amaricoccus* species have been isolated from sludge, but Pohlner et al. (2019) also found OTUs that hit to *Amaricoccus* from marine sediment (Pohlner *et al.*, 2019).

Not surprisingly, the CI approach identified numerous taxa with few or no cultured representatives. For example, the phylum Latescibacteria (aka WS3), which is commonly detected in CI studies (Youssef *et al.*, 2015; Farag *et al.*, 2017; Lloyd *et al.*, 2018), and to the best of our knowledge does not have a cultured representative, was also identified as ca. 1.07% of the CI communities. Additionally, bacteria from the widely distributed and diverse phyla Acidobacteria, Patescibacteria and Gemmatimonadetes averaged 2.7%, 0.95, and 0.6% of the relative CI community respectively, but few strains from these phyla have been cultured (Hugenholtz *et al.*, 2001; Ward *et al.*, 2009; DeBruyn *et al.*, 2011; Soro *et al.*, 2014; Lemos *et al.*, 2019).

Both CD and CI ASVs were also assessed in relation to their nearest cultured relative in the SILVA database (Figure 2.6). Not surprisingly, many of the ASVs shared greater similarity with sequences from other CI studies than with known type or cultured strains (Figure 2.6). ASVs from CD samples had a median similarity of 99.6% with cultured representatives, but there were a few distinct outliers (Figure 2.6). One ASV from CD samples was identified with SILVA as an uncultured Chitinophagales and shared only ~79% similarity with a cultured representative. When that ASV was assessed with the NCBI BLAST database, the most similar sequence was a Muribaculaceae bacterium with ~85% sequence similarity originally isolated from a pig gut. The other ASV from CD samples that was <90% similar to any culture in the SILVA database, matched with a recently cultured sponge microbe called *Xanthovirga aplysinae* (99.6% similar) (Goldberg *et al.*, 2020). There were also a handful of sequences between 90-95% similar to cultured sequences. Subsequent searching with BLAST confirmed the lower similarity for some, while others had >99% similarity to recent isolates from marine sources such as holothurians and corals. In contrast to the CD samples, the median sequence similarity for CI samples was 85.4% (Figure 2.6). Many of the ASVs <60% similar were identified as archaea with SILVA. Of those, some

shared greater similarity (~80%) with members of the Microgenomates group of bacteria in the NCBI database.

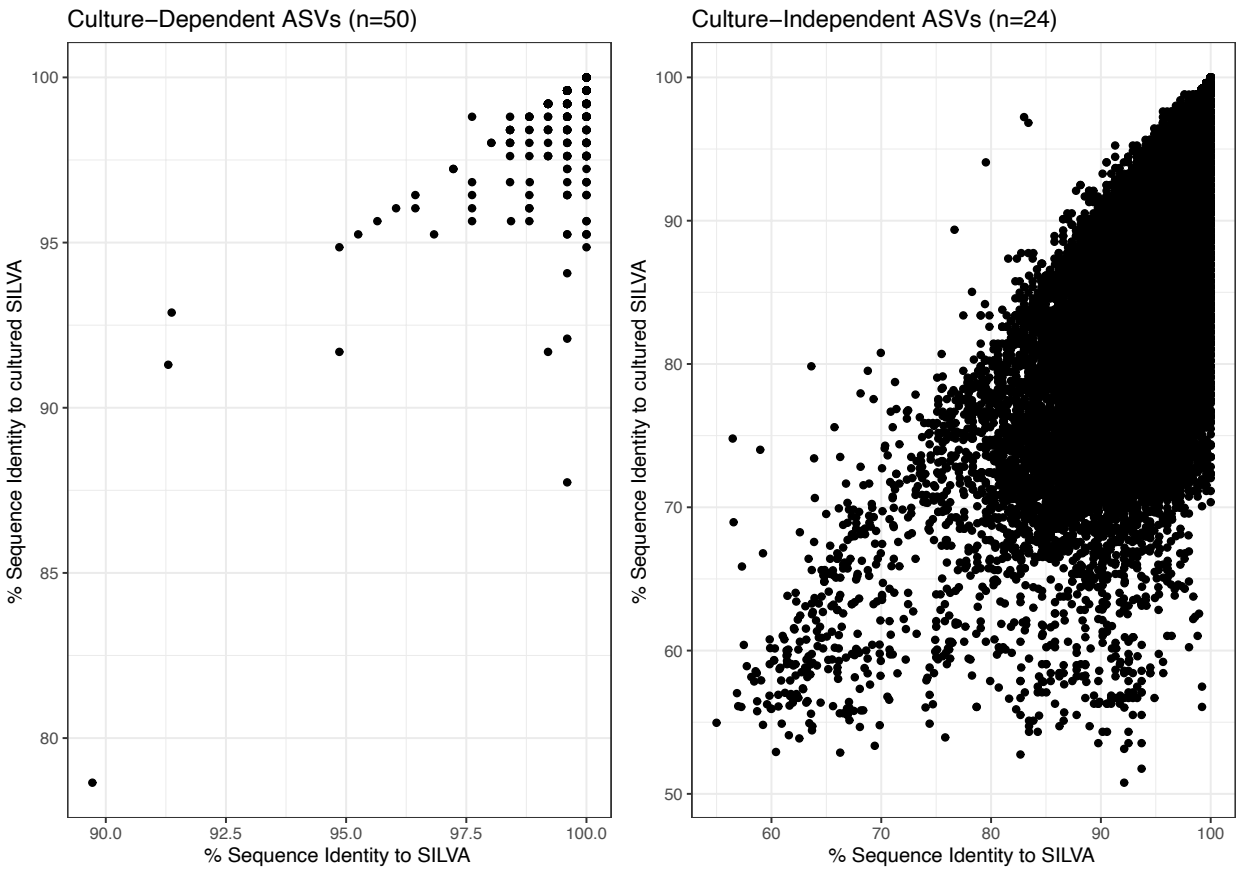

Figure 2.6: Scatterplot of amplicon sequence variant (ASV) similarity from culture-dependent and culture-independent samples. ASVs were compared to the nearest cultured or type strain extracted from the SILVA v138.1 database (y-axis) versus the nearest sequence identified with the full non-redundant SILVA v138.1 database (x-axis).

When assessing beta diversity and the presence/absence of ASVs in each community, the CD communities were found to be significantly different than CI communities (Figure 2.7a; Unweighted UniFrac PERMANOVA pseudo-F=39.826, p=0.001). There was one CI replicate that was distinct from the rest of the CI samples (Figure 2.7), likely due to the large (>20%) fraction of an unknown genus of Flavobacteriales present within that sample. When considering sample site, cultured communities showed no pattern while CI communities seemed to cluster by site

(Figure 2.7b). Weighted beta diversity analysis also showed significant separation between CD and CI communities (Figure 2.7c; Weighted UniFrac PERMANOVA pseudo-F=125.15, p=0.001) and between the nutrient rich (MA) and nutrient poor (SWA) media types (Figure 2.7c). Since there appeared to be clustering based on sites within the CI communities (Figure 2.7b & d), only CI samples were considered for subsequent CI community visualizations and comparisons.

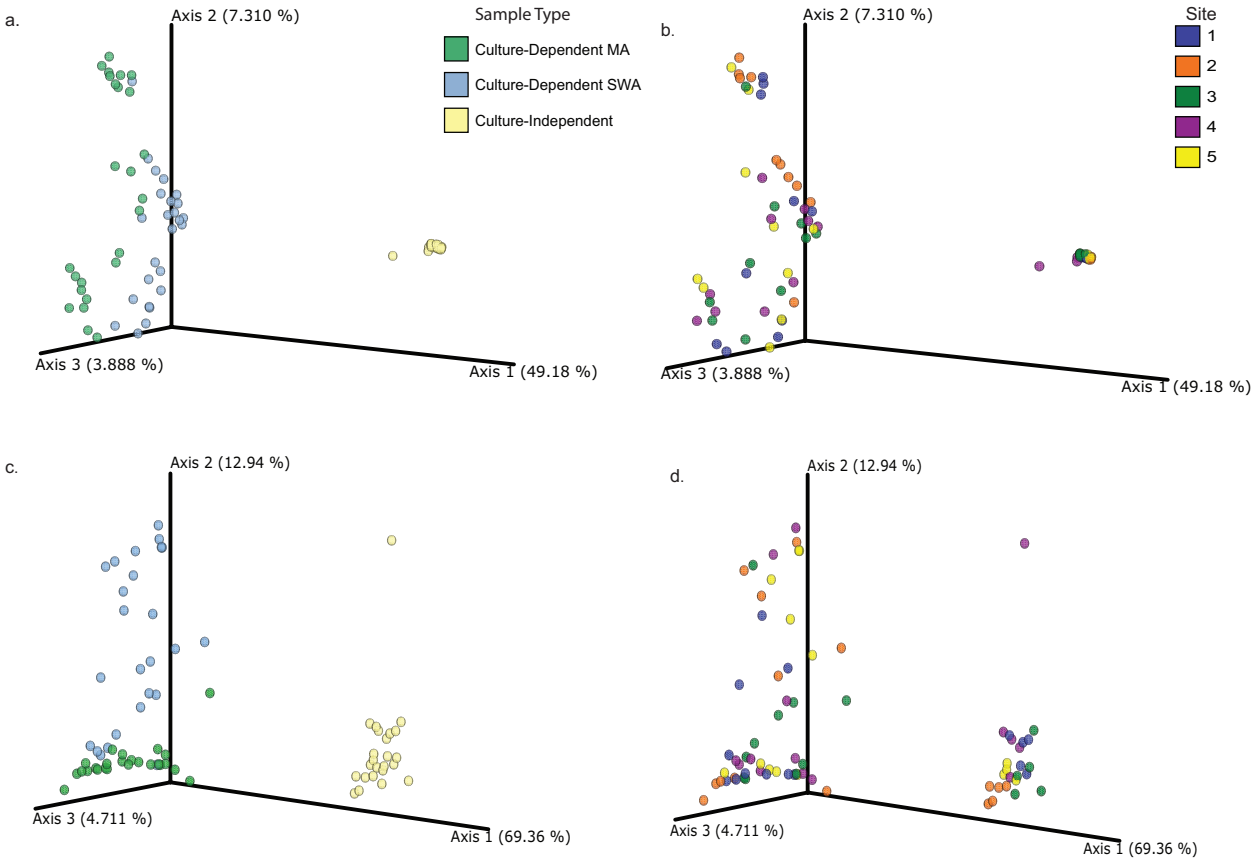

Figure 2.7: UniFrac analysis of culture-independent and culture-dependent sediment microbial communities visualized via principle coordinates analysis (PCoA). Unweighted results colored by a) sample type and b) site. Weighted results colored by c) sample type and d) site.

2.4.4 Culture-Independent (CI) Site Comparisons

The five sites sampled were within 5 km of each other (Table 2.1) but represented different habitats: an 8 m deep spur and groove reef (site 1), a 20 m deep reef slope (site 2), a 1 m deep sand patch in a seagrass bed at the mouth of a mangrove island (site 3), a 6 m deep seamount in a lagoon (site 4), and a 1 m deep sand and rubble patch at the marine station's dock (site 5). Both weighted and unweighted UniFrac analyses on the CI communities indicate a significant difference among site communities (Figure 2.8; Unweighted UniFrac PERMANOVA pseudo-F=2.982, p=0.001; Weighted UniFrac PERMANOVA pseudo-F=7.987, p=0.001).

Site 3 appeared the most distinct from the other communities in the unweighted UniFrac (Figure 2.8a) and showed higher richness than the four other sites (Figure 2.5a). Some of the taxonomic differences between sites included relatively more Proteobacteria at site 2, relatively fewer Cyanobacteria at sites 3 and 4, and about twice the relative amount of Chloroflexi at site 3 when compared to the other sites. Additionally, an analysis of composition of microbiomes (ANCOM) done at the genus level identified seven taxa that were significantly different by site, six of which were related to site 3. These were identified as *Marixanthomonas*, an unidentified genus of the BD2-7 (Family Cellvibrionales), an uncultured gammaproteobacterium that was relatively more abundant at site 3, and two taxa that were absent from site 3 (*Stanieria*, an unknown genus of Xenococcaceae, and un uncultured MBAE14 gammaproteobacterium).

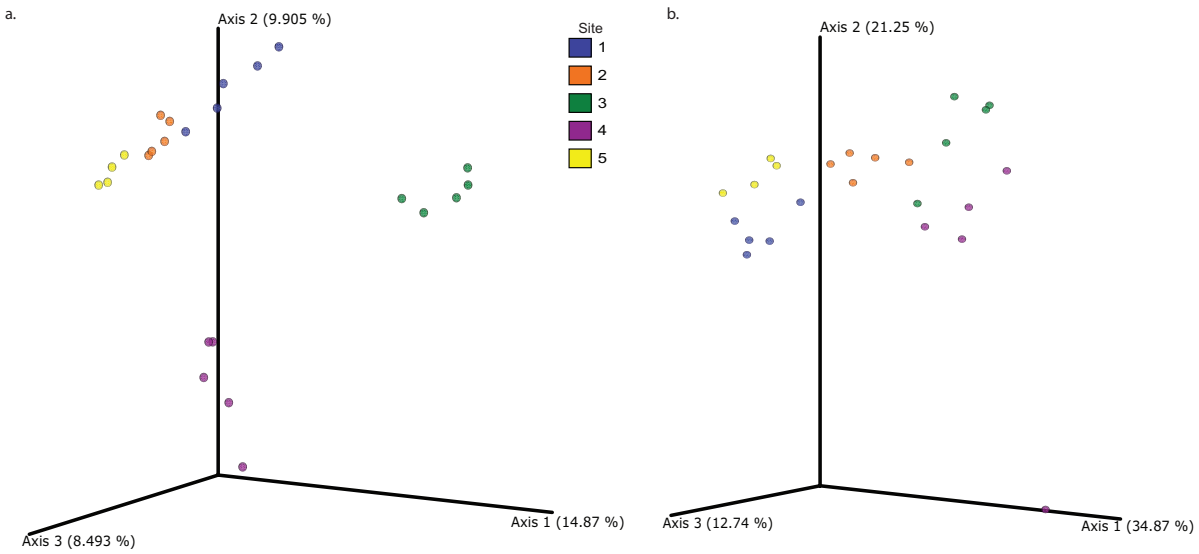

Figure 2.8: Unifrac analysis of culture-independent sediment microbial communities across five sites. a) Unweighted UniFrac and b) weighted UniFrac. Communities visualized via principal coordinates analysis (PCoA).

2.4.5 Culturing Efficiency

We assessed culturing efficiency in two ways. First, we quantified percent efficiency by determining the total number of ASVs cultured vs. the number identified using CI techniques. The alpha diversity rarefaction curves based on observed ASVs revealed that saturation was achieved using both techniques (Figure 2.1), thus ensuring that effective comparisons could be made. The number of cultured ASVs represented 1.82% and 1.33% (for SWA and MA, respectively) of the ASVs detected using the CI approach. Second, we calculated the overlap in taxonomic assignment from domain down to ASV using the two approaches (Table 2). Based on taxonomic assignment, 3.95% of the genera detected using the CI approach were also detected in culture on SWA medium while 3.31% were detected on MA. Thus, while the amount of diversity cultured will vary with the method used, our results generally exceeded the <1% value typically reported in association with the great plate count anomaly (Staley and Konopka, 1985; Martiny, 2019; Steen *et al.*, 2019).

Table 2.6: Percent cultured across taxonomic levels. Percent cultured calculations were based on taxonomic overlap in culture-dependent (CD) and culture-independent (CI) samples. Percent taxa unique to CD samples represents the number of taxa only identified with CD methods as a proportion of the total taxa identified at the corresponding level.

	Percent Cultured		Percent Taxa Unique to Culture-Dependent Samples	
Taxonomic Level	**SWA**	**MA**	**SWA**	**MA**
Domain	33.33	33.33	0	0
Phylum	7.35	7.35	0	0
Class	2.39	3.19	0	0
Order	2.91	2.75	0.46	0.15
Family	3.36	2.76	0.51	0.34
Genus	3.95	3.31	1.54	0.80
Species	3.16	2.64	1.77	1.22
ASV	0.39	0.33	0.86	0.68

2.5 Discussion

It is widely recognized that CD approaches are inadequate for measuring microbial diversity. Early comparisons of colony counts to cell counts indicated that <1% of environmental bacteria were cultured, a phenomenon referred to as the "great plate count anomaly" (Razumov, 1932; Staley and Konopka, 1985). The great plate count anomaly was the first indication that a majority of environmental bacteria were not readily cultured using standard techniques, which has subsequently driven numerous studies seeking to improve culturability (Kaeberlein *et al.*, 2002; Tamaki *et al.*, 2009; Tanaka *et al.*, 2014; Rygaard *et al.*, 2017). The advent of sequence-based CI methods brought the extent of uncultured microbial diversity into better perspective (Lynch and Neufeld, 2015; Hug *et al.*, 2016; Lloyd *et al.*, 2018). But while high-throughput amplicon sequencing methods, such as 16S rRNA gene surveys, are valuable tools for assessing microbial communities, there are biases and limitations with these methods. For instance, DNA extraction

(Brooks *et al.*, 2015), primer selection (Fischer *et al.*, 2016; Laursen *et al.*, 2017; Wear *et al.*, 2018; Willis *et al.*, 2019), PCR amplification (Brooks *et al.*, 2015) and template concentration (Kennedy *et al.*, 2014) have all been shown to impact microbial community profiles and the use of the 16S rRNA gene as a taxonomic marker can miss up to 10% of environmental sequences (Eloe-Fadrosh *et al.*, 2016). Additionally, variables used for sequence analysis, such as operational taxonomic unit clustering vs. denoising affect diversity estimates (Patin *et al.*, 2013; Callahan *et al.*, 2017; Edgar, 2017, 2018; Nearing *et al.*, 2018; Straub *et al.*, 2019) and resulting data interpretations are further complicated by the compositional nature of amplicon data (Gloor *et al.*, 2017).

Given advances in both CD and CI techniques, it is surprising though that few studies have sought to reassess bacterial diversity estimates using both approaches (but see some example comparisons in river sediment: (Pédron *et al.*, 2020), lake sediment: (Elfeki *et al.*, 2018), seawater: (Rygaard *et al.*, 2017), cheese: (Perin *et al.*, 2017), and lungs: (Dickson *et al.*, 2014)). Here, we used CD and CI techniques to explore the microbial communities in marine sediments and estimate culturing efficiency by sequencing bacteria directly from agar plates as opposed to the more traditional approach of isolating or counting colonies. The 16S rRNA gene sequences amplified from sediment eDNA required ~50,000 reads, post quality control and denoising, to approach saturation in the alpha rarefaction curves. This indicates that relatively deep sequencing is needed to capture the microbial diversity present in these complex communities. A recent study comparing global diversity of marine sediments detected over 34,000 ASVs associated with bacteria and archaea across 299 sites (Hoshino *et al.*, 2020). In total, our study found over 27,000 ASVs from CD and CI samples around Carrie Bow Cay, Belize. One major difference between the global study (Hoshino *et al.*, 2020) and ours is sequencing depth. While Hoshino et al. (2020) were able to do an impressive global analysis on a large number of samples, none of their rarefaction curves

reached saturation, suggesting that much of the rare biosphere was missed. Given the richness levels detected at fine scales in sediment (Probandt *et al.*, 2018) and that the majority of taxa identified in our sediments were in low relative abundance (<1%), deep sequencing is needed to detect ecologically relevant members of the rare biosphere and to fully profile sediment communities (Lynch and Neufeld, 2015; Jousset *et al.*, 2017).

As expected, CD richness estimates were considerably lower, averaging 54 and 39 ASVs per sample for seawater agar (SWA) and marine agar (MA) respectively, compared to >2,000 ASVs per CI sample. While we did not perform colony counts, the number of ASVs detected appeared higher than the number of colonies readily visualized by eye on the plates after eight days of incubation, suggesting that much of the diversity detected may have arisen from micro-colonies that could not be easily counted or isolated using standard practices. This result likely helps account for the observation that culturing efficiency, as defined by the number of detected ASVs, on SWA and MA relative to the CI results was 1.82 and 1.33%, respectively, or 3.15% in total. Given that only two cultivation media were used, these results support a re-evaluation of the 1% culturability paradigm. It would also be beneficial to perform colony counts in parallel to sequencing plates to help delineate methodological differences in assessing the great plate count anomaly and to confirm the presence of micro-colonies in culture. Pairwise approaches have been successfully used in some studies (e.g. Perin *et al.*, 2017; Rygaard *et al.*, 2017; Elfeki *et al.*, 2018; Pédron *et al.*, 2020), but not yet in marine sediments to our knowledge. Culturability varies depending on the ecosystem and complexity of the community, so additional studies across systems and with varying methodologies will be beneficial for proper re-assessment of the great plate count anomaly.

We were also interested in assessing culturing efficiency based on the taxa detected in culture compared to CI samples. While, we determine the taxonomic overlap between CD and CI methods from domain to ASV, we focused on the genus level given limitations with species assignment when using short regions of the 16S rRNA gene (Liu *et al.*, 2008; Větrovský and Baldrian, 2013; Yang *et al.*, 2016; Johnson *et al.*, 2019). When assessing culturing efficiency based on the genera detected with both CD and CI methods, culturability increased to 3.95 and 3.31% on SWA and MA respectively (and increased to 5.42 and 4.04% if including genera only identified with CD methods). While this highlights that cultivation may in fact be more effective than the original experiments of the great plate count anomaly, the lack of marine sediment representatives in culture was apparent given the high frequency at which our taxonomy was denoted as "uncultured" or unassigned. While we would expect culturability to increase with taxonomic level, surprisingly we found that the culturing percentage remained fairly consistent (~3%) until the phylum (~7%) and domain level (~33% since no Archaeal or Eukaryotic sequences were detected in culture). This finding supports previous research documenting the limited knowledge on marine sediment microbial diversity (Baker *et al.*, 2021).

While we further explored differences in communities at the genus level, it is important to note that there is considerable intra-genus diversity in microbes (Cordero *et al.*, 2012; Patin *et al.*, 2015), thus limiting the conclusions we can draw from studies such as ours based on a short fragment of the 16S rRNA gene. Additionally, given functional variation within microbial lineages, assessing communities at different taxonomic levels would provide differing insight. For instance, Hoshino *et al.* (2020) evaluated ASV differences in relation to the phylum level which can provide some insight into broad functional patterns (e.g. aerobic vs anaerobic metabolism) but lacks the resolution and/or associated metagenomic data to concretely answer functional questions.

Additionally, taxonomic assignment might vary based on the reference database. For instance, SILVA may only identify an ASV to the order level while a closer NCBI match would provide taxonomy to the genus level. Additionally, if taxonomy is annotated to the same level, reference databased can disagree (Pollock *et al.*, 2018). Thus, it can be difficult to extract ecological meaning from taxonomic community profiles, especially in understudied environments like sediments that lack cultured representatives and genomic data for many lineages (Baker *et al.*, 2021).

Here, we defined cultured taxa as those that were identified after applying a background threshold to remove any ASVs that had fewer than 70 reads in their corresponding samples. The three cultured phyla that occurred in >1% relative abundance were Proteobacteria, Bacteroidetes and Firmicutes. Among these, Proteobacteria also represented the largest fraction of the CI community, indicating that members of the most abundant phylum detected can be readily cultured. While both Bacteroidetes and Firmicutes were detected on SWA and MA, SWA had a considerably larger fraction of Bacteroidetes (~20% compared to ~2%) while MA had more Firmicutes than SWA (~3% compared to ~0.1%), demonstrating how these different media enrich for different taxa. Additionally, the cultivation of significantly greater microbial richness on the relatively nutrient poor medium SWA supports previous results (Watve *et al.*, 2020) and suggests that, even for sediment communities, high nutrient concentrations can be inhibitory. Interestingly, the candidate phylum PAUC34f was detected in every replicate of our CI analyses, yet this phylum remains uncultured to date (Chen *et al.*, 2020). While PAUC34f was not detected in our post-filtering CD data, it was detected in one SWA replicate prior to correcting for the inoculum control. Given that dilution to extinction in a nutrient poor medium led to the cultivation of the ubiquitous SAR11 clade of marine bacterioplankton (Rappé *et al.*, 2002), our decision to pool eDNA obtained

from the dilutions prior to sequencing may have masked evidence for the cultivation of this phylum.

Approximately 20% of the ASVs detected using the CD method could only be assigned to broad groups, such as the Roseobacter clade CHAB-I-5, or remained unclassified at the genus level. Given the accuracy of removing erroneous sequences while obtaining greater taxonomic resolution with denoising pipelines (Callahan *et al.*, 2016, 2017; Amir *et al.*, 2017; Nearing *et al.*, 2018; Prodan *et al.*, 2020), our ASVs without taxonomic assignment likely represent undescribed taxa obtained in culture. Our subtraction of inoculum controls further supports this suggestion. In addition to culturing taxa that could not be assigned to a known genus, we also cultured genera that contained few named species. For instance, the genus *Ascidiaceihabitans* only contains one species, isolated from the tunicate *Halocynthia aurantium* (Kim *et al.*, 2014), yet was detected on six SWA replicates (max ~27% relative percent) and three MA replicates (max ~2% relative percent). We also observed the genus *Endozoicomonas* in two replicates of MA (max ~2% relative percent). *Endozoicomonas* bacteria have been found in association with a wide variety of marine organisms including corals (Bayer *et al.*, 2013), tunicates (Schreiber *et al.*, 2016) and sea slugs (Kurahashi and Yokota, 2007). A recent CI study revealed the potential for diverse functional roles for *Endozoicomonas* as a symbiont while also postulating that this bacterium has a free-living stage based on its large genome size (Neave *et al.*, 2017). Our detection of *Endozoicomonas* in multiple replicates potentially represents a previously uncultured, free-living stage of this taxon.

In addition to separation of the CD vs CI communities observed in beta diversity analyses, we also saw clustering by site in the CI samples, with site 3, a sand patch near a seagrass bed and a mangrove island, seemingly the most distinct. Gribben et al. 2017 found that seagrass sediment communities can reduce the success of an invasive macrophyte, indicating their importance for the

overall health of these ecosystems (Gribben *et al.*, 2017). ANCOM results identified seven taxa at the genus level that significantly differed between sites, with three taxa found to have the highest relative abundance at site 3 and three taxa completely absent from site 3. Further studies are needed to determine what, if any, role these taxa may play in seagrass communities and what physical and chemical characteristics may play a role in structuring these communities. Changes in clustering patterns of CI communities were also observed when using unweighted vs weighted UniFrac. The unweighted UniFrac showed clearer separation across sites in the CI communities. Given that weighted UniFrac accounts for the relative abundances of taxa, this finding suggests that there are commonalities in the dominant community members across the five sites. Additionally, all significant genera identified through ANCOM were rare members of the community, with the maximum relative percent abundance of ~0.2%. Therefore, it appears that rare members of the community are likely driving site differentiation in beta diversity. While we cannot determine functional differences in our sites with this study, previous research has demonstrated the disproportionate role rare microbes may play in communities (Jousset *et al.*, 2017; Bech *et al.*, 2020). Given that the vast majority of taxa identified in our study represented <1% of the relative communities at each site, it would be valuable to explore how rare microbes may functionally impact sediment communities in future work.

One surprising result was the detection of cultured taxa that were not detected using CI methods. Similar results were reported from Mediterranean water samples (Crespo *et al.*, 2016) and recent work in a freshwater system (Pédron *et al.*, 2020). Almost all genera only detected with CD methods would likely be considered members of the rare biosphere. Eight of the genera detected only in CD samples have been previously identified as laboratory contaminants including *Corynebacterium, Kocuria, Paenibacillus, Brevundimonas, Hoeflea, Paracoccus, Sphingomonas,*

Acinetobacter, and *Psychrobacter* (Salter *et al.*, 2014)*,* however they all also have named species from the marine environment. Given the identification of marine representatives and our filtering methods, we believe these taxa to be members of our communities rather than contaminants. To the best of our knowledge, *Aureimonas* is the only genus detected exclusively from CD methods that has not been previously reported from marine samples. It was detected on one MA replicate from Site 1 and represented 1.62% of that community. *Aureimonas* is a sister genus to *Aurantimonas*, which does include marine representatives (Rathsack *et al.*, 2011), suggesting that the ASV might have been misannotated due to high sequence similarity. Alternatively, this may be the first report of the genus *Aureimonas* from the marine environment. Importantly, the CD methods revealed ASVs that could not be identified to the genus level, or were simply annotated as "uncultured", suggesting that new taxa had been cultured using these relatively simple approaches.

One of the potential limitations in using environmental enrichment methods is the difficulty in isolating pure cultures. Bacteria may fail traditional isolation attempts for a variety of reasons including obligate associations with co-occurring microbes or metabolic needs that are not met when the strain is isolated from the community. Thus the use of new techniques such as reverse genomics (Cross *et al.*, 2019) and metagenomics in conjunction with traditional methods can aid in culturing efforts (Lewis *et al.*, 2020). Additionally, innovative culturing techniques that utilize environmental conditions such as diffusion chambers (Bollmann *et al.*, 2007) and the ichip (Berdy *et al.*, 2017) also provide a means to increase culturing success. It would be interesting to further interrogate the unique taxa we detected in our culturing approach to determine if they could be isolated using traditional methods such as dilution to extinction, traditional methods with altered medium (e.g. (Rygaard *et al.*, 2017) or if new techniques might prove more fruitful. One of the

divergent strains detected in CD samples was subsequently identified as *Xanthovirga aplysinae.* Given the recent description of this genus, isolated with traditional methods on MA (Goldberg *et al.*, 2020), it is likely that continued culturing efforts in underexplored environments will yield novel diversity. It is also possible that our threshold for defining culturability was not stringent enough due to differences between sample and control sediments, in which case further efforts should focus on taxa that were detected in relatively higher abundances across multiple cultures.

In conclusion, the sediment microbial communities analyzed here were highly diverse, with the majority of genera representing rare (<1%) members of the community. The unique detection of taxa using CD methods supports the value of these techniques in conjunction with CI methods to assess community composition. Additionally, it may be valuable to further explore variation in microbial communities across spatial scales to ensure adequate sampling and diversity assessments.

2.6 Acknowledgements

I would like to thank Robert Tuttle for assisting in the collection of samples and Gregory Amos and Alexander Chase for bioinformatic advice. Additionally, I would like to acknowledge Jessica Blanton and Eric Allen for providing sequencing indices and for assistance in library construction. I would also like to acknowledge the Smithsonian's Carrie Bow Cay Field Station. This publication is a contribution to the Caribbean Coral Reef Ecosystems (CCRE) Program. Sequencing was conducted at the IGM Genomics Center, University of California, San Diego, La Jolla, CA.

This research was supported by the National Science Foundation Grant No. OCE-1235142 and based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1650112 and the Scripps Fellowship to A.M.D.

Chapter 2 is coauthored with Nastassia V. Patin and Paul R. Jensen and was recently submitted for publication. The dissertation author was the primary investigator and author of this chapter.

Chapter 3

Assessing connections between marine microbial communities, metabolites and sedimentary characteristics across varying spatial scales

3.1 Abstract

Marine sediments host rich and diverse microbial communities that remain woefully understudied. Previous research has found correlations between the microbial communities in marine sediments and various environmental characteristics, but few studies have explored those connections using next-generation sequencing or considered the role of spatial variability in community dynamics. Furthermore, studies of microbial communities rarely assess the associated environmental metabolome which includes compounds responsible for cooperation and competition. Thus, the goal for this project was to utilize 16S rRNA community analysis in conjunction with environmental metabolomics and sediment characteristics to study the microbial community of tropical marine sediments across a 1 m^2 quadrat, 10 m transects and five sites in a 12 km^2 area around Nacula Island, Fiji. Our results highlight that microbial diversity increases with spatial scale. Sedimentary characteristics including nitrate, iron and grain size composition were significantly correlated with microbial communities. Finally, while metabolomic profiles were less distinct than microbial communities, supervised DIABLO analysis was able to identify correlations between microbial genera, metabolites and sediment traits. Thus, this study establishes

baseline information about sediment communities and emphasizes the importance of accounting for spatial heterogeneity when sampling.

3.2 Introduction

Marine sediments host an incredible abundance and diversity of microbes, with cell counts between 10^8 - 10^9 per cm^3 in surface sediments (Dale, 1974; Musat *et al.*, 2006) and global estimates between 2.9 x 10^{29} – 5.39 x 10^{29} cells (Kallmeyer *et al.*, 2012; Parkes *et al.*, 2014). Recent richness estimates for archaea and bacteria in marine sediments range from 3.28 x 10^4 – 2.46 x 10^6 and 7.85 x 10^3 – 6.10 x 10^5 amplicon sequence variants (ASVs) respectively (Hoshino *et al.*, 2020). These benthic microbes are an integral component of the marine ecosystem - acting as key players in global nutrient cycling, the decomposition of pollutants, and as a food source for protozoans, meiofauna and even macrofauna (Snelgrove *et al.*, 1997; Hoshino *et al.*, 2020; Baker *et al.*, 2021). Recently, microbial communities in marine sediments have even been shown to impact the success of invasive macrophytes (Gribben *et al.*, 2017), demonstrating their potential cascading impacts on ecosystems. Despite their importance, much remains unknown about the structure and function of these complex communities.

While biotic processes, such as predation and bioturbation, undoubtedly play an important role in structuring microbial communities (Mermillod-Blondin *et al.*, 2004; Thompson *et al.*, 2017; Deng *et al.*, 2020), abiotic factors are also important. Percent organic carbon and nutrients, such as nitrogen, have been positively correlated with bacterial biomass (Dale, 1974) and a recent survey of global marine sediments found significant correlations between taxonomic composition, sedimentary organic content and dissolved oxygen (Hoshino *et al.*, 2020). Additionally, sediment grain size is inversely correlated with bacterial abundance, likely due to a loss of physical niche

space (Dale, 1974). Indeed, a single grain of sediment was found to host between 10^4-10^5 bacterial cells representing thousands of unique operational taxonomic units (OTUs) (Probandt *et al.*, 2018). In sandy sediments, the majority of bacteria are associated with the sediment grains (Rusch *et al.*, 2003) and results from epifluorescent microscopy on single grains of sand revealed increased bacterial colonization within crevasses of the grain, further suggesting the importance of physical niche space (Probandt *et al.*, 2018).

Secondary metabolites are also thought to be important for structuring microbial communities. Microbes produce chemical compounds that serve a variety of ecological functions (Ghoul and Mitri, 2016). For instance, siderophores are produced to scavenge iron and can drive cooperation, exploitation through "social cheating" and competition among bacteria (Kramer *et al.*, 2020). Alternatively, microbes can eliminate competitors through the production of antimicrobial compounds, such as bacteriocins (Riley and Gordon, 1999). While recent work has highlighted the natural product potential of marine sediments (Elfeki *et al.*, 2018; Bech *et al.*, 2020) and a variety of secondary metabolites have been isolated from cultures of marine sediment microbes (Fenical and Jensen, 2006; Dalisay *et al.*, 2013; Ziemert *et al.*, 2014; Yim *et al.*, 2017), the ecological role of these compounds in most cases remains unknown. Interestingly, when sediment microbial communities from Fiji were exposed to organic extracts of the actinomycete genus *Salinispora* in a laboratory mesocosm study, various members of the community, including predatory bacteria, were reduced while several other taxa became enriched, providing evidence that secondary metabolites can influence community diversity and structure within sediments (Patin *et al.*, 2017).

Traditionally, studies evaluating marine secondary metabolites have taken a targeted approach, monitoring for specific known compounds, such as a marine toxins (Lane *et al.*, 2010).

However recent years have seen an increase in untargeted environmental metabolomic studies. For instance, one study that took an untargeted approach to sediment metabolomics determined that the construction of a causeway led to divergence in the microbial communities and associated metabolites on either side of the causeway (Soliman *et al.*, 2017). This study however, was limited in its spatial scale and did not directly link any metabolites with specific microbes (Soliman *et al.*, 2017). Recent work out of the Jensen laboratory was able to identify specific molecules in marine sediments, determine concentrations, and correlate metabolites with known producers in the community (Tuttle *et al.*, 2019). Given their potential role in influencing microbial community structure and dynamics, it would be beneficial to gain insight into the spatial variability of secondary metabolites and test for correlations with the associated microbial communities.

Therefore, the goal of this study was to assess how microbial communities and metabolomes of marine sediments vary across spatial scales and in relation to environmental characteristics. To do this, five sites around Nacula Island, Fiji were sampled for their microbial communities using 16S rRNA, the corresponding sediment metabolome using an adsorbent resin technique and for a variety of sedimentary traits. A spatial sampling scheme was employed to facilitate site comparisons, as well as assess variation across a smaller (1 m^2 quadrat) and coarser (10 m transects) scale within one site. Correlative techniques were then used to determine linkages between microbial communities, metabolites and sedimentary traits. The results show that microbial communities are remarkably distinct from each other in both alpha and beta diversity. Multiple microbial taxa were determined to significantly vary in relative abundance based on site. Sediment metabolomes broadly overlapped and exhibit considerable variation in their richness. The majority of metabolites remain unknown, highlighting the chemical diversity in marine sediments, and some putative hits are discussed. Environmental characteristics including nutrients

and grain size composition of sediments were found to vary with sites and significant correlations were detected between microbes, metabolites and traits. Ultimately, the results provide important insight into the heterogeneity of marine sediment communities.

3.3 Methods

3.3.1 Sample Collection

All samples were collected around Nacula Island, Fiji via SCUBA in June of 2017 (Figure 3.1; Table 3.1). For fine scale sampling a 1 m^2 quadrat, divided into 16 even segments (4 x 4), was deployed at each field site adjacent to a reef area. Surface sediment from each of the 16 sections was then collected into whirl-pak (Nasco) bags denoting their location within the quadrat. To assess diversity differences along a coarser scale, surface sediments were collected at one-meter intervals along 10 m transects at sites 4 and 5. Upon returning to the surface, all sediment were stored in a cooler until processing.

To process the sediment samples, 5 - 6 ml from each bag was transferred to a pre-labeled 15 ml falcon tube containing 6 ml of RNALater® using sterile techniques. Samples were inverted to mix, parafilmed and stored at -20°C. A small scoop of sediment from each sample was also added to a cryovial containing 50% glycerol and stored at -20 °C for future culturing work if needed. The remaining sediment was kept in whirl-pak bags at -20 °C.

Figure 3.1: Map of sites sampled around Nacula Island, Fiji in June 2017.

Table 3.1: Fiji Sampling Site Information.

Site Number	Site Name	Depth (m)	Latitude (S)	Longitude (E)	Description
1	West of Nacula Island	11.58	16° 54.759’	177° 22.722’	sandy patch by reef
2	“The Zoo”	13.11	16° 56.866’	177° 23.934’	sandy patch by reef, noticeably more fish than site 1
3	Near Nacula village	10.36	16° 53.1776’	177° 24.218’	sand near reef, noticeable burrow holes in sediment
4	Off "Honeymoon Island"	10.36	16° 53.578’	177° 23.076’	coarse calcareous sediment, *Halimede* rubble, next to small reef
5	Off Blue Lagoon Resort	3.05	16° 54.796’	177° 23.258’	very fine sand near shallow reef and a seagrass bed

3.3.2 DNA Extractions & Sequencing

In order to extract DNA, 2 g of sediment was sterilely removed from the RNALater® sediment samples and divided into two ceramic bead beating tubes and extracted using a physical and chemical lysis method (Patin *et al.*, 2013). Replicates were then combined during the purification process. For the purposes of site comparisons, DNA was extracted from the four corner quadrat samples at each site (total n= 20) and sent for subsequent library preparation and sequencing through the UC Davis Host Microbe Systems Biology Core (HMSBC) in May 2018 (hereafter referred to as "corner samples"). The 16S rRNA gene was amplified using the V4 primers recommended by the Earth Microbiome Project (Thompson *et al.*, 2017) 515F: GTGYCAGCMGCCGCGGTAA (Parada *et al.*, 2016) and 806Rb GGACTACNVGGGTWTCTAAT (Apprill *et al.*, 2015) and sequenced via an Illumina MiSeq PE250 at a depth of ~100k reads per sample.

DNA was also extracted from all samples associated with site 4 (15 quadrat samples and 29 transect samples) in February 2020 (hereafter referred to as "site 4 samples"). Due to a lack of DNA and sediment sample, one corner sample (FJ17_49) was omitted from the site 4 sequencing efforts. Additionally, for transect samples, sediment was only preserved frozen in whirl-pak bags, so extractions were done on those samples rather than sediment from frozen RNALater® tubes. One transect sample (FJ17_T303) was missing and therefore is not represented in the dataset. Site 4 samples were then sent in March 2020 to Novogene (South Plainfield, NJ) for library preparation and sequencing using the above described 16S V4 primers on an Illumina Novoseq at a depth of ~100k reads per sample.

3.3.3 Community Analysis

Given that the corner samples and site 4 samples were sequenced on different platforms, the initial denoising process was performed separately using the same workflow. Sequences were imported into QIIME2 (version 2020.2) (Bolyen *et al.*, 2019) and denoised with DADA2 (Callahan *et al.*, 2016). Default DADA2 parameters were used with p-trunc-len-f 226, p-trunc-len-r 224, p-trim-left-f 20 and p-trim-left-r 21 for corner samples and p-trunc-len-f 250, p-trunc-len-r 229, p-trim-left-f 20 and p-trim-left-r 21 for site 4 samples. The resulting tables and rep-seq files were merged using the feature-table merge and feature-table merge-seqs commands in QIIME2. The merged file was then used to generate a tree for phylogenetic analyses and to perform taxonomic assignment with the SILVA v138 SSU NR99 515f-806r classifier (Quast *et al.*, 2013) based on referenced reads and taxonomy generated through RESCRIPt (Robeson II *et al.*, 2020). All taxonomic classifications reported follow SILVA v138 nomenclature. SILVA v138 included extensive taxonomic updates based on the Genome Taxonomy Database (GTD) (Parks *et al.*, 2018, 2020), LPSN (Parte, 2018; Parte *et al.*, 2020) and Bergey's. The dataset was then filtered to remove sequences identified as chloroplasts, mitochondria or eukaryotes. Subsequently, the dataset was separated into corner samples and site 4 samples and singleton sequences were removed. For comparative diversity analyses, corner samples were rarefied to 41,836 reads. For site 4 samples, four quadrat samples (FJ17_52, 56, 58 and 61) were found to have inadequate sequencing depth and removed (leaving n=11 for the quadrat and n=29 for the transect samples). The remaining site 4 samples were then rarefied to 57,299 reads for diversity analyses.

Alpha and beta diversity analyses were performed using QIIME2 (v2020.2). Faith's Phylogenetic Index (Faith, 1992) was used to determine richness estimates while evenness was assessed with Pielou's Evenness Index (Pielou, 1966). Beta diversity analyses were performed

using UniFrac, a phylogenetically informed method that can be done both unweighted (based on presence/absence of taxa) or weighted (which accounts for proportion of taxa) (Lozupone and Knight, 2005) and visualized as PCoA plots using Emperor (Vázquez-Baeza *et al.*, 2013). Richness and evenness values were imported into RStudio with R version 4.0.3 (R Core Team, 2019) and visualized as boxplots using ggplot2 (Wickham, 2016). Statistical analyses of diversity indexes were performed with QIIME2 and used the non-parametric Kruskal-Wallis test (Kruskal and Wallis, 1952) to compare richness and evenness across sites and between the site 4 quadrat and transect samples. Multivariate PERMANOVAs were performed to compare beta diversity results across sites and between the quadrat and transect samples of site 4 (Anderson, 2001). Analysis of composition of microbes (ANCOM) (Mandal *et al.*, 2015) was performed at the genus level to determine which genera could explain differences across sites using QIIME2. Finally, to compare how communities differed across spatial scales, Shannon diversity (Shannon and Weaver, 1964) within site 4 was compared to diversity across transects and other sites. A Kruskal-Wallis test was used to determine significance with Dunn's post-hoc test.

3.3.4 Metabolomics

In an effort to capture a snapshot of the sediment metabolome without collecting large amounts of sediments, I opted to use a resin capture method based on previous work in environmental metabolomics (Lane *et al.*, 2010; Tuttle *et al.*, 2019). Resin bags were made using miracloth sewn into 18 x 22 cm dimensions with 200 g of Diaion® HP-20 resin added before bags were hand-sewn shut. The HP-20 resin was then activated by submerging each bag in 100% HPLC-grade methanol (MeOH) for 5 minutes, followed by 50:50 MeOH :H_2O for 5 minutes and 100%

H_2O for 5 minutes. Bags were then placed into another 100% H_2O bath for a minimum of 10 minutes before being transferred and stored damp in a Ziploc® bag at room temperature.

Resin bags were transported in sealed Ziploc® bags via SCUBA in an effort to limit their interaction with seawater metabolites. After collecting sediment samples, five resin bags were buried underneath the surface sediment where the quadrat had previously been placed. The location was marked using a buoy and resin bags were left for approximately 24 hrs. Resin bags were then removed and placed into a sealed Ziploc bag®, brought to the surface and placed in a cooler for transport back to the field station. Resin bags were then stored in a -20 °C freezer until extractions could occur at SIO. A control resin bag was prepared at SIO, stored in a Ziploc® bag and processed in the same manner as experimental resin bags.

In order to extract the resin bags, they were first left to thaw for 48 hours at 4 °C. Each resin bag was rinsed with 1 L of distilled water five times to thoroughly remove any excess salts. Bags were then soaked in 200 ml of fresh HPLC-grade MeOH five times to remove metabolites from the resin. The solvent was filtered through P8 fluted paper (Fisher Scientific) and then dried via rotary evaporation. Mass of the crude extract was determined before crude extracts were resuspended in 100% MeOH to a concentration of 20 µg/ml. From there, a 1 µg/ml vial was made and filtered via 0.2 um spin columns and run on a Bruker ion-trap LC-MS (positive mode with H_2O:acetonitrile gradient; Phenomenex Luna C18 column, 5 um particle size, 100A pore size, 150 mm length).

3.3.5 Metabolomic Analysis

The resulting LC-MS/MS files were converted to .mzXML format and uploaded into MZMine 2 v3.27 (Pluskal *et al.*, 2010) for processing. Mass detection for both MS1 and MS2 used

the centroid method and noise levels were set to 1.0E5 and 1.0E3 respectively. The minimum time was set to 0.01, minimum height of 3.0E3 and *m/z* tolerance to 20.0 ppm for building the chromatograms. For chromatogram deconvolution the baseline cutoff was set to have a minimum peak height of 2.0E5, a duration of 0.01 to 3.0 min and a baseline level of 1.0E5. The *m/z* center calculation was median with a *m/z* range of 0.02 and a 0.1 min retention time for MS2 pairing. The isotopic peak grouper was set to have a *m/z* tolerance of 20 ppm, retention time tolerance of 0.1 min and a maximum charge of three with the most intense peak set as the representative isotope. Finally, join aligner was used with a 20 ppm *m/z* tolerance, 75% weight for *m/z*, 25% weight for retention time and 0.1 min retention time tolerance to align all detected peaks together and the peak list rows filter was applied to set minimum peaks in a row to 2 and minimum peaks in isotopic pattern to 2 peaks. Gap filling was performed with the peak finder method set to an intensity of 10%, 20 ppm *m/z* tolerance and 0.2 min retention time. The resulting peak list was exported as a .csv file for downstream analysis.

MS1 data were then filtered to remove any peaks detected in the first column blank sample and the method control sample in an effort to ensure the proper removal of any contaminants associated with the LC-MS/MS column and/or the extraction method used. This process reduced MS1 features from 22,713 down to 4,593. To compare metabolite richness across samples by MS1 features, the number of observed MS1 peaks for each sample was determined. Subsequently a one-way ANOVA test was performed to determine if metabolite richness was significantly different across sites. Residuals were checked to ensure adherence to ANOVA assumptions and a Tukey's HSD post-hoc subsequently performed. Metabolomic profiles (MS1) were also visualized in multivariate space with principal component analysis (PCA) in R. Site differences were determined with the Adonis command in R from the vegan package performed using the Canberra

method (Anderson, 2001; Oksanen *et al.*, 2020). Finally, MS1 peak intensity data was uploaded into MetaboAnalyst v5.0 (Chong *et al.*, 2019) and then normalized using cube-root transformation and pareto-scaling. A one-way ANOVA was performed to detect any metabolites that differed significantly based on site with a Fisher LSD post-hoc test.

Raw .mzXML files for all samples in addition to the method control and first solvent blank were uploaded into the Global Natural Product Social Molecular Networking (GNPS) platform (Wang *et al.*, 2016) for MS2 analysis using classical molecular networking. Default parameters were used with a precursor mass tolerance of 2.0 Da and a fragment ion mass tolerance of 0.5 Da. The resulting molecular network was visualized in Cytoscape v3.7.2 (Shannon *et al.*, 2003) and all nodes associated with the solvent blank and resin control were removed. Nodes were colored based on site detection with parent mass indicated on the nodes and cosine score indicated on the edges. A PCoA of MS2 data was also generated through the GNPS platform to assess differences in site metabolomic profiles. The solvent blank was removed, and the dataset rerun to better visualize the PCoA based on sites.

3.3.6 Sediment Characteristics

In order to quantify sediment composition and quality, a variety of trait assays and measurements were employed. For site comparisons, triplicate samples were taken from each corner section of the quadrats (5 sites x 4 corners x 3 replicates = 60 samples). For the site 4 samples, the organic content procedure was performed in triplicate while the sediment composition was done in duplicate due to time and material constraints. All assays were performed with sediment that was dried at 60 °C for a minimum of 12 hours to get an accurate starting dry mass.

Sediment organic content was determined using the loss on ignition (LOI) approach (Dale, 1974; Kreeger *et al.*, 2010). In summary, approximately 1 g of dry sediment was placed into an aluminum boat and then combusted in a muffle furnace at 550°C for 4 hours. The remaining inorganic ash was then massed, and ash-free dry mass calculated as a proxy for organic content. Percent calcium carbonate in sediments was determined by employing an acid titration method (Black, 1965; Dean Jr, 1974). In summary, approximately 1 g of dry sediment was added to a 20 ml glass scintillation vial. Hydrochloric acid (0.5 M) was then added in small amounts and the vial swirled causing a reaction that produces CO_2 bubbles. This process was repeated until the reaction stopped. Sediment samples were then rinsed three times with distilled water to remove the acid and dried overnight at 60 °C. Samples were then massed and the mass difference was used as a proxy to calculate the percent calcium carbonate lost. One sample from site 2 (FJ17_20) was missed from calcium carbonate analysis.

Sediment grain size composition was determined by employing a Fieldmaster ® sieve set containing six standardized mesh sizes: 4000 µm (fine gravel), 2000 µm (very fine gravel), 500 µm (coarse sand), 250 µm (medium sand), 125 µm (fine sand) and 63 µm (very fine sand). Due to instability when shaking the entire sieve stack together, the sieves were divided and shaken in two sets. First, 5 g of dry sediment was added to the sieve set containing the four largest sieve sizes and shaken for 22 min at ~280 rpm on a rotary shaker (New Brunswick). Next, the remaining finer sediment from the base collection pan was moved into the 125 µm sieve and shaken with the 63 µm sieve for an additional 22 min. To determine the percentage of sediment in each size fraction, the sediment was removed from each sieve and massed. The percentage was then determined based on the proportion of mass in each sieve over the total collected mass. The total collected mass was then also used in relation to the original starting mass to determine recovery efficiency.

Nutrient analysis was also completed on the corner samples from each site. Dry sediment was submitted to the Soil Analytical Laboratory at UC Davis and ammonium (NH4-N), nitrate (NO3-N), and iron (Fe total) concentrations were determined following standardized protocols (Keeney and Nelson, 1982; Sah and Miller, 1992; Hofer, 2003; Knepel, 2003). In order to supply the proper amount of sediment for nutrient analysis, some of the sediment that had been used and regathered during the sediment composition protocol was sent in addition to unused dry sediment.

3.3.7 Sediment Analysis

Sediment characteristics were visualized with ggplot2 v3.3.2 in RStudio v4.0. The jitter_point function was used to display variation in individual replicates. For statistical analyses, the average value for each corner sample was used (n=4 per site) rather than each replicate to avoid issues with pseudo-replication. To statistically compare differences in sediment characteristics across sites, one-way ANOVAs were performed with R and Tukeys HSD post-hoc test subsequently done on characteristics that were found to be significantly different by site. ANOVA residuals were checked to ensure no violations in statistical assumptions. If assumptions were violated, the non-parametric Kruskal-Wallis test was then performed followed by Dunn's post-hoc test (Dunn, 1964) using the FSA package in R (Dinno, 2017; Ogle *et al.*, 2020). Given that sediment size composition is a compositional dataset, it was plotted using the multivariate PCA approach with the vegan package v2.5-7 (Oksanen *et al.*, 2020) in RStudio. PERMANOVA and pairwise PERMANOVA was performed in RStudio using the adonis command from vegan and the pairwise.adonis2 command from the pairwiseAdonis library (Martinez Arbizu, 2017) via devtools v2.3.2 (Wickham *et al.*, 2020) to test site differences in sediment composition.

3.3.8 Correlation Analyses

In an effort to evaluate connections between the community, metabolomics and sediment composition, multivariate correlation analyses were employed. For correlation analyses, community data was filtered in QIIME2 to maintain only features with a minimum frequency of 20 (selected since n=20), reducing the original dataset from ~25,000 features down to 9,995 features. Diversity analyses were then redone with rarefaction set to 40,369 reads. The metadata distance-matrix command was used to generate distance matrixes based on Euclidean distance for each sediment characteristic except grain size composition. Since grain size is compositional data in nature, it was imported into RStudio and converted into a Euclidean distance matrix with the dist function. The distance matrix was then imported into QIIME2. The metabolomics dataset was imported into RStudio and converted into a Canberra distance matrix. The Canberra distance matrix was then imported into QIIME2 for subsequent analysis.

To first test for correlations between datasets, Mantel tests with spearman rank correlation coefficients were performed in QIIME2. Weighted UniFrac distance was used for the microbial community input as it accounts for relative proportion of community members. A supervised approach (based on site) was then used to determine if correlations between microbial genera and metabolites could be extracted via the DIABLO method in the MixOmics package (Rohart *et al.*, 2017) in R. Since DIABLO requires a paired dataset, the metabolomics data was haphazardly reduced to four samples per site and matched with community samples from the same site. DIABLO also advises the use of normalized data, so the transformed and scaled metabolomics dataset from MetaboAnalyst was utilized. For the purposes of identifying connections, the filtered microbial community dataset was input at the genus level (794 genera total) and transformed using centered log-ratio (CLR) with the microbiome package (Lahti and Shetty, 2019). Metadata was

also scaled using the base scale function in R prior to DIABLO analyses. For DIABLO, a sparse partial least squared discriminant analysis (sPLS-DA) was run on the paired datasets with site as the identifying factor. The network function was then used with a correlation cutoff of 0.7 to detect connections. Due to the difficulty in visualizing all the datasets and the web of connections at once, networks were subsequently generated based on paired datasets. First, the microbe-metabolite network was visualized, followed by a network focused on genera and sediment characteristics. In an effort to further isolate and visualize important connections, the nutrient dataset was run separately from the sediment size composition dataset. To reduce the number of connections for subsequent analysis, the cutoff value in the nutrient network was raised to 0.8 while the sediment size composition network was increased to 0.9. To assess connections at site 4, all genera from the site 4 samples were used and the sediment size composition was tested in conjunction with organic content with the cutoff set to 0.8 given the smaller dataset dimensions.

3.4 Results

3.4.1 Microbial Diversity Across Sites

Sediments from five sites around Nacula Island, Fiji, covering about 12 km^2 were sampled for their microbial communities (Figure 3.1, Table 3.1). Sediment communities exhibited considerable richness, 24,759 amplicon sequence variants (ASVs) were observed in total, averaging between 2,200 to 3,290 depending on site (Figure 3.2a). Sites 3, 4 and 5 had significantly greater richness, both in terms of observed ASVs (Observed ASVs Kruskal-Wallis H=10.36 p=0.035) and in phylogenetically informed richness (Figure 3.2b; Faith's PD Kruskal-Wallis H=12.29 p=0.015). Sites were all fairly even, although site 5 was significantly less even than the other sites (Figure 3.2c; Pielou's Evenness Kruskal Wallis H=12.53 p=0.014). Sites were

significantly different in terms of beta diversity, both when considering presence/absence of taxa (Figure 3.3a; Unweighted UniFrac PERMANOVA pseudo-F= 3.00 p=0.001) and the relative proportions of ASVs (Figure 3.4b; Weighted UniFrac PERMANOVA pseudo-F= 20.43 p=0.001).

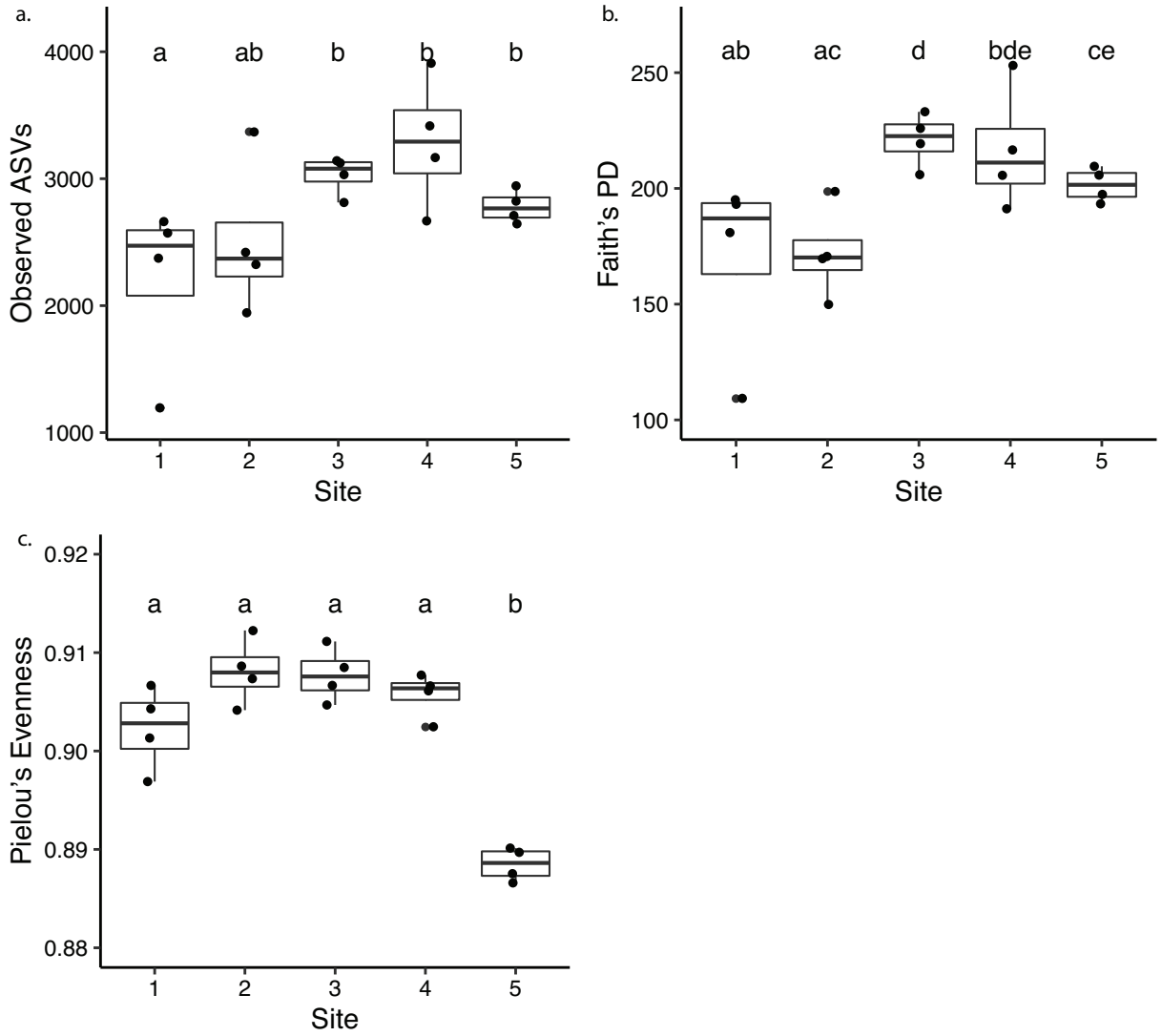

Figure 3.2: Alpha diversity of sediment microbial communities across five sites around Nacula Island, Fiji. a) Observed amplicon sequences variants (ASVs) b) Faith's phylogenetic diversity (PD) index and c) Pielou's evenness. Letters denote significant differences (alpha=0.05).

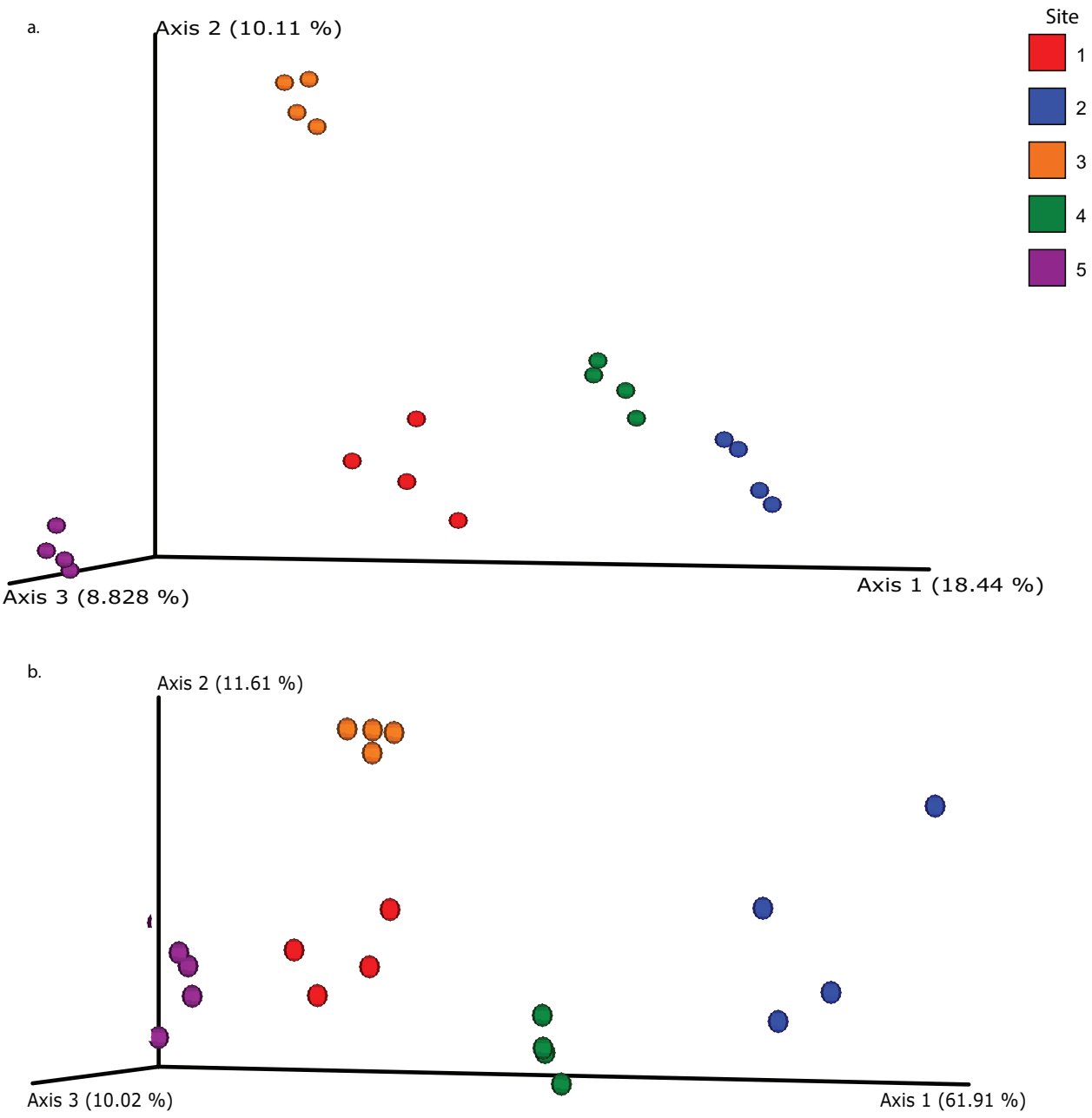

Figure 3.3: Beta diversity of sediment microbial communities across five sites around Nacula Island, Fiji. a) unweighted UniFrac and b) weighted UniFrac.

A total of 78 phyla (13 Archaea and 65 Bacteria) were detected across sites. Proteobacteria were the most abundant with an average relative percent ranging from 42 to 48% at each site. Desulfobacterota was the second most abundant phylum across sites 1, 3, 4 and 5 (average range 8.15-15.52%) with the exception of site 2 (average 0.97%). The second most abundant phylum at

site 2 was Crenarchaeota which was on average 8.62% of the community. Bacteroidota (site 1 and 5) and Planctomycetota (site 2, 3 and 4) were the third most abundant phyla detected. The majority of phyla were <1% of the relative community on average with only 17 phyla surpassing 1% at each site (Figure 3.4).

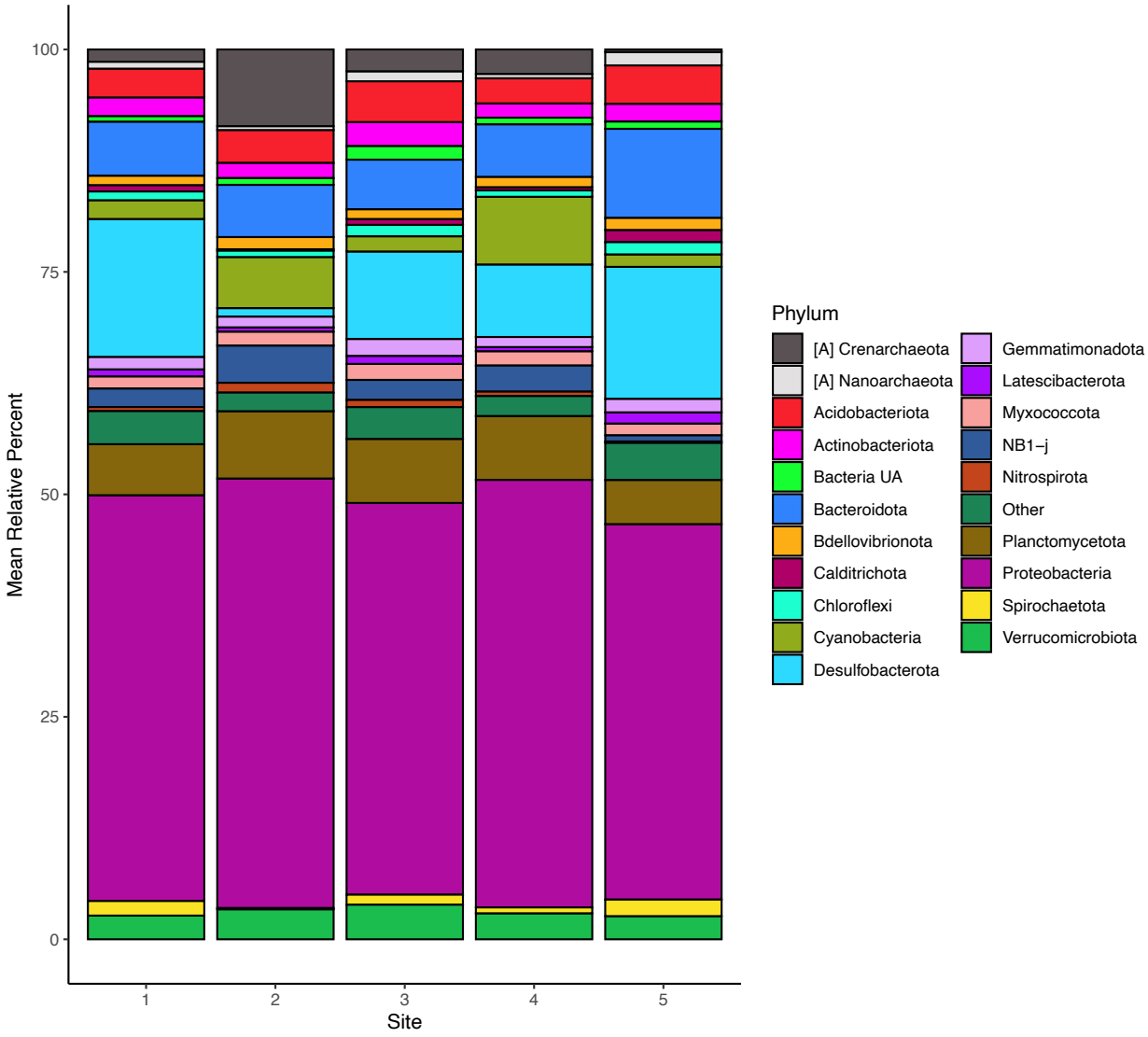

Figure 3.4: Mean relative percent of phyla detected across five sites around Nacula Island, Fiji. Taxonomy assigned with SILVA v138. Archaea are denoted as [A].

ASVs were assigned to 1,239 genera in total (1,190 Bacteria, 49 Archaea). Of the genera assigned, 98% accounted for <1% of the relative community on average across sites. The most abundant genus detected across all five sites was *Woeseia* (Protobacteria), which averaged 6.66 ± 0.44% of the community per site. Some of the other genera that represented >1% of the relative community included an unidentified Gammaproteobacterium (3.51 ± 0.55%), an uncultured Kiloniellaceae (3.32 ± 0.90%), an unidentified BD7-8 Gammaproteobacterium (2.92 ± 0.37%) a Sva0081 sediment group (Desulfobacterota) genus (2.73 ± 0.70%) and an unidentified genus associated with the NB1-j clade (2.42 ± 0.57%). ANCOM performed at the genus level identified 24 genera that significantly differed based on sites and included 11 Proteobacteria, 3 Nitrospirota, 3 Desulfobacterota, 1 Cyanobacteria, 1 Bacteroidota and 5 Crenarchaeota (Figure 3.5). While many of the significant taxa could not be identified to the genus level, the genera that were identified included *Thiohalobacter, Thiohalophilus, Congregibacter, Thermodesulfovibrio, Nitrospira, Desulfoconvexum, Muriicola, Cenarchaeum, Candidatus Nitrosopumilus and Candidatus Nitrosopelagicus* (Figure 3.5).

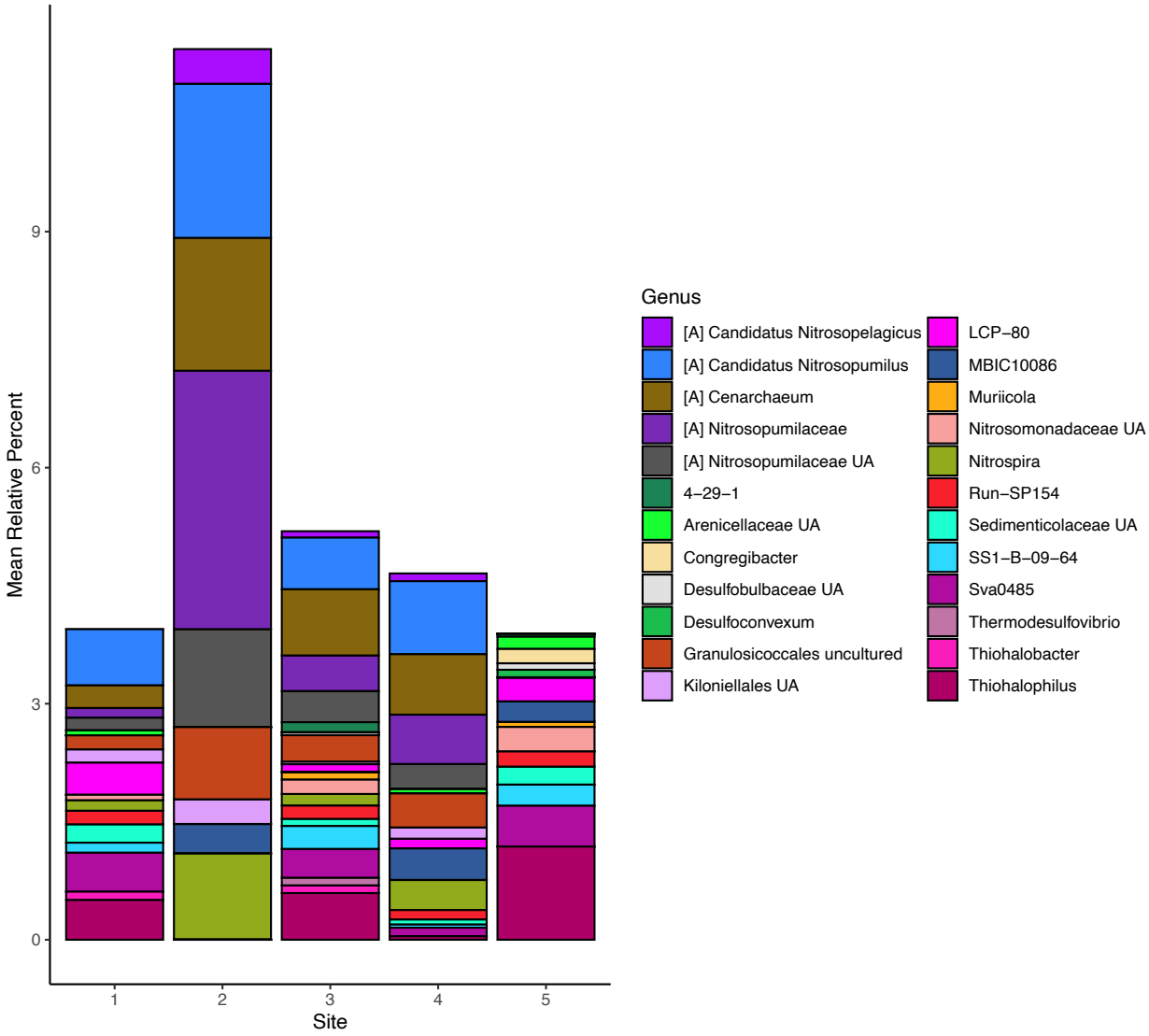

Figure 3.5: Mean relative percent of genera identified as significant with ANCOM analysis. Taxonomy assigned with SILVA v138. Archaea are denoted as [A].

3.4.2 Sediment Metabolomics

Extraction of resin bags in methanol resulted in approximately 23-30 mg of crude MeOH soluble extract per 200 g of resin. After processing LC-MS/MS spectra and filtering to remove any peaks detected in the control sample or the column wash, 4,593 MS1 peaks remained (20%). The number of metabolites detected averaged 2,054 ± 11 per site. Site 1 had significantly fewer compounds, with an average of 1972 ± 15 (Figure 3.6). Substantial overlap in MS1 feature profiles

was observed (Figure 3.7), although site was found to be significantly different (Adonis PERMANOVA F=1.03, p=0.004) with a pairwise adonis post-hoc tests indicating that site 1 was driving the significant result. Site 1 was different from sites 2, 3 and 4. Site 4 was also found to be significantly different than site 5 (Figure 3.7). Results comparing MS1 features across all sites performed with MetaboAnalyst v5.0 detected two features based on a one-way ANOVA on site (Figure 3.8).

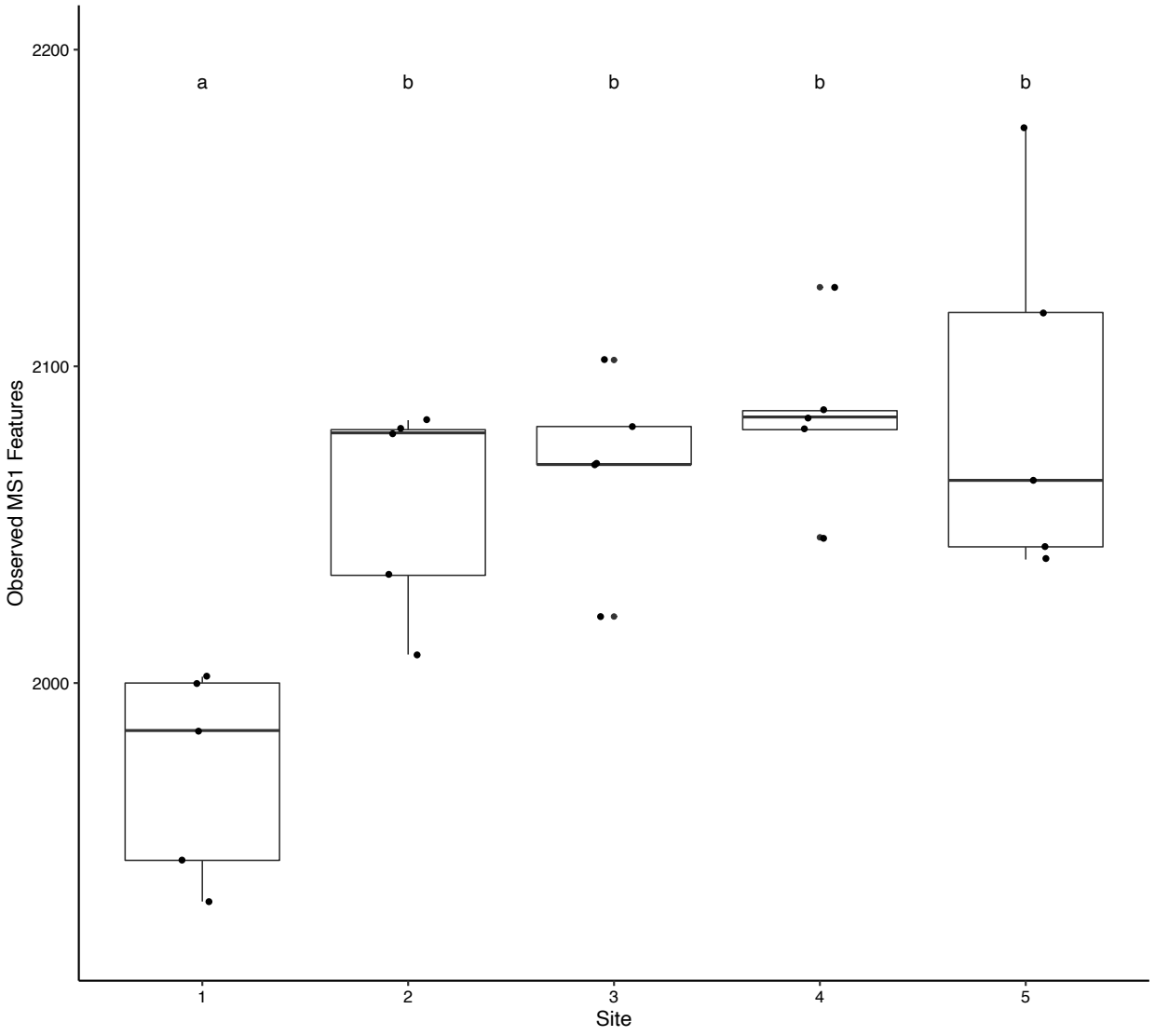

Figure 3.6: Richness of sediment metabolites (MS1 features) observed across five sites around Nacula Island, Fiji. Letters denote significant differences (alpha=0.05) determine through Tukey's HSD post-hoc test following a one-way ANOVA with site as factor.

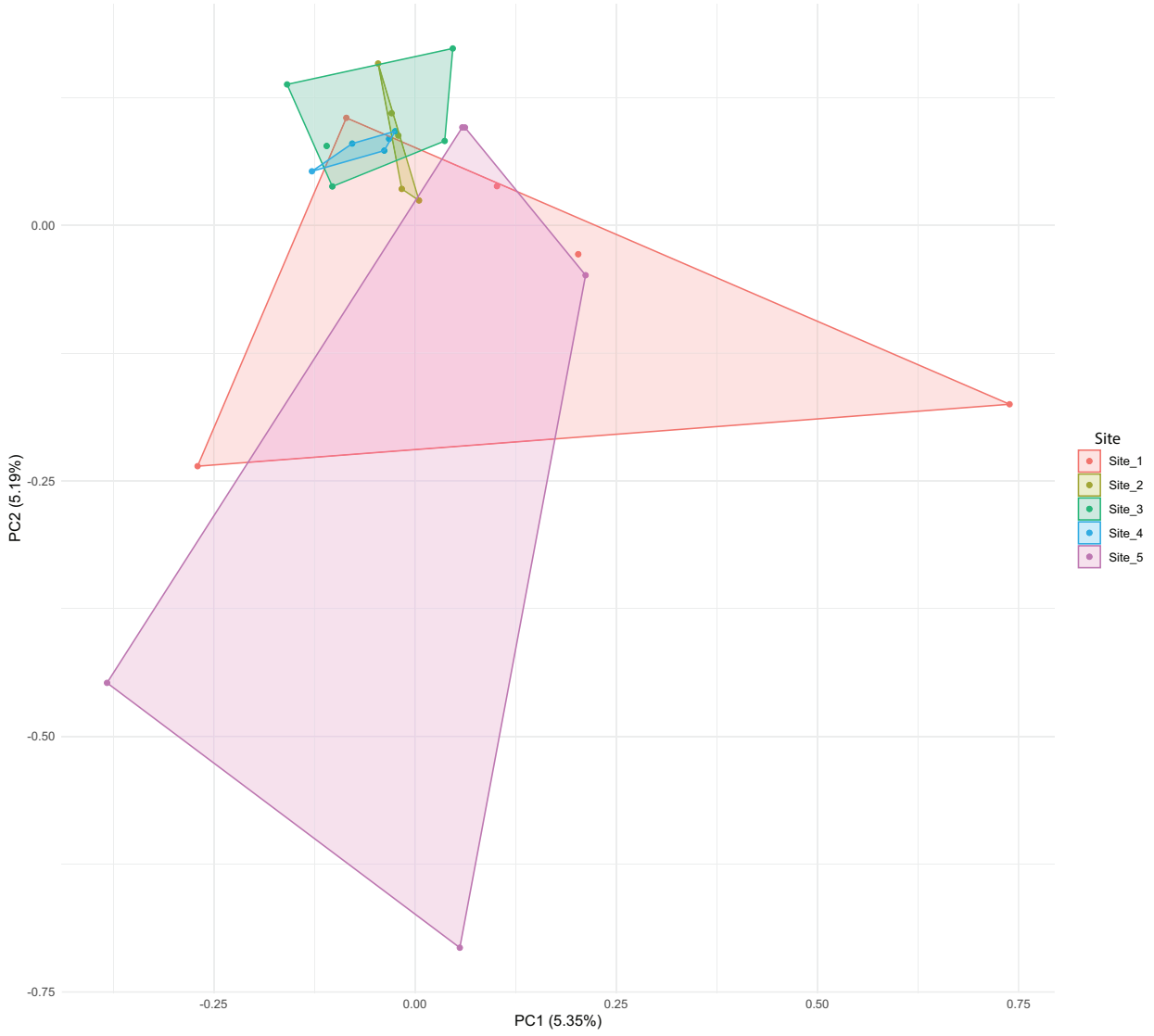

Figure 3.7: PCA of metabolite (MS1 features) composition across sediments from five sites around Nacula Island, Fiji.

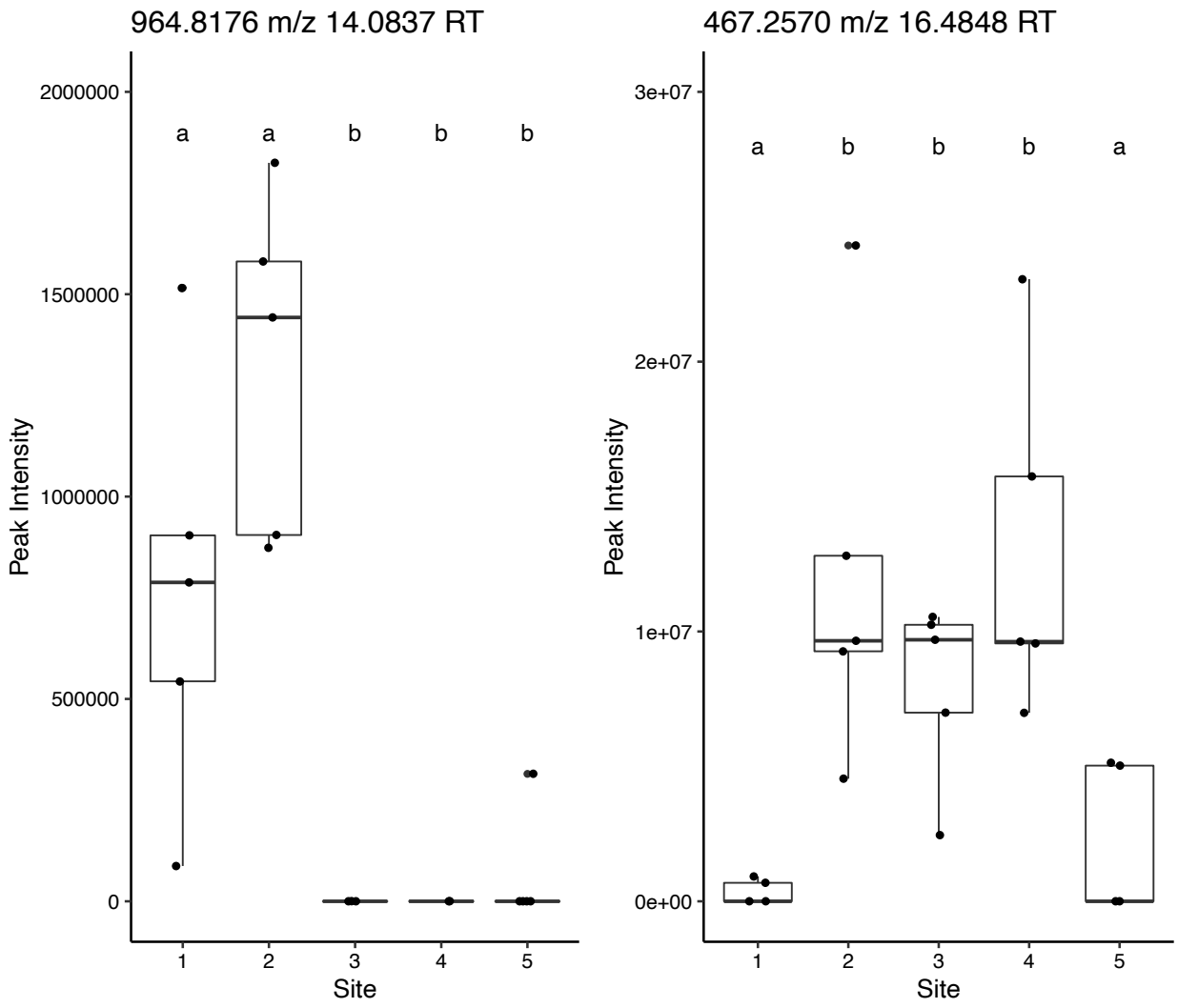

Figure 3.8: Peak Intensity of significant features identified with MetaboAnalyst by site. Letters denote significant differences (alpha=0.05) determine through Dunn's post-hoc test following a Kruskal-Wallis test with site as factor.

Classical molecular networking identified 942 different parent masses (denoted as "nodes") in the initial dataset. After filtering to remove nodes associated with method controls, 526 different parent masses were detected with 313 (59.5%) being detected only once, i.e., unrelated to any other nodes based on a 0.7 cosine score (Figure 3.9). Only 19 nodes (3.6%) could be putatively identified through the GNPS platform. These included natural products such as the cyanobacterial compound Bastimolide C and the lichen associated Methoxyhaemoventosine.

Anthropogenically-associated compounds such as cilostazol and dioctyl sulfosuccinate were also putatively identified (Appendix B). As with the MS1 data, MS2 diversity plotted in multivariate space showed overlap across sites (Figure 3.10).

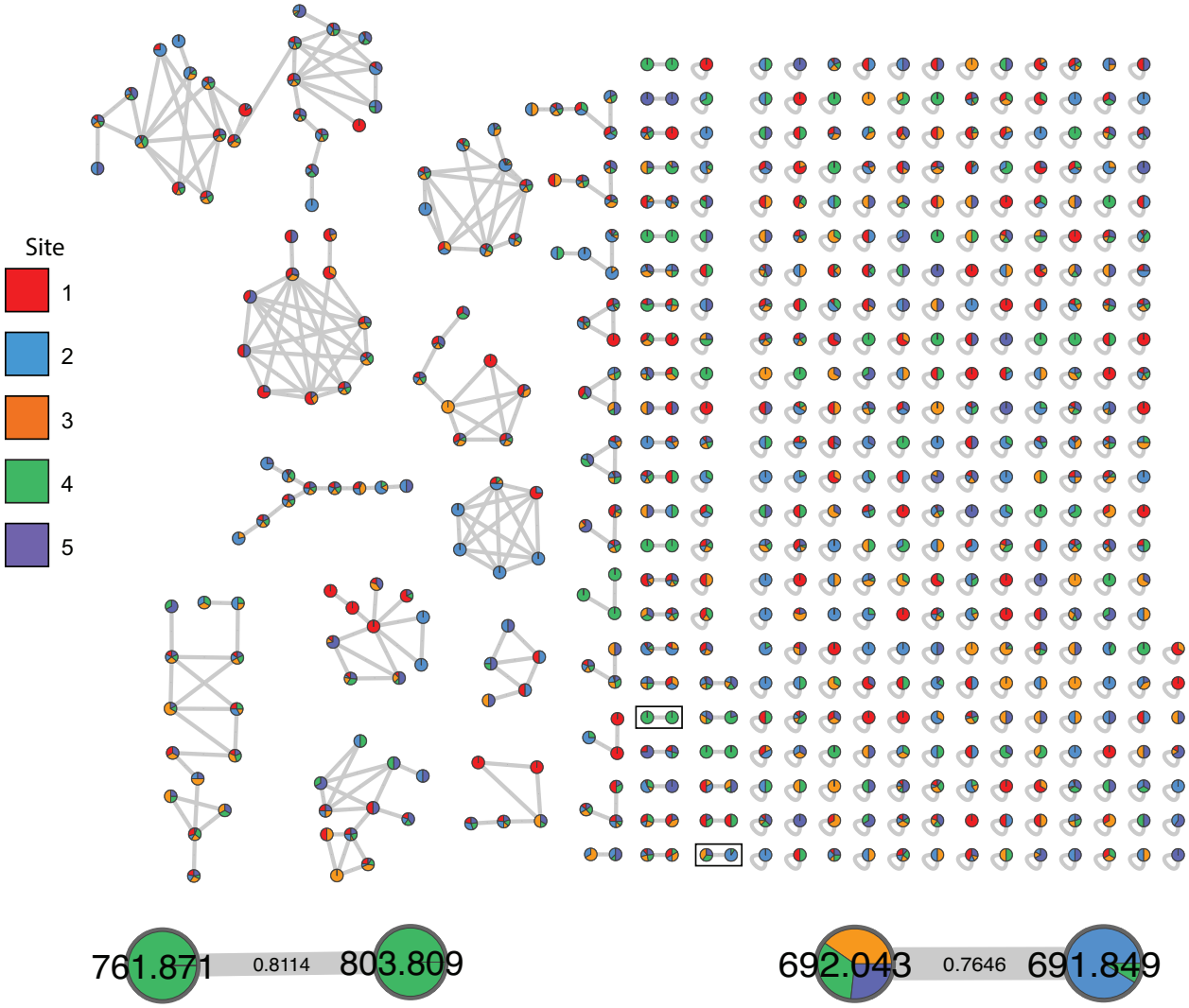

Figure 3.9: Molecular network generated with GNPS based on MS2 fragmentation patterns of crude resin extracts. Nodes represent parent masses while edges denote cosine values >0.7. Nodes are colored to depict the sites where they were detected. The two enlarged doublets include putative hits to Bastimolide C (761.87) and Methoxyhaemoventosine (691.849).

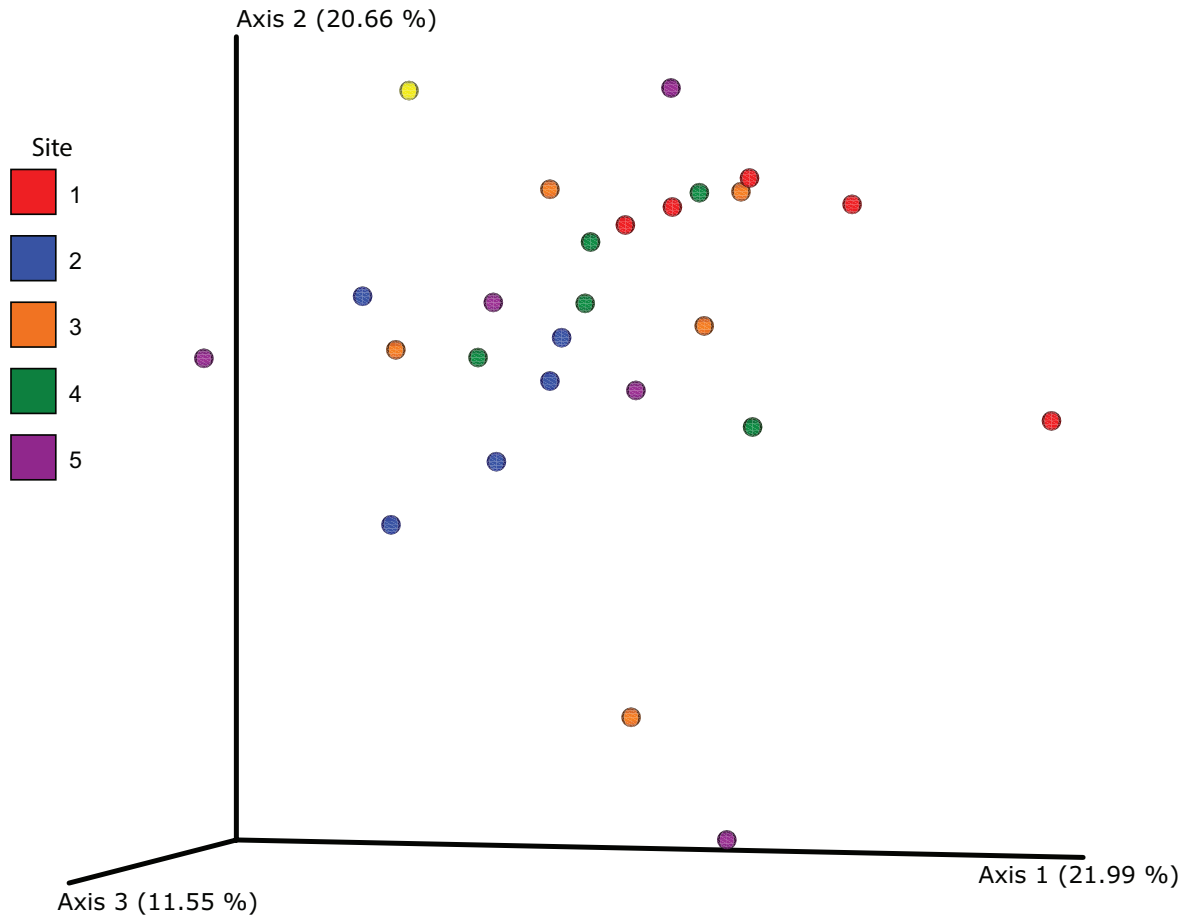

Figure 3.10: PCoA of sediment metabolomes from crude resin extracts generated with GNPS. Points are based on MS2 fragmentation patterns and colored to denote site of origin.

3.4.3 Sediment Characteristics

Sediment organic content averaged 4.2% across samples (Figure 3.11a) and was not significantly different by site (ANOVA $F_{4,15}$=1.655; p=0.213). Percent calcium carbonate was also fairly consistent across sites (Figure 3.11b), averaging 27.2% (ANOVA $F_{4,14}$=0.108; p=0.978). Nutrients, however, significantly differed depending on site. Ammonium concentrations ranged

from 0.29 to 1.19 ppm (Figure 3.11c) with the highest concentrations observed at site 3 and site 4 (ANOVA $F_{4,15}$=10.15; p<0.001). Site 2 in contrast, was found to have significantly higher nitrate levels compared to all other sites (Kruskal Wallis H=14.75; p=0.005), averaging 0.93 ppm while average nitrate concentrations at sites 1, 3, 4 and 5 ranged from 0.34 to 0.55 ppm (Figure 3.11d). Iron was also found in significantly different concentrations based on site (Kruskal Wallis H=17.39; p=0.002) with the highest concentrations observed at site 3 (average 2,577.33 ppm) and site 5 (average 1,375.96 ppm) compared to sites 1, 2 and 4 which averaged 332.04 to 495.92 ppm (Figure 3.11e). Sites were also significantly different from each other based upon the sediment size composition (Figure 3.12; PERMANOVA pseudo-F=4.95, p=0.016). Sites 1, 2 and 4 exhibited greater proportion of very fine gravel and coarse sand compared to sites 3 and 5 which was dominated by fine and very fine sand (Figure 3.12). Sites 3 and 4 also had greater relative proportions of medium sand compared to the other sites (Figure 3.12).

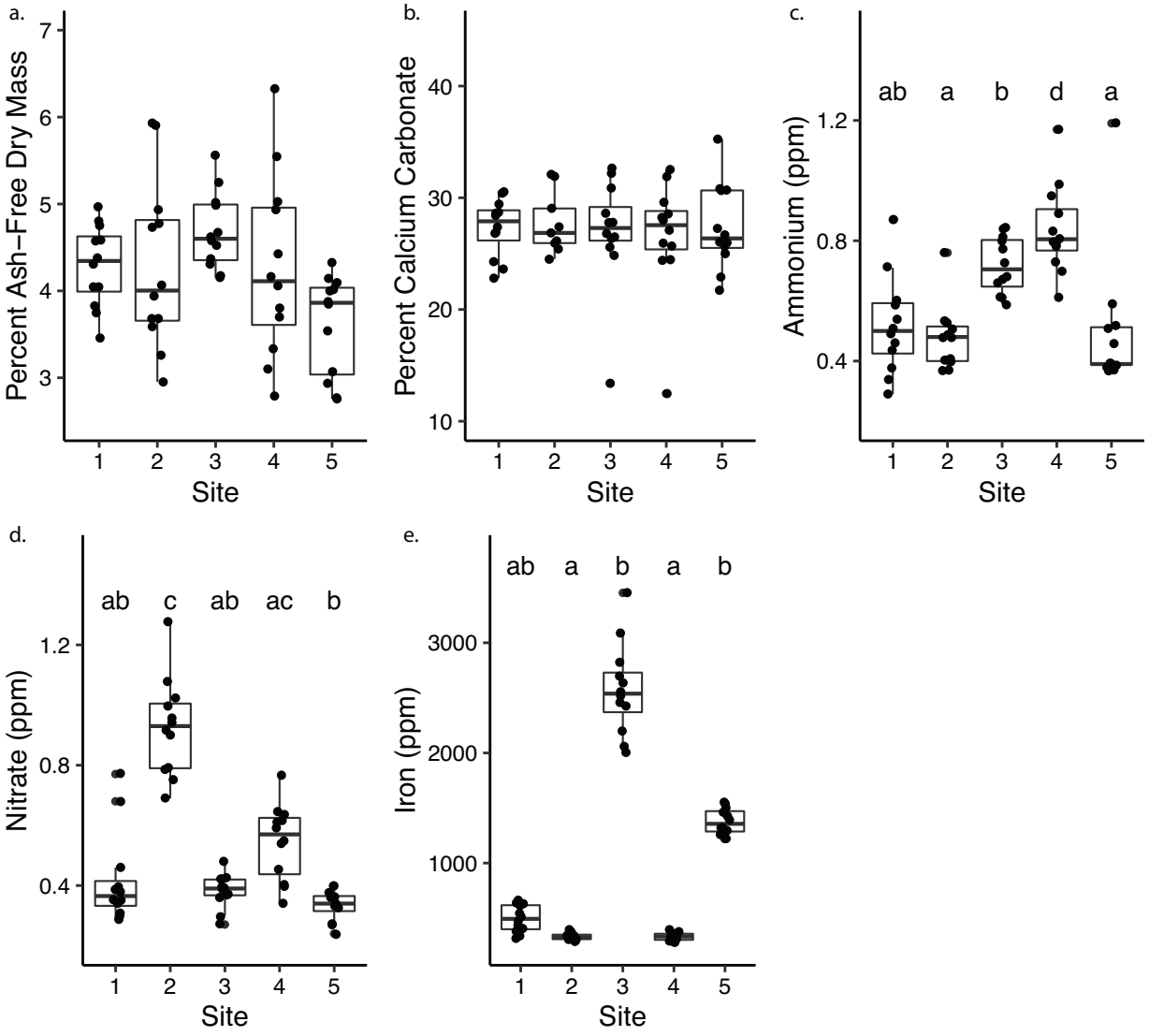

Figure 3.11: Boxplots of sediment traits measured by site. All replicates are plotted as individual points to visualize variance. Traits observed included a) Percent ash-free dry mass b) percent calcium carbonate c) Ammonium concentrations d) Nitrate concentrations and e) Iron concentrations. Letters denote significance (alpha=0.05).

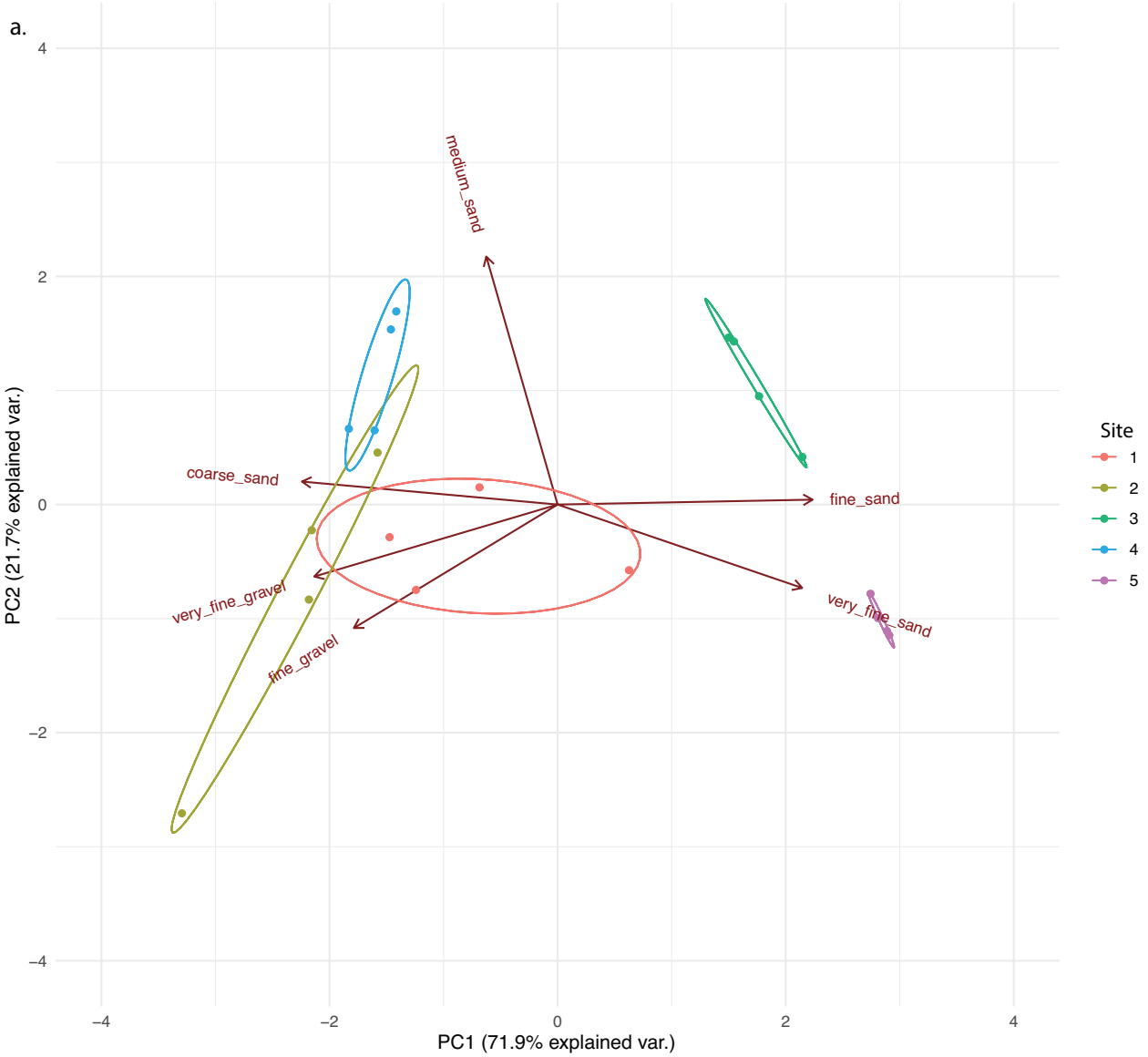

Figure 3.12: Sediment grain size composition across five sites around Nacula Island, Fiji. a) Biplot with sites denoted by color and b) mean relative percent of grain size classes by site.

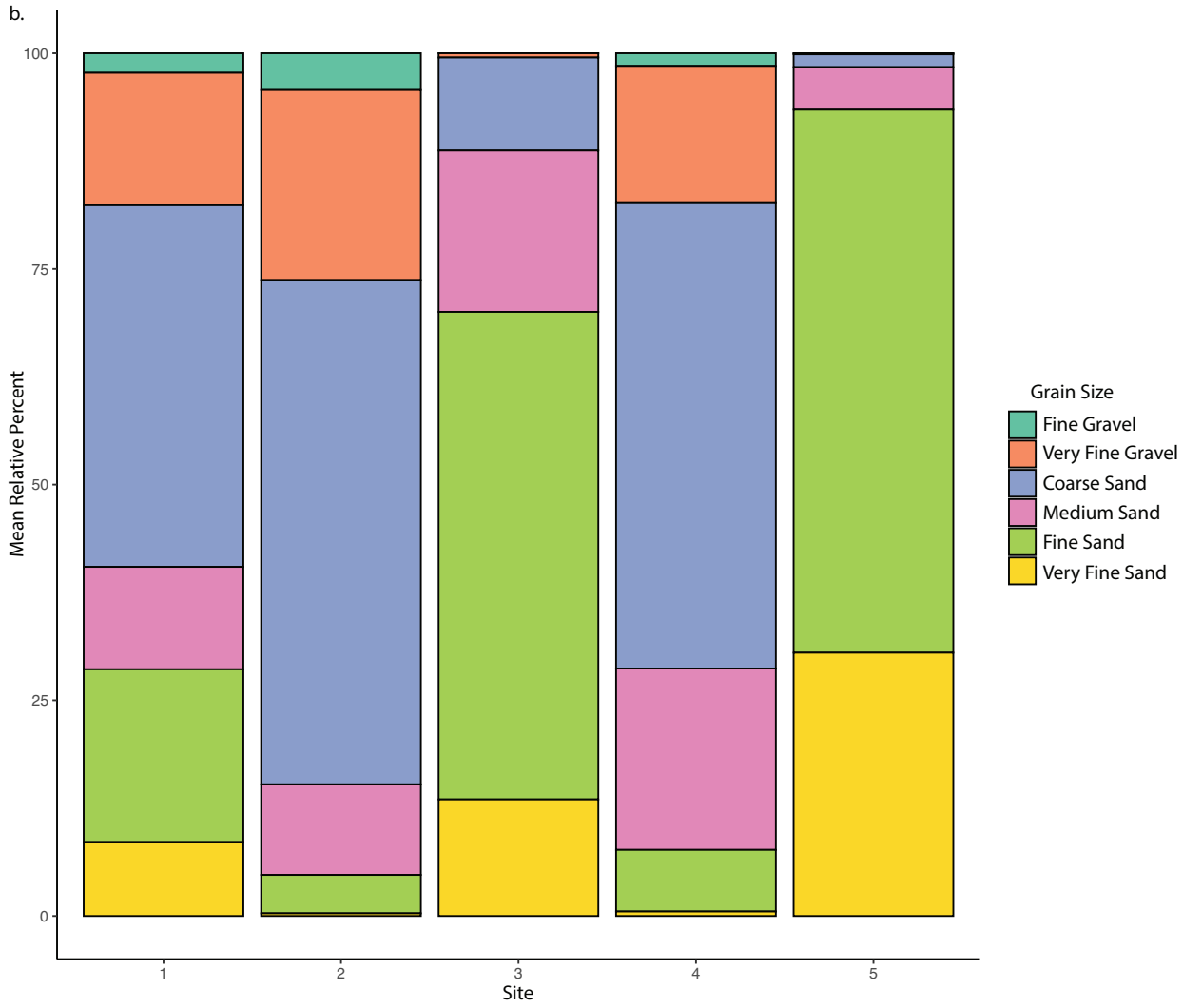

Figure 3.13: Sediment grain size composition across five sites around Nacula Island, Fiji, Continued.

3.4.4 Connecting microbes, metabolites and sediment characteristics

To first assess overall correlations between the microbial communities and the corresponding sediment metabolites and characteristics, multivariate mantel tests were performed. Filtered microbial communities (9,995 ASVs) accounting for proportion (weighted UniFrac) were tested against the filtered metabolomic dataset (filtered to remove any metabolites associated with control samples) and no correlation was detected ($p=0.406$, $rho=0.088$). When microbial

communities were tested for correlations with sediment characteristics, nitrate, iron and sediment size composition were found to be significantly correlated while organic content, calcium carbonate content and ammonium showed no correlation with microbial communities (Table 3.2).

Table 3.2: Correlations between sediment microbial communities and sediment characteristics. Results were generated from multivariate Mantel's tests with Spearman's rank correlation coefficient performed on the 16S rRNA community (weighted UniFrac) versus sediment characteristics.

Trait	Spearman's Rho	P-value
Organic content	-0.017	0.840
CaCO3	-0.012	0.904
Ammonium	0.075	0.406
Nitrate	0.475	0.001
Iron	0.311	0.009
Grain Size Composition	0.676	0.001

Next, to extract significant correlations between datasets, the supervised DIABLO method from mixOmics was employed. The sparse partial least squares discriminant analysis (sPLS-DA) of the paired genera-metabolite dataset was generated with each dataset (genera, metabolites, traits, sediment size composition) as a component and site as the known factor. DIABLO plots were generated to visualize how well the sPLS-DA model maximized correlations between datasets (Appendix C). Network analysis was then performed with the sPLS-DA on pairwise datasets for better visualization. First, microbe-metabolite interactions were analyzed, and five metabolites were found to have significant correlations with 13 genera

(cutoff = 0.7) (Figure 3.13a). While all the metabolites require future work for identification and isolation, the genera include members of Proteobacteria, Desulfobacterota, Planctomycetota, Myxococcota, Acidobacteriota, Spirochaetota and Latescibacterota (Figure 3.13a). Notably, one of the metabolites (964.8176 *m/z*) identified in network analysis was also identified previously

with MetaboAnalyst as a significant driver of site differences. This metabolite was negatively correlated with an unculture Gammaproteobacterium, an uncultured member of the Sandaracinaceae family and Nitrosomonadaceae IS-44. The Sandaracinaceae family of myxobacteria, described for their starch-degrading ability in soils (Mohr *et al.*, 2012), has also been identified as a "core" sediment microbe (Zeng *et al.*, 2017; Probandt *et al.*, 2018) and is of interest for natural product discovery (Garcia *et al.*, 2018) while Nitrosomonadaceae are ammonia-oxidizing bacteria associated with both nitrogen and iron cycling (Prosser *et al.*, 2014).

DIABLO was also used to identify connections between microbial genera and sediment characteristics (Figure 3.12b & c). Nitrate was correlated with 13 genera (10 negative correlations, 3 positive correlations), while iron was positively correlated with four genera in network analysis (Figure 3.13b). Genera linked with nitrate included multiple taxa associated with nitrogen cycling such as the proteobacterial family Nitrosomonadaceae, Nitrospira (phylum Nitrospirota) and the cyanobacterial genus *Xenococcus*. Three of the genera associated with iron were also associated with nitrate and included Latescibacteraceae, *Desulfurivibrio* and *Thiohalophilus* (Figure 3.13b). Latescibacteraceae (phylum Latescibacterota aka Latescibacteria, formerly WS3) is thought to play an important role in carbon turnover through its degradation of algal cell walls (Youssef *et al.*, 2015) and has also been linked with iron reduction in the rhizosphere of rice (Zhang *et al.*, 2019). Both *Desulfurivibrio* and *Thiohalophilus* are involved in sulfur cycling and *Thiohalophilus* is also able capable of growth via denitrification (Sorokin *et al.*, 2015; Melton *et al.*, 2016).

Since grain size composition was highly correlated with the microbial community, I increased the network analysis cutoff to 0.9 to reduce connections (Figure 3.13c). Twenty-five genera had significant interactions with size composition (Figure 3.13c). Very fine sand formed its own network cluster and had the greatest number of correlations, with 14 genera across six

phyla, including Crenarchaeota linked to this grain size (Figure 3.13c). All genera associated with very fine sand were negatively correlated, suggesting a loss of them with increasing proportion of very fine sand. Medium sand was also in its own cluster with two genera, both of which were positively correlated. This pattern is also reflected in community data which show *Rhodopirellula* and the Proteobacteria group Ga077536 in the highest relative abundances at sites 3 and 4 which were also the sites with the greatest proportion of medium sand. The third and final network cluster included very fine gravel, coarse sand and fine sand linked to nine genera. Of the nine genera in this cluster, three of them were significantly correlated to two grain sizes (Figure 3.13c). Interestingly, the four genera associated with fine sand were positively correlated while all taxa associated with coarse sand and very fine gravel were negatively correlated. Multiple genera identified in networks were linked to multiple traits, potentially indicative of covariation between sediment grain size and nutrients.

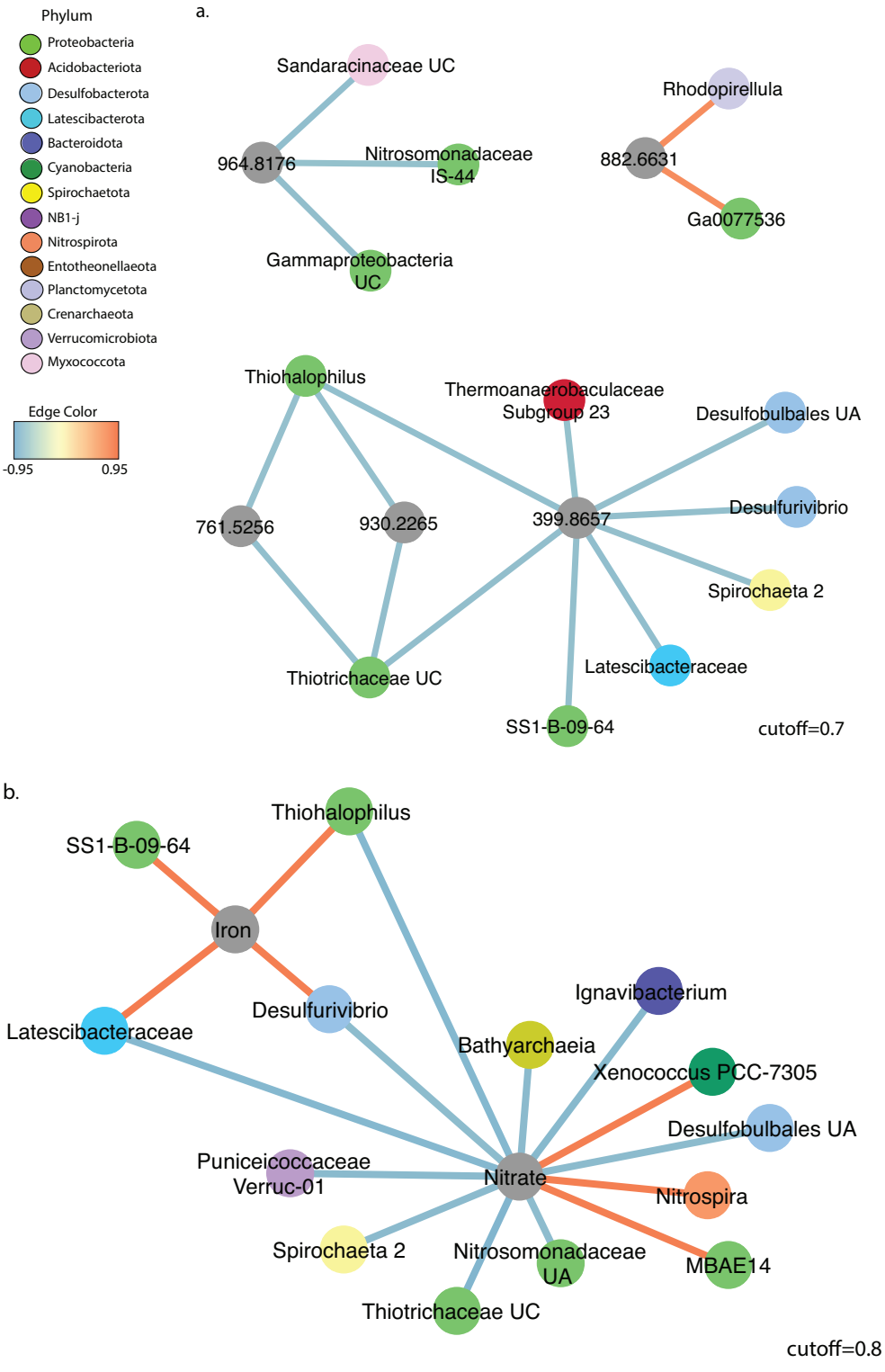

Figure 3.14: DIABLO networks representing significant correlations between microbial genera and a) metabolites denoted by their mass, b) sediment nutrients and c) grain size. "UC" = uncultured while "UA" = unannotated genus. Edges are colored based on correlation value.

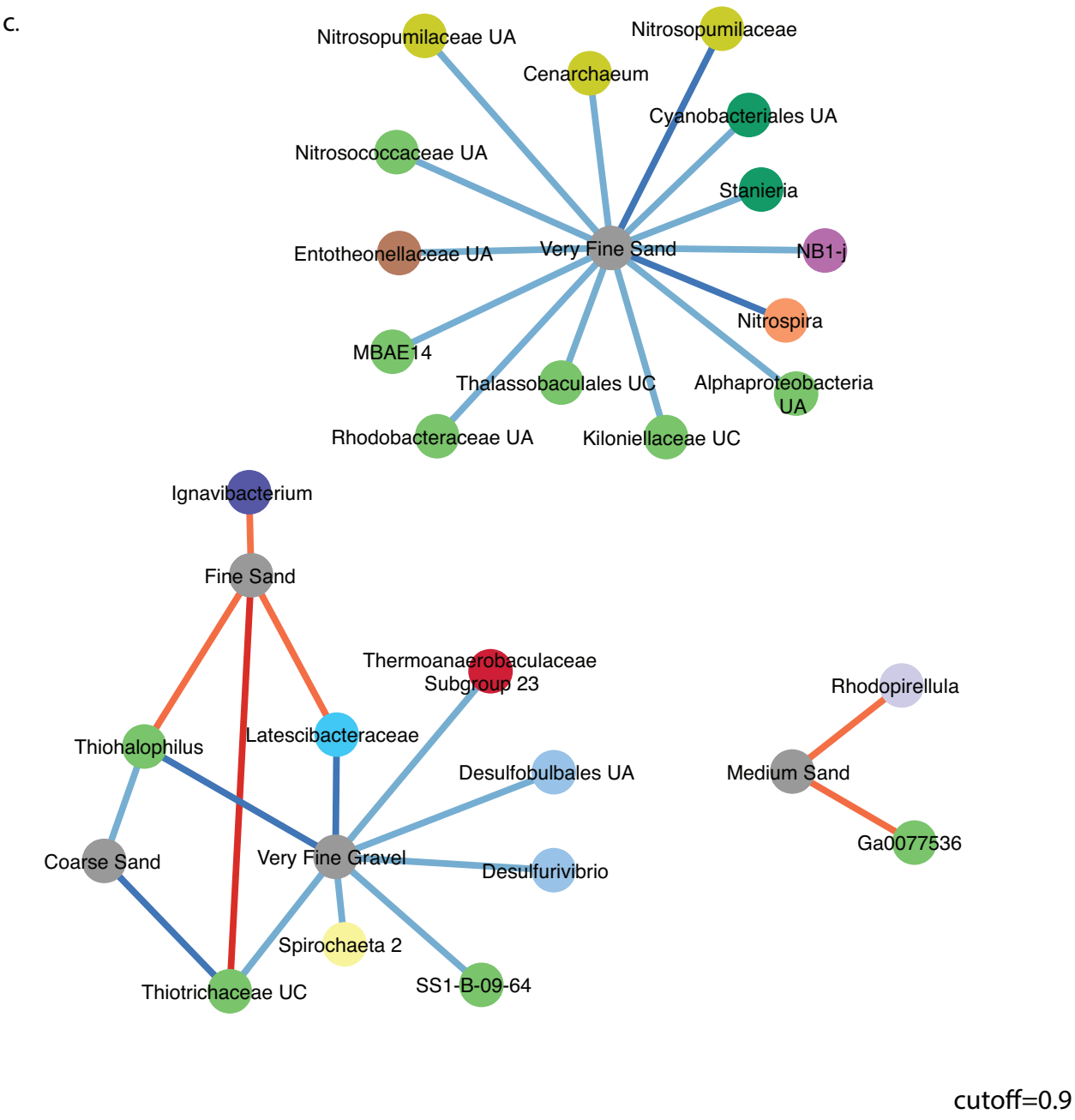

Figure 3.13: DIABLO networks, Continued.

3.4.5 Site 4 Communities and Traits

While the use of corner samples for site comparison was able to provide insight into variations in communities around Nacula Island, Fiji, I was also interested in exploring community

dynamics across a finer spatial resolution. Therefore, all samples associated with the site 4 quadrat and transects were analyzed for their microbial communities and for a subset of the sediment characteristics. By assessing samples within the 1 m^2 quadrat, I was able to determine how variable a community was on a relatively small scale. The transects, which involved sampling every meter for ten meters provided coarser spatial representation at one site. For comparative purposes, I opted to keep the transects separate for much of the analysis. This facilitated visualizing both differences between the quadrat and transects as well as differences among the transects. In total, 21,068 ASVs were detected at site 4 with an average of 1,213 ASVs per sample. Since ~25,000 ASVs were identified across the five sites in Fiji, the high richness at site 4 highlights the impact of sampling size on richness estimates. One well sampled site, in this case site 4, rivals the overall richness associated with a few samples from across five different sites. Within site 4, richness determined by observed ASVs in the quadrat samples was significantly lower (1,119 ± 77 ASVs on average ± standard error) compared to the transects (1,249 ± 25 ASVs) (Figure 3.14a; Kruskal-Wallis $H=8.368$, $p=0.004$). Phylogenetically informed richness was also significantly different, with the quadrats less rich than the transect samples (Figure 3.14b; Kruskal-Wallis $H=5.654$, $p=0.0174$). Transects were also significantly more even in the community composition (Figure 3.14c; Kruskal-Wallis $H=12.453$, $p<0.001$).

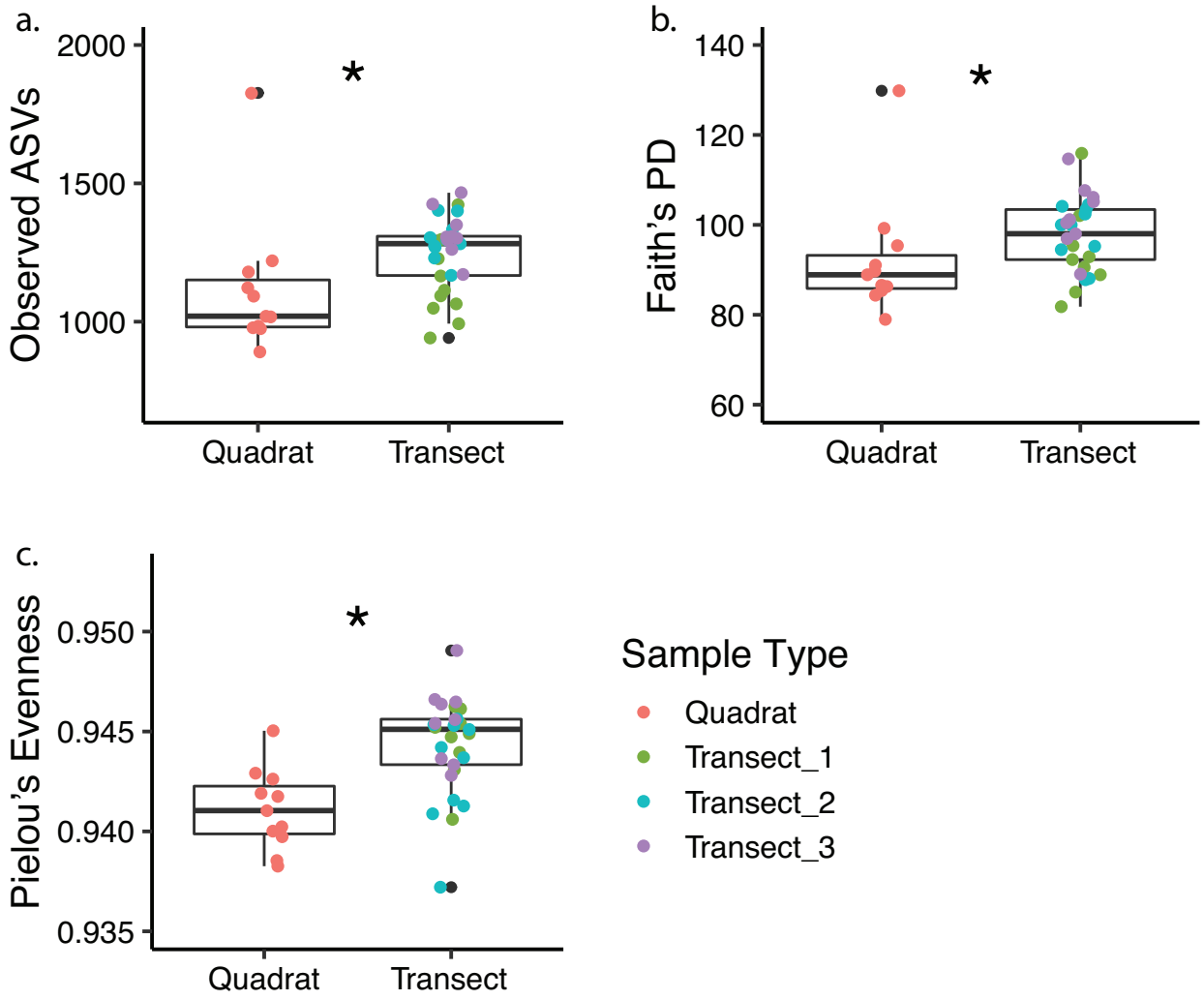

Figure 3.15: Alpha diversity measures of a) observed amplicon sequence variants (ASVs), b) phylogenetically informed richness and b) evenness of sediment microbial communities from site 4. Asterisks denote significance (alpha=0.05) in a Kruskal-Wallis test.

When taxonomy was assigned, 70 phyla (59 Bacteria, 11 Archaea) were detected across the site 4 quadrat and transect samples. The most abundant phyla were Proteobacteria (34%), Planctomycetota (15%) and Cyanobacteria (8%) for the quadrat samples and Proteobacteria (26%), Planctomycetota (19%) and Bacteroidota (9%) for the transect samples (Figure 3.15). Six of the phyla detected were unique to only one transect while all phyla detected within the quadrat had representatives within the transect samples. Thus, sampling across a coarser gradient is better

for capturing microbial diversity present at one site when compared to taking multiple samples from a much smaller patch of sediment. The unique phyla included the archaeon Hadarchaeota (Transect 2: 0.0006%), and the bacteria Methylomirabilota (Transect 2: 0.0012%), BHI80-139 (Transect 3: 0.0015%), WOR-1 (Transect 1: 0.0004%), NKB15 (Transect 1: 0.00099%) and Cloacimonadota (Transect 1: 0.0014%).

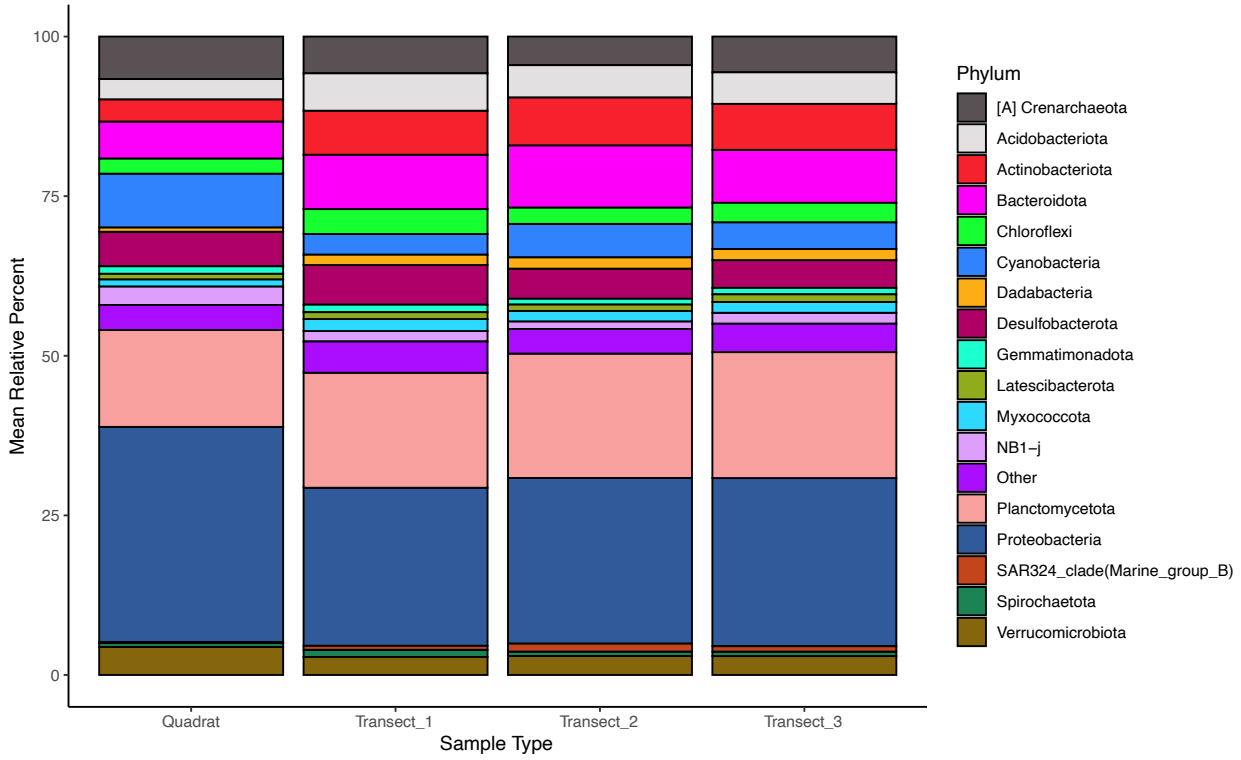

Figure 3.16: Mean relative percent of phyla detected across sample types. Archaea are denoted with [A]. Phyla that averaged >1% are denoted with their own color while all phyla that averaged <1% of the relative community were collapsed into the "other" category.

At the genus level, 941 genera (36 Archaea; 905 Bacteria) were observed at site 4 with *Woeseia* having the highest relative abundance across all samples (quadrat: 4.45%, transects: 3.43%). Interestingly, Woeseiaceae has previously been identified as a major core community member in sand grains from the North Sea (Probandt *et al.*, 2018). While an uncultured Kiloniellaceae genus (4.28%) and a genus from the clade BD7-8 (3.25%) had the next highest

relative abundances for the quadrat samples, the second and third most relatively abundant genera for transect samples were an uncultured Actinomarinales (3.24%) and *Blastopirellula* (3.10%). Only ~26 genera represented an average relative percent >1% of the communities, indicating 97% of genera represented <1% of their relative communities. Both the quadrat as well as the transects had genera unique to them with 72 genera (3 Archaea, 69 Bacteria) detected only within the quadrat samples, 48 genera (1 Archaea, 47 Bacteria) within transect 1 only, 50 genera (2 Archaea, 48 Bacteria) within transect 2 only and 46 genera (2 Archaea, 44 Bacteria) unique to transect 3 samples. This variation in unique genera highlights the heterogeneity in sediment communities. The transects together had a greater number of unique taxa when compared to the quadrat samples, however, the transects were not equivalent. Thus, sampling location impacts community composition results which in turn can affect the interpretation of how the community might be functioning.

The beta diversity of microbial communities within site 4 showed a significant delineation between quadrat and transect samples, both when comparing presence/absence of taxa (Figure 3.16a; unweighted UniFrac PERMANOVA pseudo-F=3.88, p=0.001) and when accounting for relative composition (Figure 3.16b; weighted UniFrac PERMANOVA pseudo-F=21.00, p=0.001). Transect 1 was also found to be significantly different from transect 2 and 3 in unweighted analysis while only different from the quadrat and transect 2 in weighted analyses. This suggests that while the transect communities are distinct from each other in terms of the diversity present, there is much overlap in the dominant community members present within the transect samples. Interestingly, since transect 1 and 2 were significantly different in weighted analysis as well, the dominant community members were also likely different. This is supported by the taxonomic analysis, which identified multiple differences in genera comprising >1% of the communities.

Transect 1 included an uncultured Anaerolineaceae genus (phylum: Chloroflexi), subgroup-17 of Vicinamibacteria (phylum: Acidobacteriota), and subgroup-10 of Thermoanaerobaculaceae (phylum: Acidobacteriota) while transect 2 included a genus from the SAR324 clade (Marine group B), an unannotated Alphaproteobacterial genus, and a genus from the NB1-j clade in >1% relative abundance. While a difference in six genera may not seem like much, only 27 genera are found in >1% abundance in transects 1 and 2, thus changes to a few genera can have a considerable impact on the community composition.

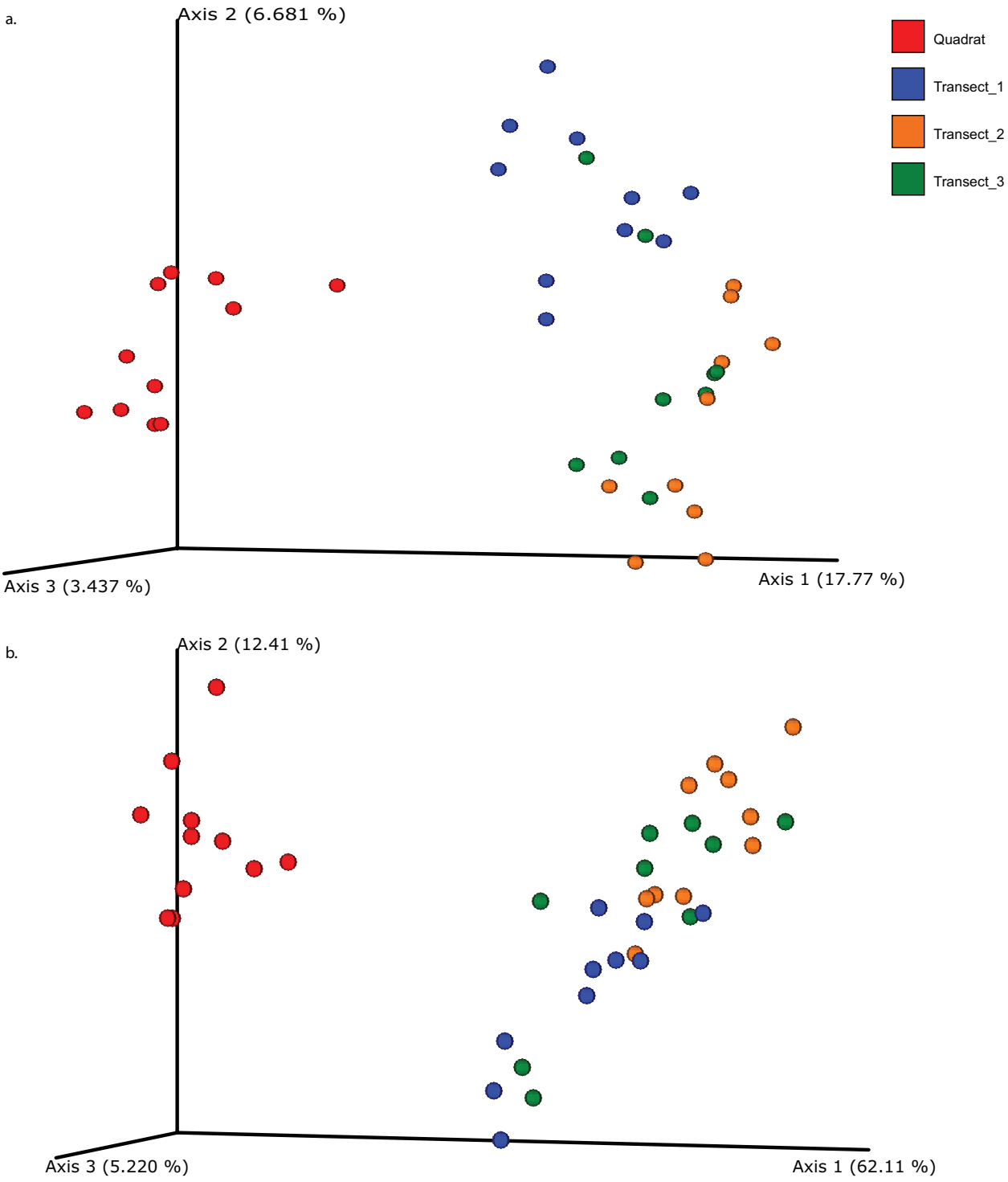

Figure 3.17: Beta diversity across quadrat and transect samples from site 4 in Fiji. a) unweighted UniFrac accounts for presence/absence of taxa while b) weighted UniFrac incorporates proportion of each taxa.

In an effort to identify which genera significantly drive beta diversity differences seen among samples, ANCOM analysis was run at the genus level. ANCOM identified 13 genera that were significantly different based on sample type (Figure 3.17). Since the majority of variation in communities was largely delineated based on sample type (quadrat versus transect) (Figure 3.16), it was not surprising that the significant taxa identified through ANCOM seemed to be based on differences between the quadrat and transects rather than among the three transects. While all significant genera could be considered relatively rare (<1% of the community), the majority (11/13) of significant genera were seen in higher relative abundance in quadrat samples when compared to the three transects (Figure 3.17). Thus, while the transect samples encompassed greater diversity (Figure 3.14), the average abundance of many taxa detected in transects was less than the quadrat. This result can also be reflective of the inherent patchiness of marine sediments. By heavily sampling one small area (1 m^2 quadrat), I may have been repeatedly sampling the same "patch" of a community. In contrast, sampling across many meters of transects would facilitate capturing a variety of microbes in difference patches- giving smaller average relative proportions of each taxa but an overall better representation of diversity. More than half of the taxa identified as significantly higher in quadrat samples were cyanobacteria and included Xenococcaceae, Cyanobacteriia, *Acrophormium*, *Symploca*, a Cyanobacteriaceae and Nodosilineaceae MBIC10086 (Figure 3.17). Thus, it is possible that the quadrat may have been set on top of a microbial mat. The only two genera identified from ANCOM that were significantly higher in transect samples were the Proteobacteria *Ahrensia* and *Photobacterium* (Figure 3.17).

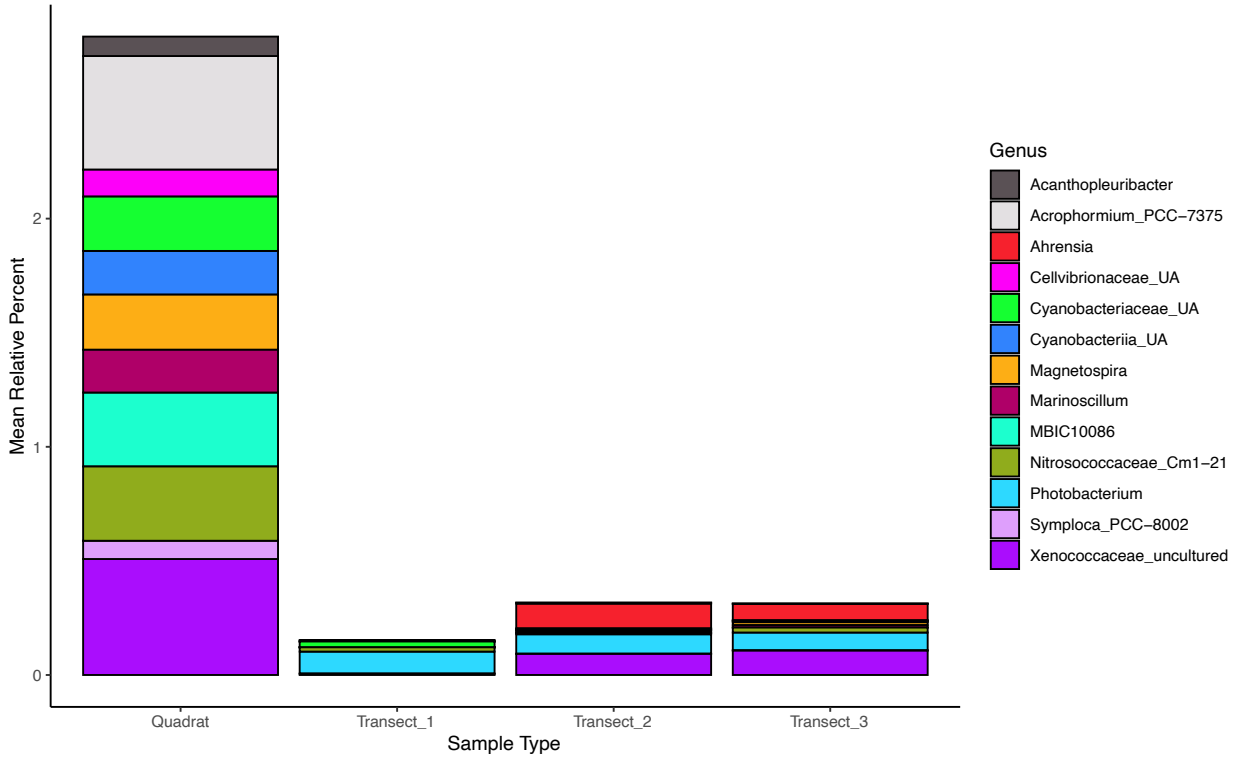

Figure 3.18: Mean relative percent of genera identified in ANCOM analysis as being significantly different based on sample type.

While the spatial resolution for sedimentary characteristics were limited to site comparisons in many cases, I was able to determine organic content and grain size composition for all site 4 samples. Interestingly, while organic content was not significantly different in site comparisons, it was significantly different within site 4. The site 4 quadrat averaged 4.40 ± 0.13% ash-free dry mass which was significantly higher than the transect samples, which averaged 4.04 ± 0.05% ash-free dry mass (Figure 3.18; Kruskal-Wallis H=8.375, p=0.004). A Mantel's test also found a significant correlation between the microbial community at site 4 and the organic content (Mantel's spearman rho=0.276, p=0.001). Therefore, it is possible that the limited samples from across sites were not representative of fine-scale differences in organic content of sediments. This could have important implications for understanding communities since organic content is an important driver of microbial community dynamics in marine sediments (Hoshino *et al.*, 2020).

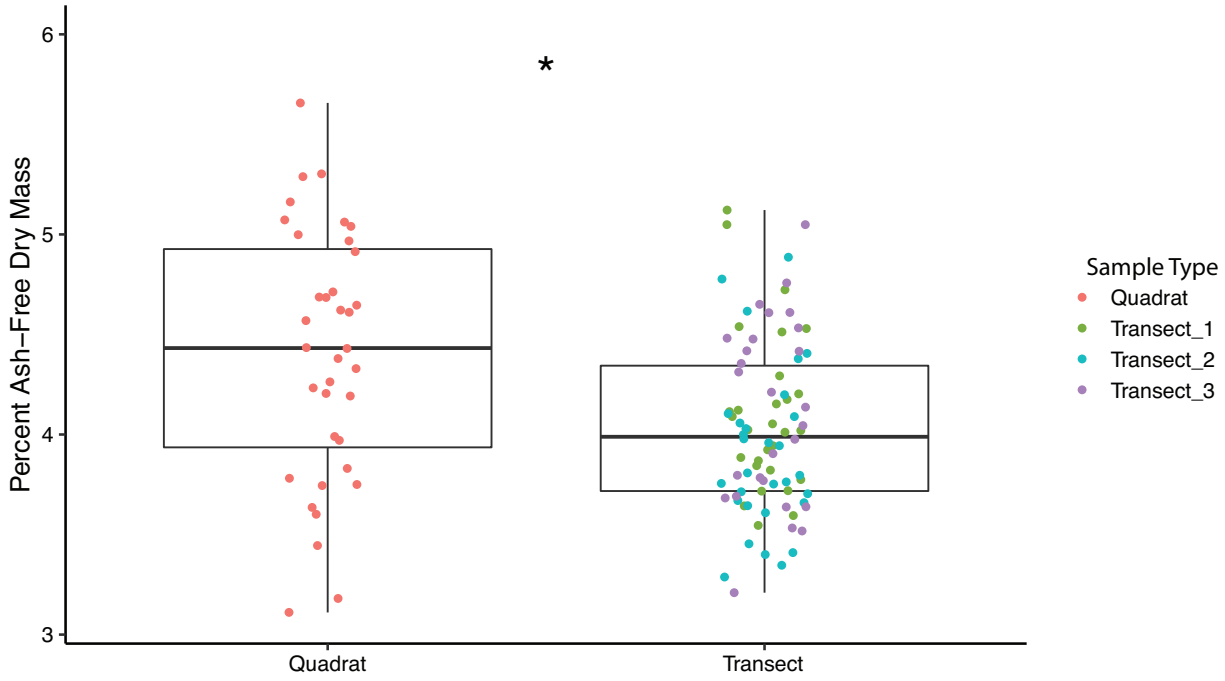

Figure 3.19: Organic content determined by percent ash-free dry mass in samples within site 4. An asterisk denotes significance (alpha=0.05).

Sediment grain size composition was also found to be significantly different between quadrat and transects from site 4 (Figure 3.19a; Adonis PERMANOVA pseudo-F=4.005, p=0.001). A subsequent pairwise adonis confirmed that the quadrat sample was different from each transect while transect 1 was also distinct from the other transects. Transect 1 had relatively more medium sand than the other samples while transects 2 and 3 had the most coarse sand and very fine gravel (Figure 3.19). The quadrat in contrast, had fairly even proportions of medium sand and very fine gravel (Figure 3.19b).

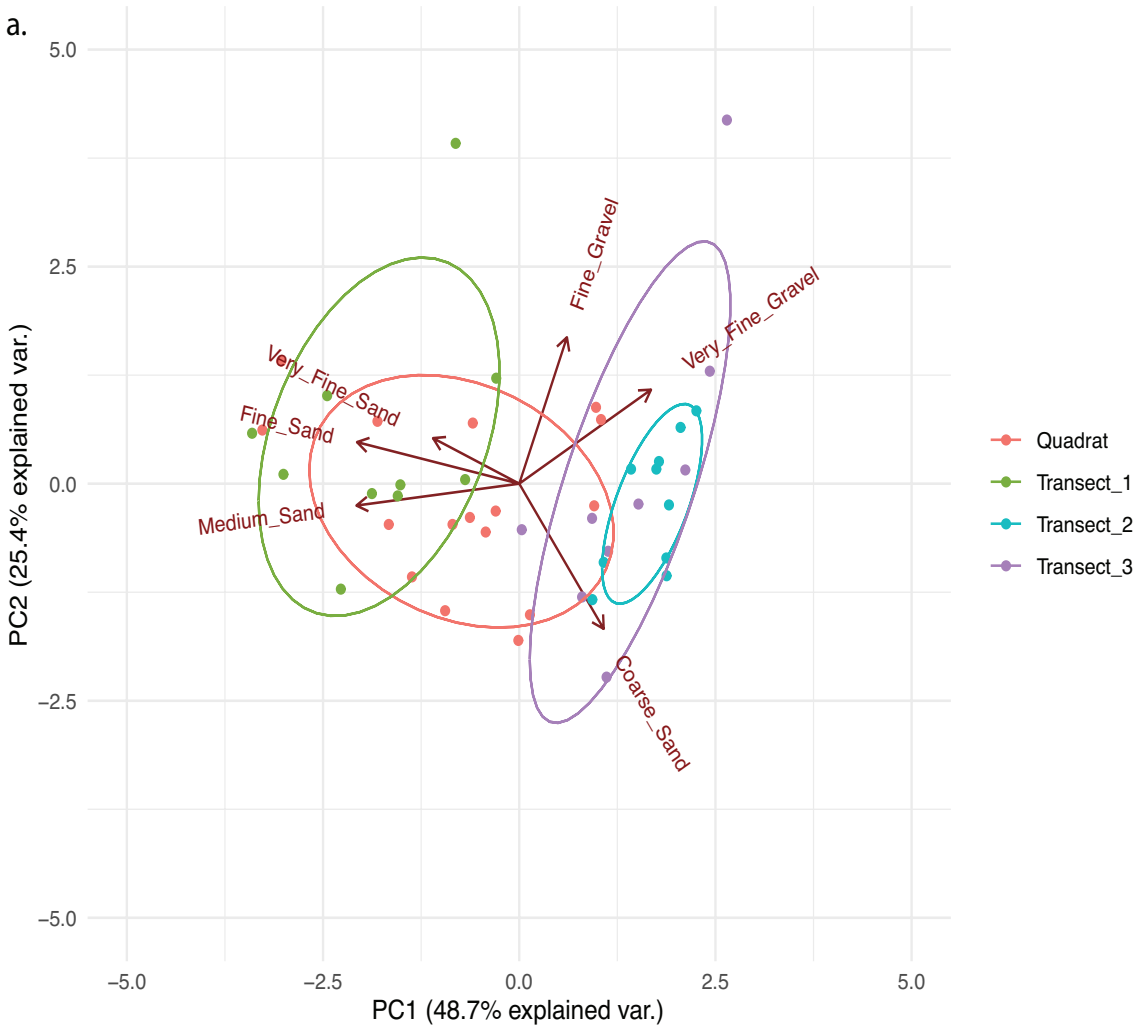

Figure 3.20: Sediment grain size composition from site 4 in Fiji. a) Biplot with sample type denoted by color and b) mean relative percent of grain size classes by site.

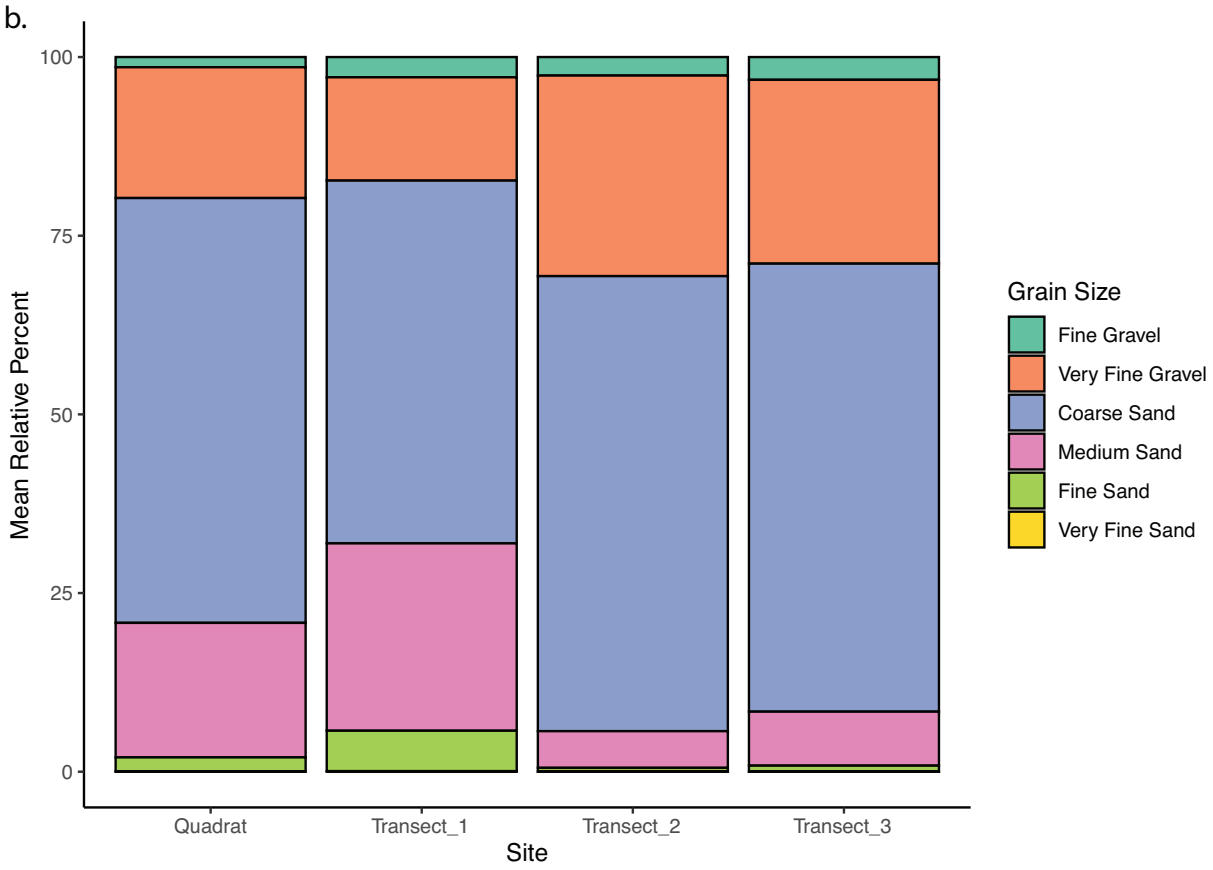

Figure 3.21: Sediment grain size composition from site 4 in Fiji, Continued.

Paired DIABLO analysis run on site 4 genera by sediment traits (organic content and sediment size composition together) with sample type as the factor, had a fairly weak overall correlation (Appendix C) but was able to detect connections between individual genera with medium and/or fine sand (Figure 3.20). Fine sand was negatively correlated with four genera including the proteobacteria *Ahrensia* and *Filomicrobium*, the plantomycete *Bythopirellula* and a member of the SAR324 clade. Medium sand was correlated with 19 genera across 10 phyla and included the four genera linked with fine sand (Figure 3.20). Three of the 19 genera associated with medium sand showed positive correlations, including the acidobacterium *Acanthopleuribacter*, an uncultured Ardenticatenales and the MBIC10086 Nodosilineaceae cyanobacterium. Since DIABLO analysis did not link organic content to taxa, the differences in

organic content may not have been significant enough to result in specific taxonomic correlations. Alternatively, there could be an impact at a different taxonomic level (e.g. certain ASVs are significantly correlated but the relationship is lost at genus level resolution), however further work would be needed to explore microbe-trait relationships across every taxonomic level. Grain size can be indicative of niche space, but also a variety of other abiotic traits (oxygen, nitrate, iron, etc.). Further research is needed to determine why medium and fine sand are significantly correlated with so many diverse genera and untangle which traits are responsible for the grain size connections.

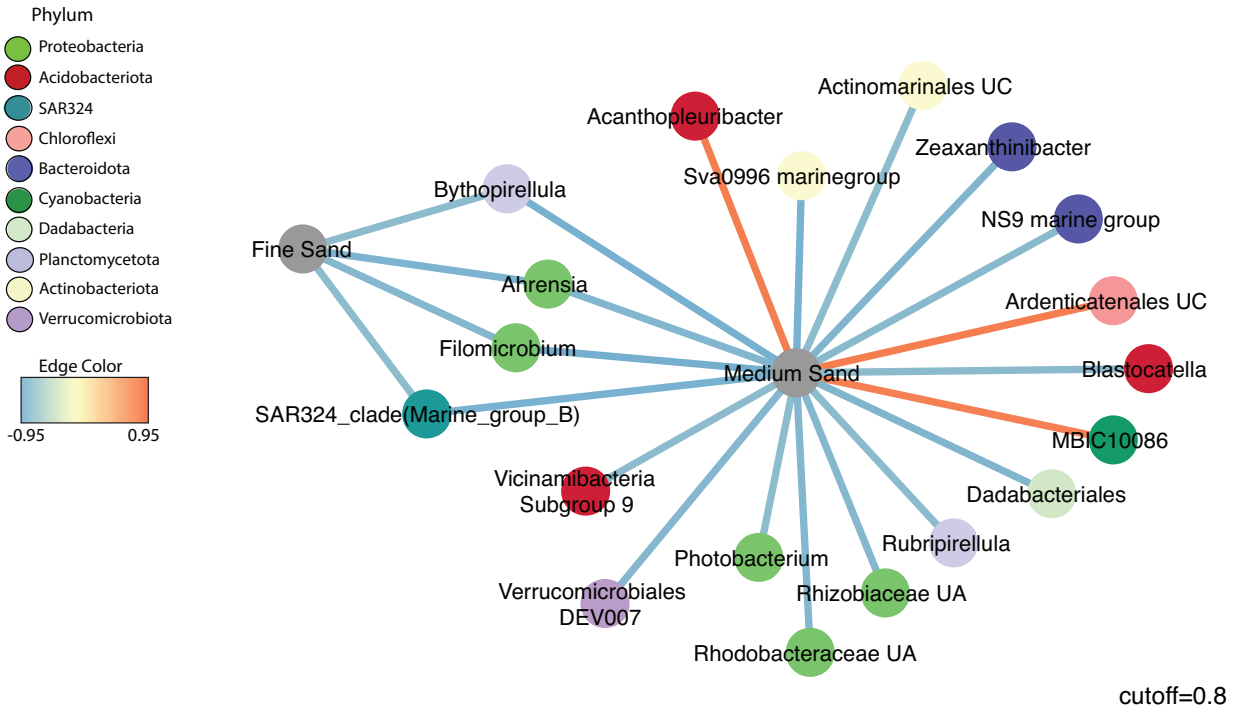

Figure 3.22: DIABLO network representing significant correlations between microbial genera and grain size from site 4 samples. "UC" = uncultured while "UA" = unannotated genus. Edges are colored based on correlation value.

Finally, I used the results from site 4 to generate a comparison of diversity by spatial scale. Shannon diversity (which incorporates both richness and evenness) was assessed between the quadrat samples at site 4 (n=11), the transect samples from site 4 (n=29) and across the 5 sites

using corner samples (n=20). Spatial scale had a significant effect on microbial diversity, with diversity significantly higher when assessing across sites compared to within site 4 (Figure 3.21; Kruskal-Wallis H=34.473, p<0.001).

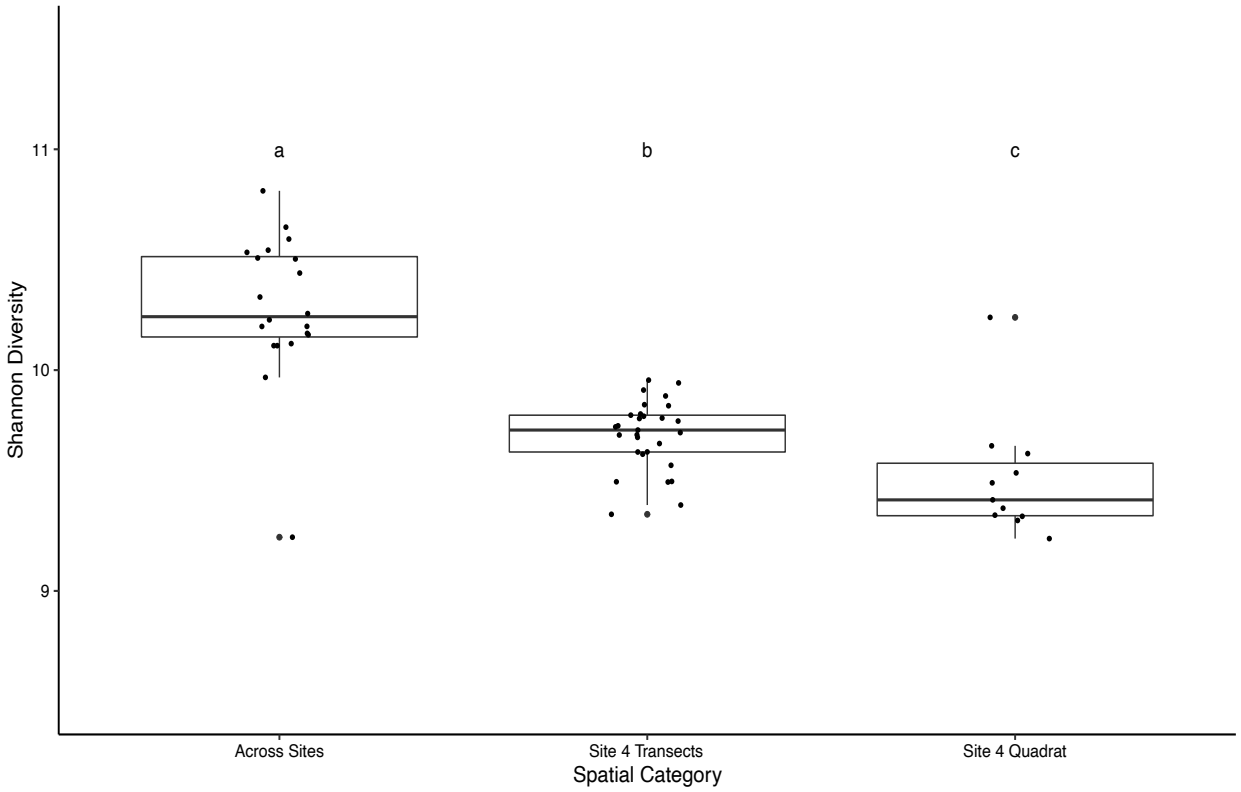

Figure 3.23: Boxplot of diversity across spatial scales sampled in this study. Across sites represent samples taken from five sites in a 12 km^2 area, site 4 transects include three 10 m transects with samples taken every meter and site 4 quadrat represents only samples from 1 m^2. Letters denote significance (alpha=0.5).

3.5 Discussion

Recent work has highlighted the vast diversity of microbes in marine sediments, from thousands of unique OTUs on a single grain of sand (Probandt *et al.*, 2018) to nearly 35,000 ASVs identified in a global survey of marine sediments (Hoshino *et al.*, 2020). My study further highlights the high microbial diversity present in sediments and how richness increases with spatial scale. When a single 1 m^2 quadrat was well sampled, 7,599 ASVs were detected. While that

number is not much higher than what Probandt *et al.* (2018) detected on single grains of sand, the addition of three transects that sampled an additional 30 m in the same area, resulted in a total of ~22,600 ASVs observed. When expanding the study further out and sampling four corner samples across five sites in Fiji (~12 km^2), almost 25,000 ASVs were observed, about 10,000 less than what was documented by Hoshino *et al*. (2020). It would be interesting in future work to compare the microbial diversity detected across global studies and assess whether there is a "core" sediment microbiome, similar to what was done with global samples of marine sponges (Thomas *et al.*, 2016), and to further compare how sediments from different habitats (e.g. latitudes, depths, etc.) differ from each other. It would also be useful to perform a meta-analysis on microbial community studies to better compare datasets since bioinformatic pipelines have changed considerably in the last few years.

Despite the limited distance between sites, microbial communities in this study were distinct from each other- both when accounting for the relative proportion of taxa and taxa presence/absence. This highlights that while there may be some overlap in the dominant members of the community (e.g. Proteobacteria represent almost half the relative community at each site), there are also significant differences within the communities. With the exception of site 5, which was a shallow sand bed nearshore and adjacent to a small seagrass bed, the other sites were fairly similar in terms of the depth and all appeared to host healthy coral reef communities. While we did not study the macroecology of the sites, the fact that we see such disparate communities speaks to how little is known about the connections between marine sediments and the surrounding ecosystems. ANCOM analysis identified 24 genera that showed significant patterns in their relative proportions across sites. Site 2, which was a bit deeper and further offshore, had significantly more Crenarchaeota compared to the other four sites. Many of the Crenarchaeota

identified are associated with nitrogen cycling, while bacteria associated with nitrogen cycling, such as *Thermodesulfovibrio*, was seen in greater relative abundance at the other sites. It would be interesting in future work to either extrapolate potential function through programs such as PAPRICA (Bowman and Ducklow, 2015) or gather metagenomic and/or transcriptomic data for these systems and assess whether the sites are distinct functionally or if some functional trade-off is occurring within communities.

Given the clear delineation in communities, I expected to see clear site delineation in the sediment metabolomes, but that was not the case. A significant pattern was not observed in metabolite richness by site and metabolomic profiles largely overlapped with each other. The one exception was site 1, which showed a significantly different pattern in the sediment metabolome based on MS1 data. Surprisingly, MetaboAnalyst only identified two metabolites as significantly different based on site. One was nearly absent at site 1 but present at the other sites, while the other was highest at sites 1 and 2. Further work is needed to identify these compounds and their potential role within the sediment communities. MS2 patterns were also consistent with overlap in the sediment metabolomes from each site. Only 3.6% of the metabolites present were assigned putative hits with the GNPS platform. While additional work needs to be done to validate these putative hits, this result highlights both the potential chemical novelty in marine sediments and the need for additional natural product representation in metabolomic databases.

This study was the second in the Jensen lab to employ resin bags for untargeted sediment metabolomics. Tuttle *et al.* (2019) was able to successfully identify compounds of interest from Belize sediments, however there are a few notable differences between the studies. Tuttle *et al.* (2019) combined their replicate extracts and fractionated them prior to running them on a more sensitive instrument than the one used in this study. Given the complexity of marine environments,

it would be beneficial for deconvolution and compound identification to fractionate my samples moving forward and run them on an instrument with higher precision. Additionally, Tuttle *et al.* (2019) removed the resin from the bags prior to extraction while I opted to extract the entire bag in an effort to increase extraction efficiency. While I did not expect a dramatic difference between extracting the resin in or out of the miracloth bags, more than half of the metabolites detected were associated with the control samples. Additionally, the chromatograms showed a large cluster of high intensity peaks associated with the resin control, so while I attempted to remove control-associated peaks in the data processing steps, I suspect that the contaminates may have swamped out much of the real signal present within the crude extracts. This is another reason why fractionation of samples would be beneficial moving forward.

Some other potential modifications that could aid in the improvement of the *in situ* resin metabolomics method include leaving the resin bags in the environment longer and doing more bags at smaller scales to get better spatial resolution. Indeed, most of the methods developed to find correlations between omics datasets require explicitly paired samples. When the resin bags were buried and retrieved in this study, I did not denote where exactly within the quadrat they were located, making it difficult for downstream correlative analyses. While sediment extractions, such as those done in Soliman *et al.* (2017), may produce greater quantities of extracts while avoiding the pitfalls of the resin capture method, I think continued efforts at improving resin metabolomics would be valuable for multiple reasons. Firstly, it would be beneficial to have a non-destructive means of sampling for sediment metabolites and secondly, direct sediment extractions include compounds from eukaryotes like nematodes, that are living within the sediment. Thus, effective resin extractions are valuable in cases where the interest is in capturing diffusible molecules being actively secreted, which was my goal in this study.

Despite the overlap in sediment metabolomes across sites, I was able to optimize the correlation between genera and metabolites through DIABLO and extract five metabolites that were significantly correlated with 13 different genera. All of the metabolites had multiple connections, but the number of connections varied. It is possible that the number of connections is indicative of the overall importance to the systems. For instance, if a metabolite only has one connection it could indicate the presence or absence of the correlated genus. However, if a metabolite has many connections it may be indicative of having a larger overall impact on the community since it is associated with multiple community members (accounting for both positive and negative correlations). For instance, metabolite 399.8657 *m/z* was negatively correlated with eight genera, suggesting that the presence of this molecule corresponds with a decrease in those taxa. It is possible that other members of the microbial community produce this compound and it in turn suppresses growth of the genera identified in network analysis. The lack of positive correlations with the metabolite can be indicative that a variety of bacteria produce this compound, that there is not a linear relationship between the producer and the compound or that the compound itself is not linked to a microbe in my dataset. All of these possibilities would need further work to explore. If a compound could be identified, it would be really interesting to perform a transcriptomics study to assess expression of the gene cluster responsible for compound production since it may be a compound that is up or downregulated depending on what other community interactions are occurring.

One of the compounds identified in network analysis was also identified in MetaboAnalyst. This metabolite (964.8176 *m/z*) was linked with an uncultured gammaproteobacterium, Nitrosomonadaceae IS-44 and an uncultured Sandaracinaceae. Sandaracinaceae are a group of myxobacteria of particular interest for natural product discovery (Garcia *et al.*, 2018), however the

negative correlation suggests that the presence of this compound is associated with a decrease in abundance of this myxobacteria. Thus, it is unlikely that the Sandaracinaceae is producing this compound. However, it would be interesting if this compound is produced by other community members in response to the presence of this natural product producer. If this compound could be identified and linked to its producer (whether a microbe or a eukaryotic producer), it would provide insight into the Sandaracinaceae which could in turn aid in culturing efforts. The development of software tools for metabolomics, such as GNPS (Wang *et al.*, 2016), and for explicitly studying connections between omics datasets, such as MixOmics (Rohart *et al.*, 2017), have undoubtedly been valuable for the field. As these tools are updated, new tools are made, and databases are expanded, more information will be gleaned from experiments such as mine. Additionally, paired and multi-omic statistical techniques are more accurate with higher sample sizes, so it would be valuable to increase the number of samples collected or possibly combine studies with similar systems to aid in linking microbes to metabolites.

Of the sediment characteristics evaluated in this study, significant correlations were found with nitrate, iron and sediment grain size. These correlations coincide with the significant differences in nitrate and iron concentrations by site, as well as the significant delineation in sites based on grain size composition. Nitrate networked with thirteen different genera, including an unannotated genus of Nitrosomonadacea, an unannotated genus of Thiotrichaceae, *Desulfurivibrio*, *Nitrospira*, *Xenococcus* and *Thiohalophilus* which are all know to be involved in microbial nitrogen cycling. For instance, all cultured representatives of the Nitrosomonadacea family are lithoautotrophic ammonia oxidizers that convert ammonia to nitrite (Prosser *et al.*, 2014). While the unidentified Nitrosomonadacea is not using or producing nitrate directly, its correlation is indicative of nitrogen metabolism occurring in these sediments. This taxon was also

seen in the highest relative abundance at site 5 which coincides with the lowest concentration of nitrate, suggesting that there are organisms present that readily uptake nitrate. *Thiohalophilus,* which accounts for >1% of the community at site 5, performs nitrite reduction to nitrate (Sorokin *et al.*, 2015) and taxa such as *Desulfurivibrio* can grow via nitrate reduction (Thorup *et al.*, 2017). Thiotrichaceae, which was observed in >4% of the site 5 community, also include members known for their ability to accumulate nitrate intracellularly (Prokopenko *et al.*, 2013). Thus, the networks identified a consortium of taxa that together can perform all the steps for denitrification.

Iron was positively correlated with four different genera, the Protobacteria clade SS1-B-09-64, Latescibacteraceae, *Thiohalophilus* and *Desulfurivibrio*. Unfortunately, SS1-B-09-64 has not been characterized, but its connection to iron suggests that there is some association with iron concentrations. Latescibacteraceae belong to phylum Latescibacterota (aka Latescibacteria aka WS3) and this phylum includes iron-reducing members (Zhang *et al.*, 2019). Interestingly, Latesibacteria were also recently identified in a metagenomic study focused on the evolution of magnetotactic bacteria, suggesting that members of this phylum may produce magnetosomes (Lin *et al.*, 2018; Amor *et al.*, 2020). *Desulfurvibrio* and *Thiohalophilus,* which were linked to both nitrate and iron are more well known for their role in sulfur cycling (Marietou *et al.*, 2005; Sorokin *et al.*, 2015), a process that is frequently coupled with nitrate reduction. Marine biogeochemical cycles including nitrogen, iron and sulfur are interconnected in many ways, adding to the difficulty in untangling microbial interactions within sediments (Rios-Del Toro *et al.*, 2018). Considering that multiple sulfur oxidizing bacteria were identified, it would be interesting in future work to also measure sulfur levels within the sediments. Recently, a study on tropical estuarine sediments identified iron as one of the most important geochemical properties impacting the structure of nitrogen fixing bacterial communities (Thajudeen *et al.*, 2017). Thus, continued research is needed

to further characterize the marine microbial lineages involved in important nutrient cycles. While this study identified taxa significantly correlated with nitrate and iron, it is also possible that some of the taxa identified are not directly linked with nutrients but rather co-occur with nitrogen and/or iron cycling taxa for some unknown reason. Microbial networks aimed at identifying correlation among taxa within communities could potentially aid in delineating which microbes co-occur. Another potential future direction would be to spike sediment communities in laboratory mesocosms with different forms of iron and/or nitrogen and assess changes in abundance of community members.

There was noticeable overlap in the genera that networked with iron and nitrate when comparing to the genera that networked based on sediment composition. For instance, all four taxa positively correlated with iron were negatively correlated with very fine gravel. When sediment traits were compared, nitrate and iron were both found to also be correlated with grain size. Thus, it is possible that some of the network links are the result of covariance in the traits measured, rather than a direct connection with the genera identified. Covariance in abiotic factors makes it difficult to untangle what traits actually drive the patterns seen here, especially in cases with multiple links. More work is needed to explore the connections identified in the grain size network, but some taxa unique to this network include *Ignavibacterium* which was positively correlated with fine sand and a variety of taxa, including Entotheonellaceae, *Crenarchaeum*, *Stanieria* and NB1-j, negatively correlated with very fine sand. *Ignavibacterium* belongs to phylum Bacteroidota (aka Bacteroidetes) and was seen in the highest relative abundance at site 5, a site dominated by fine sediment. Given site 5 was the closest to shore and quite shallow relative to the other sites, it would also likely be more exposed to physical disruptions such as wave action. Bacteroidota in turn, have many adaptations to facilitate attachment to particles (Fernández-Gómez *et al.*, 2013),

which could aid in maintaining their niche in very fine nearshore sediments. It is difficult to extrapolate the potential associations with very fine sand. Entotheonellaceae for instance, is typically associated with sponges but has also been identified as an indicator for mangrove sediments (Barreto *et al.*, 2018). Since site 5 had the greatest proportion of very fine sand and site 5 was also one of the more distinct sites in terms of macroecology (shallow nearshore site by a seagrass bed), it is likely that the negative correlation with very fine sand is indicative of a trait that was not captured in this study. Transects were performed at site 5 as well, so it would also be interesting to further assess the communities at this site and see how variable they are relative to the patterns exhibited in site 4.

Of the sediment traits evaluated in this study, grain size was the most highly correlated with microbial community composition. Multiple studies in different systems have documented the effects of grain size on soil and sediment communities, with sometimes conflicting results. For instance, richness was found to increase with coarser grain size in streams (Santmire and Leff, 2007) while richness decreased with coarser grain size in soils (Sessitsch *et al.*, 2001). Multiple studies have also documented community differences driven by grain size, with some reporting specific taxonomic preferences based on size (Sessitsch *et al.*, 2001; Jackson and Weeks, 2008; Zheng *et al.*, 2014; Hemkemeyer *et al.*, 2018). There are a variety of reasons why grain size composition is important for community structure. Grain size, sphericity and the presence of crevasses on individual grains impacts microbial colonization and available niche space (Probandt *et al.*, 2018; Ahmerkamp *et al.*, 2020). Additionally, cell abundances are also influenced by their exposure to porewater and oxygen, which in turn is influenced by the sediment composition (Ahmerkamp *et al.*, 2020). Median grain size in addition to dissolved oxygen concentrations were determined to be the driving factors in community differences across intertidal sediments in China

(Zheng *et al.*, 2014) and particulate organic matter contributed to bacterial preferences in soils (Hemkemeyer *et al.*, 2018). Additionally, the role of biotic factors, such as susceptibility to predators, has not been assessed but likely plays a role as well.

Unfortunately, I was not able to assess nutrients within the site 4 samples, but grain size was correlated with the microbial communities present. Network analysis identified a variety of genera that were linked to grain size, and in particular medium and fine sand were identified as important size classes within site 4. It would be interesting to further assess patchiness or variability in traits within one site in future work. For instance, site 4 transects had greater richness, but the majority of genera that drove differences between the transect and quadrat samples were in higher relative abundance in the quadrat samples. This result suggests that the quadrat communities may have been enriched in some taxa relative to what coarser sampling would capture, for instance the quadrat may have been set over a patch of cyanobacteria (cyanobacteria was about twice as abundant in the quadrat versus the transect samples). It would therefore be interesting to do more targeted sequencing studies that assess how diverse specific taxa are across space and determine if clonal complexes are observed.

One of the primary objectives of this study was to determine how microbial community diversity changes across spatial scales. By assessing diversity within a 1 m^2 quadrat, along three 30 m transects and across five sites spanning a 12 km^2 area, I was able to demonstrate that increasing spatial scale results in increased diversity. While this is not a surprising result, I think this work highlights the impressive diversity present in marine sediments and the value of sampling multiple sites within a location. Additionally, I believe this study also demonstrates the inherent patchiness of sediments and how sampling strategies can result in dramatically different results. For instance, if someone sampled a few spots within a one-meter squared area, their interpretation

of that sediment community would be very different than if those replicates were taken a few meters apart from each other. Furthermore, it is valuable to sequence samples to saturation, meaning that sediment samples should be routinely sequenced to depths of at least 50,000 reads per sample as opposed to the common practice of 10,000 or less. While there are trade-offs in terms of sample size and sequencing depth, inadequate sequencing results in underestimations of diversity since most ASVs are rare. Given the inherent biases in 16S rRNA amplicon sequencing (Gonzalez *et al.*, 2012; Eloe-Fadrosh *et al.*, 2016; Straub *et al.*, 2019), sequencing to saturation will still be biased, but it would give us a better picture overall of how these communities compare across habitats.

Marine sediments are estimated to contain more phylogenetically novel uncultured cells compared to most biomes including soil, seawater and host systems (Lloyd *et al.*, 2018). With improvements in sequencing technologies, environmental studies will continue to shed light onto the novel diversity present within these environments, but much work is needed to untangle the complexities of these systems. To date, we still do not know much about the microbial lineages that have been identified, even the ones that fall within well-known phyla (Baker *et al.*, 2021). This study was able to document differences in the microbial communities, environmental metabolomes and the sediment characteristics at five sites around Nacula Island in Fiji. Significant correlations were drawn between microbial genera present, specific metabolites in the environment, and traits such as nitrate and grain size. The major limitation is the lack of knowledge on these systems and in environmental metabolomes. Future work will seek to better identify the chemistry present and utilize the constant improvements in sequencing and chemical databases to re-analyze the dataset, shedding additional light on the relationships identified in this study.

Finally, the use of metagenomes to determine functional community profiles and mine for biosynthetic gene clusters will help validate the linkages identified in this study.

3.6 Acknowledgements

I would like to thank members of the Jensen Laboratory including Kaitlin Creamer, Henrique Ramalho Machado and Douglas Sweeney for their assistance in preparing for field sampling. I would also like to thank Krystle Chavarría, Dulce Guillén Matus and Mitchell Muskat for their help with chemical extractions. Thanks to Jennifer Gonzalez, Andrew Mehring, Carlos Neira and Lisa Levin for their aid in using the muffle furnace. Additionally, I would like to acknowledge my collaborators from field sampling including Deanna Beatty, Samantha Mascuch, Julia Kubanek, Mark Hay, Samson Viulu and Joape Ginigini.

This research was supported by NIH ICBG Grant U19-TW00740, the Mia Tegner Fellowship and the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1650112 to A.M.D.

Chapter 3 is coauthored with Leesa Klau, Alexander Bogdanov, Monica Cisneros and Paul R. Jensen. The dissertation author was the primary investigator and author of this chapter.

Chapter 4

Comparing sediment microbiomes and metabolomes in the fringing and back reefs of Mo'orea, French Polynesia

4.1 Abstract

Coral reefs are globally important ecosystems in decline as a result of anthropogenic factors including overfishing, eutrophication and disease. As coral-dominated habitats become degraded, they are often overrun by competing macroalgae, changing the dynamics of the entire ecosystem. The increased nutrient availability associated with these coral-algal phase shifts facilitates growth of bacteria, including pathogens, leading to the "microbialization" of coral reefs. Despite the critical role marine sediment microbial communities play in ecosystem function, little is known about their role in these shifting reef communities. The aim of this project was to establish a microbial baseline in tropical marine sediments using the fringing and back reefs of Mo'orea as a proxy for algal and coral-dominated systems. In addition to assessing changes in the microbial communities present, sediment metabolomics were also used to evaluate how the chemical landscape of these environments may be changing. Results indicate that the fringing reef is significantly less rich and diverse when compared to the microbial communities of back reef sediments. In contrast, fringing reefs exhibited richer sediment metabolomes. Beta diversity associated with microbial communities and sediment metabolomes was also found to be significantly different between back and fringing reefs. Analysis of composition of microbes (ANCOM) identified nine microbial genera associated with community differences while paired-omics analysis was used to extract significant correlations between microbes, metabolites and

environmental characteristics. Thus, this study identified key differences in fringing and back reefs within a relatively small area (~1 km^2). Future work evaluating sediments across reefs on a larger scale and through time would provide meaningful insight into how and why these communities vary and how changes to the benthic environment impacts ecosystem function.

4.2 Introduction

Coral reefs are critically important ecosystems facing global decline due to climate change, pollution, overfishing and disease (Odum 1955, Garner, Pandolfi). The shift from once healthy coral-dominated reefs to degraded algal-dominated habitats has corresponded with an increase in microbial abundances in reef waters, changes in the microbial community to include more opportunistic and pathogenic bacteria, and in the predictive metabolism of these systems (Haas *et al.*, 2016). Research over the last several years has focused on the importance of the coral microbiome for resilience to environmental stressors (Glasl *et al.*, 2019; van Oppen and Blackall, 2019). For instance, members of the mucosal microbiome were found to inhibit the growth of pathogenic bacteria (Ritchie, 2006; Krediet *et al.*, 2013; Raina *et al.*, 2016). However, coral microbiomes are not consistent across environments and can vary in relation to their surrounding benthic environment, thus impacting coral susceptibility to disease (Kelly *et al.*, 2014). The majority of research on coral reef microbial communities has focused on the microbiome of corals and the surrounding seawater. Surprisingly, marine sediments, which are major contributors to the structural benthos in some parts of coral reefs and exhibit greater bacterial richness than the seawater (Thompson *et al.*, 2017), have not been studied much in relation to their role in maintaining coral reef health.

In terrestrial environments, the microbial community in soils has been shown to play an

integral role in plant health by suppressing pathogens. Suppressive soils are well studied due to their importance for agriculture (Cha *et al.*, 2016), with both "general" and "specific" suppression activity observed (Raaijmakers and Mazzola, 2016). General suppression is the result of competitive interactions among the entire microbial community while specific suppression is attributed to the enrichment of particular community members who produce antimicrobial compounds (de Souza *et al.*, 2003; Raaijmakers and Mazzola, 2016). Environmental perturbations have been shown to not only alter the microbial diversity and community structure of the soil, but also affect their suppressive nature (Van Bruggen and Semenov, 2000), supporting the idea that secondary metabolites and community structure are intertwined. However, to my knowledge, no one has assessed if marine sediments have the natural capacity to suppress pathogens.

Similar to soils, marine sediments host rich and complex microbial communities that can impact above-ground processes. For instance, sediment microbial communities associated with *Zostera* seagrass beds were shown to inhibit growth of the invasive seaweed, *Caulerpa* (Gribben *et al.*, 2017). In contrast, sediments associated with *Caulerpa*, facilitated growth of *Caulerpa*, suggesting that below-ground microbial processes can have significant cascading impacts on the above-ground marine ecosystem (Gribben *et al.*, 2017). While changes in sediment microbial communities corresponded with sulfur metabolism, the study did not explore the role of direct interactions among the microbial taxa present in sediment communities. Given the complexity of sediment communities, it is difficult to study all of the interactions occurring in these environments, but small molecules are likely playing an important role. Microbes in marine sediments produce a variety of small molecules, known as secondary metabolites, that aid in community interactions, such as communication and competition. For example, the production of siderophores aids in the acquisition of growth-limiting iron (Butler, 2005) while the production of

antimicrobial compounds, such as bacteriocins, can act as a defense against competing bacteria (Riley and Gordon, 1999). Many natural products have been isolated from marine sediment bacteria, however, in most cases the ecological roles of these molecules are not known. One study that exposed sediment communities to organic extracts from *Salinispora* showed significant changes to the community composition, including the reduction of predatory bacteria (Patin *et al.*, 2017). Thus, secondary metabolites may act to maintain community structure and keep opportunistic bacteria in check.

Therefore, I hypothesized that microbial communities associated with "healthy" coral-dominated environments would be significantly different than sediment communities associated with more "degraded" algal-dominated environments. More specifically, I hypothesized that reefs with higher algal cover would exhibit less bacterial diversity and metabolomic complexity, thus impacting the functional roles of secondary metabolites and facilitating the proliferation of opportunistic bacteria including pathogens. While research on coral disease is extensive (Peters, 2015), the concept that microbial communities associated with coral-dominated reef sediments play an essential role in the suppression of coral pathogens and other deleterious bacteria has not been tested.

To test my hypotheses, I designed an experiment to assess sediment communities associated with reefs in the NSF sponsored Mo'orea Coral Reef Long Term Ecological Research Network (MCR LTER). This site is routinely monitored and data, including coral cover, is publicly available. However, sediment microbial communities in Mo'orea are not monitored, so this study also sought to establish baseline data in this important system. Given increased nutrient input closer to shore (Donovan *et al.*, 2020), the fringing reefs of Mo'orea are typically more algal-dominated while the back reefs tend to exhibit higher coral cover. Thus, for the purposes of my

study, I compared fringing and back reefs as a proxy for algal- and coral-dominated communities, respectively. I then utilized correlative and paired-omic techniques to determine what environmental factors were associated with changes in the communities and establish linkages between microbes and metabolites.

4.3 Methods

4.3.1 Community Sampling

To assess coral- and algal-dominated reefs without introducing confounding variables associated with different geographic locations, sites were selected within the NSF sponsored Mo'orea Coral Reef Long Term Ecological Research Network (MCR LTER). Due to time and logistical constraints, all sites in this study were in and around LTER 2, encompassing ~1 km^2. Eight sites were chosen, half associated with the fringing reef and half associated with the back reef (Figure 4.1a). The fringing reef is closer to shore and typically associated with higher nutrient levels while the back reef has lower nutrient levels and higher coral coverage (Donovan *et al.*, 2020).

For community sampling, I opted to use a similar approach to the spatial sampling scheme described in Chapter 3; all samples were collected via SCUBA. In summary, a 1 m^2 quadrat was deployed at each site. The quadrat was subdivided into 16 sections for a finer spatial gradient of sampling (Figure 4.1b). Four 4 m transects were also deployed at each site and oriented out from each side of the quadrat. Samples were collected every meter along each transect (Figure 4.1c). To sample the sediment, sterile whirl-pak bags (Nasco) were used to scoop the top layer of sediment from each collection point. Sediment samples were then transferred on ice back to the field station and approximately 7 g of sediment was transferred into a 15 ml falcon tube containing ~7 ml of RNALater ® using sterile techniques. A small amount (~1 g) of sediment was also transferred into

cryovial tubes containing seawater and 20% glycerol for future culturing efforts, if needed. All tubes and the remaining sediment in the whirl-pak bags were then stored at -20 ºC until transport back to Scripps Institution of Oceanography (SIO) on dry ice. Once at SIO, samples were kept stored at -20 ºC.

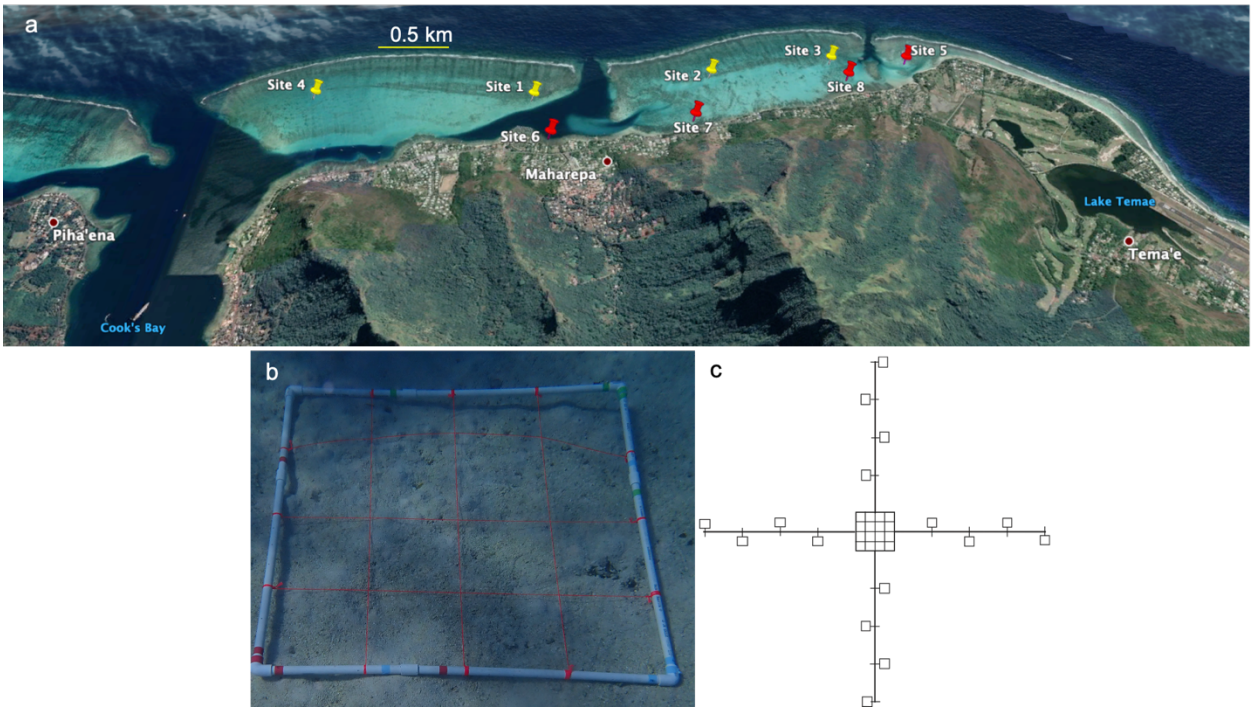

Figure 4.1: Mo'orea sampling locations and spatial scheme. a) eight sites were sampled using a b) 1 m^2 quadrat subdivided into 16 sections and c) four 4 m transects deployed adjacent to each quadrat side with sample collection every meter. Back reef sites are denoted with yellow markers while fringing reefs are denoted with red markers.

4.3.2 Environmental Variables

Metadata associated with each site along with depth and distance from shore were gathered while in the field. Heatmaps of percent coral cover (*Acropora* and *Pocillopora*) and nutrient estimates based on N:C from prior studies (Burkepile and Adam, 2019; Donovan *et al.*, 2020) were used to extract estimates for data related to coral and algal cover (Table 4.1).

Table 4.1: Environmental trait values associated with each sampling site in Mo'orea, French Polynesia. Sample numbers are denoted in brackets. Percent coral cover (*Acropora* and *Pocillopora*) and the N:C nutrient estimate was derived from Donovan *et al.* 2020 and Burkepile and Adams, 2019.

Site	Depth (m)	Distance to shore (m)	% *Acropora* cover	% *Pocillopora* cover	Nutrient estimates (N:C)	Reef Type Designation
Site 1 [MO18_1 - 32]	1.83	315	0	25	0.025	Back Reef
Site 2 [MO18_33 - 64]	2.13	426	20	70	0.0225	Back Reef
Site 3 [MO18_65-96]	1.83	322	20	50	0.0225	Back Reef
Site 4 [MO18_97 - 128]	3.05	694	0	30	0.02	Back Reef
Site 5 [MO18_129 - 160]	1.83	116	0	65	0.0225	Fringing Reef
Site 6 [MO18_161 - 192]	0.61	44.2	0	10	0.03	Fringing Reef
Site 7 [MO18_193 - 224]	1.83	55.7	0	25	0.0225	Fringing Reef
Site 8 [MO18_225 - 256]	1.83	130	0	25	0.0225	Fringing Reef

4.3.3 DNA Extractions and Sequencing

For DNA extraction and sequencing, all quadrat samples (n=128) were processed as well as the tips of all transects (n=32). Since there was room on the sequencing run for additional samples, the mid-point transect samples from site 3 (n=4) and all remaining transect samples from site 6 (n=12) were also processed for sequencing (site 3 was representative of greater coral cover while site 6 had the highest algal cover). This resulted in a total of 176 samples. For DNA

extraction, ~2 g of sediment from each sample in RNALater ® was sent to the Knight laboratory in collaboration with the Center for Microbiome Innovation at UC San Diego. Extractions were performed using the Qiagen MagAttract PowerSoil DNA KF kit. Samples were then sequenced using the Illumina MiSeq PE150 platform with the V4 primers recommended by the Earth Microbiome Project (Thompson *et al.*, 2017) 515F: GTGYCAGCMGCCGCGGTAA (Parada *et al.*, 2016) and 806Rb GGACTACNVGGGTWTCTAAT (Apprill *et al.*, 2015).

4.3.4 Community Analysis

Community analyses were performed using a similar pipeline to Chapter 3. Sequences were first imported into QIIME2 (version 2020.2) (Bolyen *et al.*, 2019) and denoised with DADA2 (Callahan *et al.*, 2016). Default DADA2 parameters were used with p-trunc-len-f 150 and p-trunc-len-r 150 for all samples. To reduce the computational load, the DADA2 step was performed with samples divided into groups based on site. After denoising, the resulting tables and rep-seq files were merged using the feature-table merge and feature-table merge-seqs commands in QIIME2. Merged files were then used when creating phylogenetic trees and performing taxonomic assignment. Taxonomic assigned was done with the SILVA v138 SSU NR99 515f-806r classifier (Quast *et al.*, 2013) based on referenced reads and taxonomy generated through RESCRIPt (Robeson II *et al.*, 2020). Sequences identified as chloroplasts, mitochondria or eukaryota were filtered from the dataset and any remaining singletons were also removed.

Unfortunately, due to high variability in sequencing depth across samples, rarefaction depth for comparative analyses was less than ideal. Based on rarefaction curves (Figure 4.2), I selected to rarefy at a depth of 16,055 reads, which was close to saturation, but resulted in the loss of six samples across five different sites (Samples 033 & 060 from site 2, 104 from site 4, 138

from site 5, 171 from site 6 and 207 from site 7). Alpha and beta diversity analyses were performed using QIIME2 (v2020.2). Alpha diversity was assessed based on observed amplicon sequence variants (ASVs) and the Shannon Diversity metric, which accounts for richness and evenness of samples (Shannon and Weaver, 1964). Alpha diversity values were imported into RStudio with R version 4.0.3 (R Core Team, 2019) and visualized as boxplots using ggplot2 (Wickham, 2016). The non-parametric Wilcoxon test and Kruskal-Wallis test (Kruskal and Wallis, 1952) were used in R to statistically compare alpha diversity across reef types and sites, respectively. When appropriate, a Dunn's post-hoc test was then performed to determine which groups were significant. For continuous variables, the diversity alpha-correlation function in QIIME2 was used with Spearman's correlation to assess how richness and diversity varied with depth, distance to shore, percent *Acropora* cover, percent *Pocillopora* cover and nutrient (N:C) concentrations.

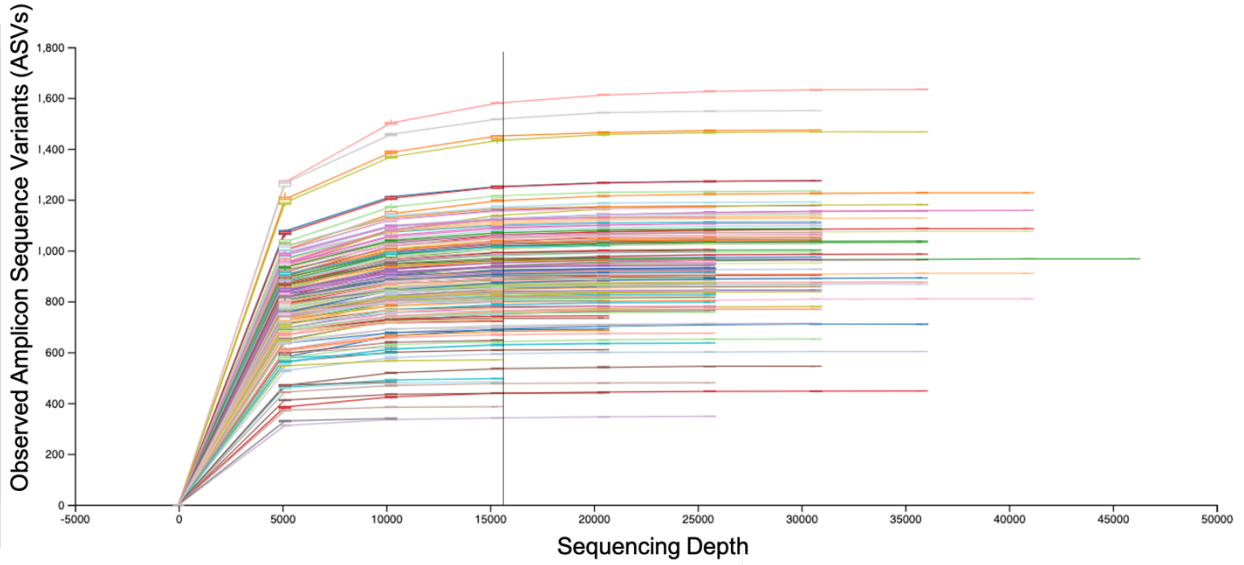

Figure 4.2: Alpha-rarefaction plot of observed amplicon sequence variants (ASVs) by sequencing depth across samples. The vertical line denotes the point of rarefaction for comparative analyses.

Beta diversity analyses were performed using the phylogenetically informed UniFrac measure, with both unweighted (based on presence/absence of taxa) and weighted (which accounts

for proportion of taxa) (Lozupone and Knight, 2005) results visualized as PCoA plots using Emperor (Vázquez-Baeza *et al.*, 2013). For testing statistical differences in beta diversity, multivariate PERMANOVAs were performed in QIIME2 with categorical variables such as site and reef type (Anderson, 2001). For continuous variables, such as depth and distance to shore, variables were converted into Euclidean distance matrices with the metadata distance-matrix command in QIIME2. Mantel tests with spearman's correlation were then performed in QIIME2 to determine significance with Spearman's correlation. In an effort to determine which microbial genera drove differences in reef type patterns, the analysis of composition of microbes (ANCOM) (Mandal *et al.*, 2015) was performed using QIIME2.

4.3.5 Metagenomics

Twenty samples from the 16S rRNA extractions were also sequenced via HiSeq for metagenomic analysis in partnership with UCSD Center for Microbiome Innovation. Two center quadrat samples were selected for each site and the remaining four samples represent the transect tips from site 6. Samples were sequenced in triplicate and averaged ~70 M reads/sample after quality filtering with BBMap. Dr. Alexander Chase analyzed the primary metabolism and glycoside hydrolase enzymes for the samples. These results were visualized as a heatmap with scaled abundance for each pathway or enzyme.

4.3.6 Metabolomics

In an effort to capture the sediment metabolome, I utilized an updated version of the small molecule *in situ* resin capture (SMIRC) method described in Chapter 3. The miracloth material used in Chapter 3 seemed to contribute to the large control peaks, so I replaced it with heat-sealed

nitex following previous targeted studies (Lane *et al.*, 2010). Additionally, while large resin bags provided increased quantities of extracts, they were cumbersome in the field and difficult to pair with sediment community samples. Therefore, a metabolomics scheme was designed with small 10 g resin bags so each resin bag could be paired with an individual sediment community sample. I also included additional controls including seawater controls, which were resin bags attached to the site marker buoy line, and additional method controls.

All resin bags were prepared at SIO and activated following the same protocol used in Chapter 3 using HPLC-grade methanol and Milli-Q water. Rinsed resin bags were then stored in plastic bags for transport. Since there were 256 resin bags set to be deployed with the sediments at each site, the bags were sealed with a loop at the top that a string would fit through. Prior to deploying the resin bags in the field, I strung bags in the proper distance increments so that there would be four strings of bags for within the quadrat and four strings of bags associated with the four transects at each site. Resin bags were transported in Ziploc ® bags and buried via SCUBA within the surface sediment after sediment samples were collected. Tent stakes were used to anchor the ends of each string down into the sediment. Resin bags were left for 24 hours and then were retrieved and transported on ice back to the Gump Marine Station. The resin bags were then individually bagged in plastic to maintain the proper sample ID number. Resin bags were stored at -20 °C until transport on dry ice back to SIO where they were stored at -20 °C until extractions occurred.

4.3.7 Chemical Extractions and Mass Spectrometry

In order to process the resin bags in a high-throughput manner, I opted to extract them while they were still in their nitex bags. Prior to extractions, resin bags were left to thaw overnight

at -4 °C. A vacuum-flash apparatus was set-up with 50 ml plastic syringes screwed to the ports and syringe plungers removed. Each nitex bag was then inserted into a syringe and flushed with 200 ml of Milli-Q water to remove excess salts. Next, 20 ml of HPLC-grade MeOH was added to the syringe and allowed to soak for five minutes before ports were opened and vacuum applied. This process was repeated with 20 ml of HPLC-grade acetone. Test tubes with crude extracts were then placed in a speedvac and concentrated down. Concentrated extracts were then filtered through Whatman 41 filter paper while being transferred into 20 ml glass scintillation vials. Extracts were dried under nitrogen and crude extract mass was determined.

Crude extracts were then redissolved with 200 ul of methanol filtered through 0.2 um spin columns and added into HPLC vials. HPLC-grade water was then added to bring the solvent concentration to 50% MeOH. Samples were then delivered to the Dorrestein Laboratory for analysis in association with the UCSD Center for Microbiome Innovation. Samples were run using a Bruker Daltonics maXis Impact and C18 RP-UHPLC. LC-MS/MS data was acquired with positive polarity acquisition.

4.3.8 Metabolomics Analysis

Files in .mzXML format were uploaded into MZMine 2 v3.27 (Pluskal *et al.*, 2010) for processing following the same pipeline as in Chapter 3. Noise levels were set to 1.0E3 and 1.0E1 for MS1 and MS2 data, respectively. Minimum height was 3.0E3 for chromatogram building and deconvolution was performed with a baseline cutoff of 1.0E3, minimum height of 3.0E3, *m/z* tolerance to 20.0 ppm and 0.01 minimum time for building the chromatograms. For chromatogram deconvolution the baseline cutoff was set to have a minimum peak height of 1.5E3, 0.01 to 3.0 min duration, and 1.0E3 baseline. Median *m/z* center calculation with a *m/z* range of 0.02 and a

0.1 min retention time was used for MS2 pairing. Isotopes fragmentation patterns were recognized with a maximum charge of three, 0.1 min time tolerance and 20 ppm *m/z* tolerance with the isotopic peak grouper. Peaks were then aligned using join aligner with 75% weight for *m/z*, 25% weight for retention time, 0.1 min retention time and 20 ppm *m/z* tolerance. Finally, peak list rows filter was used to set 2 peaks minimum for an isotopic pattern and 2 peaks minimum in a row. The resulting peak lists was then exported into a .csv file for analysis with MetaboAnalyst v5 (Chong *et al.*, 2019).

Prior to analysis with MetaboAnalyst, MS1 features were filtered to remove any features also identified in the resin controls or the initial solvent blank. This process reduced the total features from 5,640 to 2,370. Since there were less than 5,000 features, none were filtered in the initial upload to MetaboAnalyst. Data were then log transformed and scaled using pareto scaling. A non-parametric Wilcoxon test was then performed comparing back reef samples with fringing reef samples (alpha=0.05 and unequal group variance selected). Volcano plot and random forest analyses were used to extract information on the number of metabolites driving differences in fringing versus back reef sediments. Normalized MS1 data was also used in Mantel tests to assess correlations between metabolites, microbial communities and environmental characteristics. Finally, data was uploaded into the Global Natural Product Social Network (GNPS) platform (Wang *et al.*, 2016). GNPS utilizes MS/MS fractionation patterns to identify compounds and connect related molecules together. Classical molecular networking was used with the precursor ion mass tolerance and fragment ion mass tolerance set to 0.02 Da. The cosine score was reduced to 0.6 in line with maintaining a false discovery rate below 1%. After the network was generated, all nodes associated with control samples were removed.

4.3.9 Paired-Omic Analyses

In an effort to extract significant connections between microbes, metabolites and environmental characteristics, DIABLO analysis was performed using the MixOmics package (Rohart *et al.*, 2017) in R. DIABLO uses a supervised multivariate approach with sparse partial least squares discriminant analysis (sPLS-DA). Microbial data (collapsed based on genus level annotation) was normalized with center log-ratio (clr) transformation with the microbiome package (Lahti and Shetty, 2019) in R, environmental data was scaled with the base scale function in R and normalized MS1 data from MetaboAnalyst was used for correlative analyses. DIABLO networks were first made with all three datasets at the default correlation cutoff = 0.7 and the supervising factor set to reef type. Due to a high number of connections between metabolites and one environmental parameter, the network was rerun with a stricter cutoff (0.8) to better visualize the important connections for that cluster. To better assess direct connections between microbial genera and metabolites, DIABLO analysis was rerun with just the genera and metabolite datasets and at a cutoff = 0.6. This network was run with sample ID, site and reef type as the supervising factor to compare resulting network connections.

4.4 Results

4.4.1 Sediment Microbial Communities of Mo'orea

Sequencing of 16S rRNA genes from 176 sediment samples collected at eight sites within ~1 km^2 of the LTER in Mo'orea revealed a total of 29,612 amplicon sequence variants (ASVs). The average number of ASVs was significantly different based on site (Kruskal Wallis H=30.627, $p<0.001$) and ranged from 786 ± 34 ASVs (site 6) to 1,092 ± 43 ASVs (site 1) (Figure 4.3a). Diversity was also determined to be significantly different based on site (Figure 4.3b; Shannon

Diversity Kruskal Wallis H=29.933, p<0.001). Of the eight sites that were surveyed, four were categorized as fringing reefs (sites 5 - 8) while the other four were categorized as back reefs (sites 1- 4). When alpha diversity was assessed as a function of reef type, ASV richness was significantly higher in the back reef when compared to the fringing reef (Figure 4.4a; Wilcoxon test W=4782.5, p<0.001). Diversity was also significantly higher in the back reef when using Shannon Diversity which accounts for both richness and evenness (Figure 4.4b; Wilcoxon test W=4319, p=0.026). Both depth and distance from shore were significantly correlated with ASV richness but were not correlated with diversity (Table 4.2).

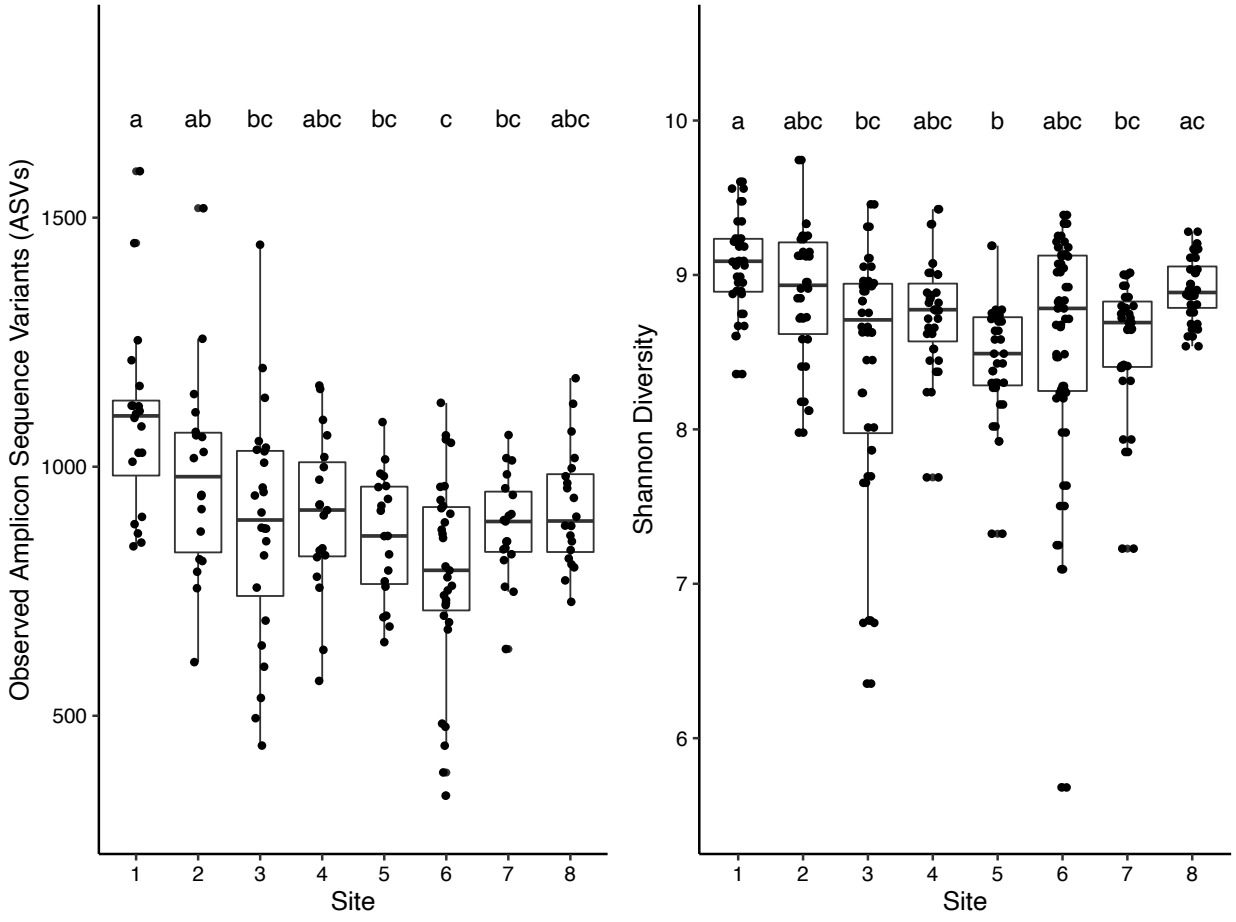

Figure 4.3: Alpha diversity of sediment microbial communities from eight sites around the Mo'orea LTER 2 region. a) richness of amplicon sequence variants (ASVs) and b) diversity determined with Shannon Diversity metric. Letters denote significant (alpha=0.05).

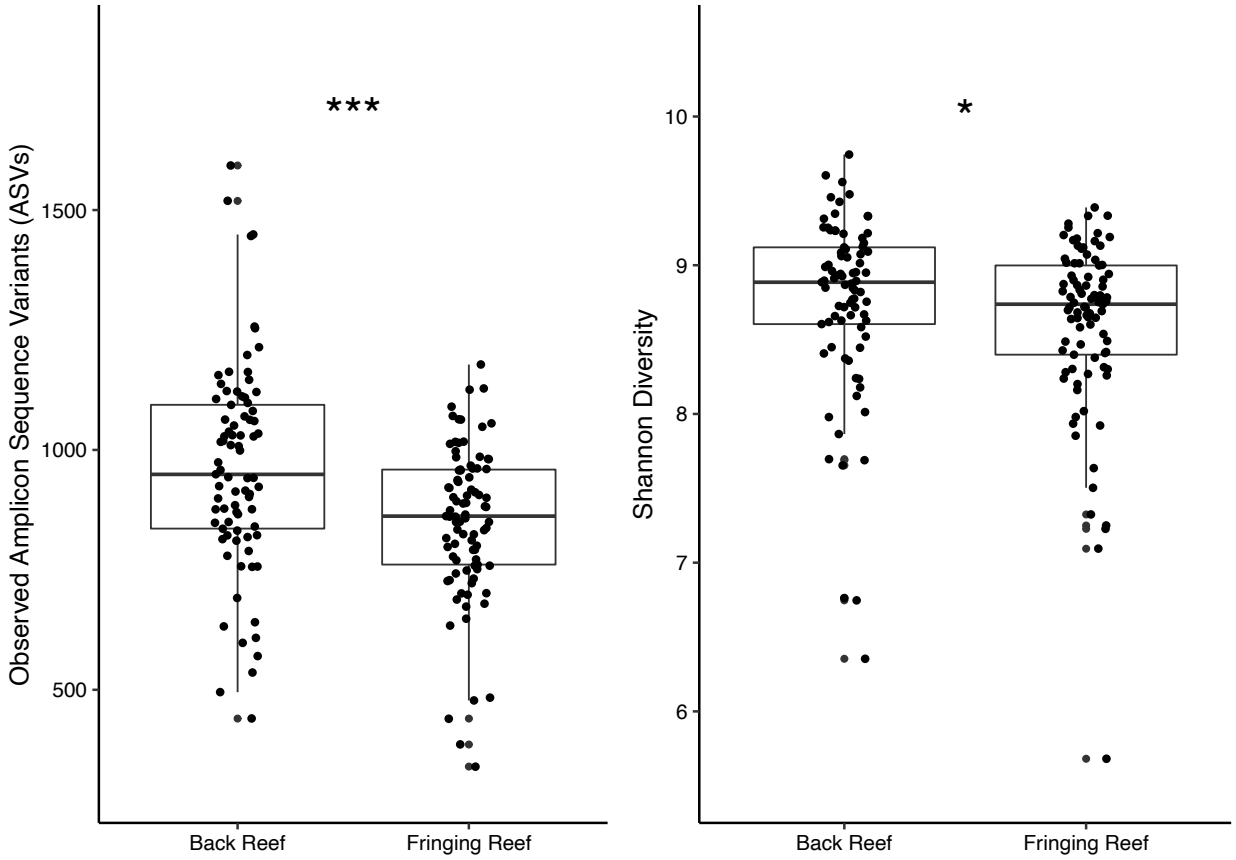

Figure 4.4: Alpha diversity of fringing and back reefs around the Mo'orea LTER 2 region. a) richness of amplicon sequence variants (ASVs) and b) diversity determined with Shannon's Diversity metric. Significance from a Wilcoxon test (*) = $p<0.05$ while (***) = $p<0.001$.

Table 4.2: Correlations between alpha diversity metrics and environmental traits determined with Spearman's correlation. (*) = p<0.05, (**) = p<0.005, (***) = p<0.001.

Alpha Diversity Metric	Trait	Spearman's rho	P-value
Observed ASVs	Depth (m)	-0.2133	0.0052 (*)
	Distance to shore (m)	0.2492	0.0010 (**)
	% *Acropora*	0.0705	0.3613
	% *Pocillopora*	0.1179	0.1258
	N:C estimates	-0.0514	0.5056
Shannon Diversity	Depth (m)	-0.0596	0.4397
	Distance to shore (m)	0.1235	0.1087
	% *Acropora*	-0.0083	0.9141
	% *Pocillopora*	-0.0978	0.2045
	N:C estimates	0.1086	0.1585

Beta diversity was also found to be significantly different based on site when accounting for relative proportion (Figure 4.5a; Weighted UniFrac PERMANOVA pseudo-F=7.629, p=0.001) and when assessing the presence/absence of ASVs (Figure 4.5b; Unweighted UniFrac PERMANOVA pseudo-F=7.629, p=0.001). Interestingly, with unweighted UniFrac, site clustering was more readily distinct by eye, with site 6 in particular standing out from the rest of the sites (Figure 4.5b). Reef type was also considered a significant factor in beta diversity using weighted (Figure 4.6a; Weighted UniFrac PERMANOVA pseudo-F=6.997, p=0.001) and unweighted analyses (Figure 4.6b; Unweighted UniFrac PERMANOVA pseudo-F= 5.051, p=0.001). The weighted PCoA showed the greatest variation along axis 1 (33.08%); however, when environmental traits were visualized, there was no obvious trait responsible for the variance along that axis. Similar to the unweighted PCoA by site, the unweighted PCoA by reef type

revealed a more distinct clustering pattern, suggesting that the overlap in proportionally dominant taxa may obscure some of the significant patterns (Figure 4.6b). Mantel tests were used to assess environmental traits in relation to beta diversity. Depth, percent cover of *Acropora*, percent cover of *Pocillopora* and nutrient estimates were all significantly correlated with weighted beta diversity while depth, distance to shore, percent cover of *Pocillopora* and nutrient estimates were correlated when beta diversity remained unweighted (Table 4.3).

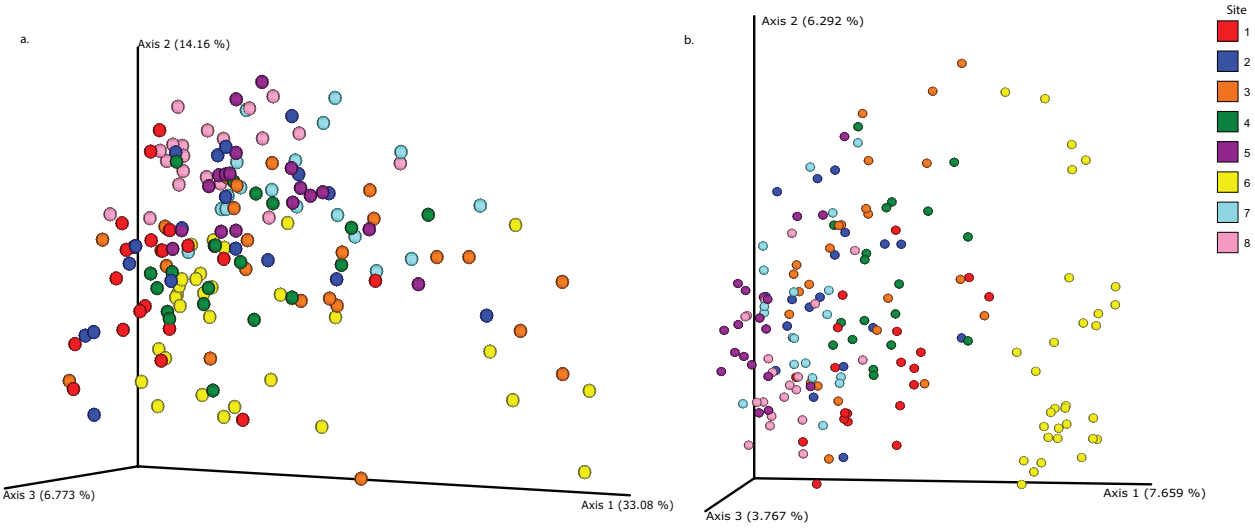

Figure 4. 5: Beta diversity of sediment microbial communities from eight sites around the Mo'orea LTER 2 region when assessed with a) weighted UniFrac and b) unweighted UniFrac. Site was significant for both weighted (p=0.001) and unweighted (p=0.001) UniFrac. Sites 1 – 4 were back reef while sites 5 – 8 were fringing reef.

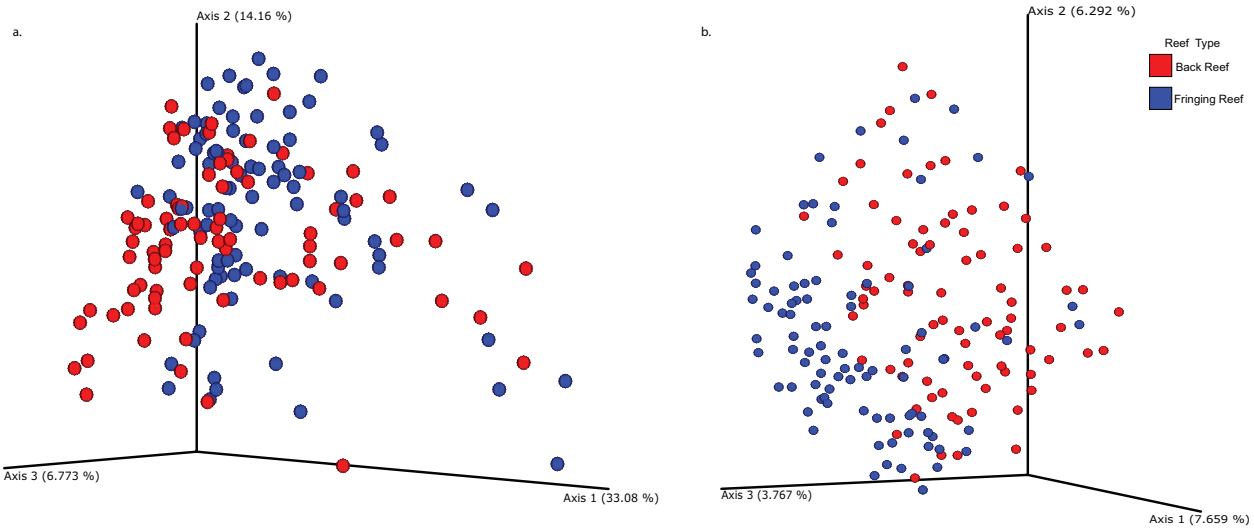

Figure 4.6: Beta diversity of sediment microbial communities from fringing and back reefs around the Mo'orea LTER 2 region when assessed with a) weighted UniFrac and b) unweighted UniFrac. PERMANOVA results by reef type were significant for both weighted (p=0.001) and unweighted (p=0.001) UniFrac.

Table 4.3: Correlations between beta diversity metrics and environmental traits. Mantel tests with Spearman's correlation were used to assess significance. (*) = p<0.05, (**) = p<0.005, (***) = p<0.001.

Beta Diversity Metric	Trait	Spearman's rho	P-value
Weighted UniFrac	Depth (m)	0.2130	0.001 (**)
	Distance to shore (m)	0.0509	0.079
	Percent *Acropora* cover	0.0991	0.012 (*)
	Percent *Pocillopora cover*	0.1753	0.001 (**)
	N:C	0.2407	0.001 (**)
Unweighted UniFrac	Depth (m)	0.5620	0.001 (**)
	Distance to shore (m)	0.2149	0.001 (**)
	Percent *Acropora* cover	0.0378	0.310
	Percent *Pocillopora* cover	0.2724	0.001 (**)
	N:C	0.5647	0.001 (**)

When taxonomy was assigned to ASVs, 73 phyla (14 Archaea, 59 Bacteria) were identified. Phyla observed in the highest relative amounts included Proteobacteria (40%), Cyanobacteria (17%), Bacteroidota (11%), Desulfobacterota (10%) and Planctomycetota (4%). The majority of phyla, ~85%, were found on average to be <1% of the microbial community present across sites. At the genus level, ASVs were assigned to 1,247 genera (48 Archaea, 1,199 Bacteria). Seventeen genera averaged >1% of the sediment communities, with an unannotated Rhodobacteraceae (4.16%), *Woesia* (4.14%), Gammaproteobacteria unannotated (3.71%), an

uncultured Desulfosarcinaceae (3.70%) and *Xenococcus* PCC-7305 (2.89%) in the highest relative amounts. At the genus level, 98.6% of genera averaged <1% of the community and 59.3% of the community averaged <0.01%. To determine which genera were associated with differences between the fringing and back reef, ANCOM analysis was employed. ANCOM identified nine genera including members of the Proteobacteria, Cyanobacteria, Actinobacteriota, and Desulfobacterota (Figure 4.7). *Actinomarinales*, an uncultured Chromatiaceae, Gammaproteobacteria Run-SP15, an uncultured Granulosicoccales and *Stanieria* PCC-7437 were significantly more abundant in back reef samples. In contrast, *Actibacterium*, an uncultured Desullfobulbaceae and *Planktothricoides* SR001 were in higher abundances in the fringing reef (Figure 4.7). All nine of these taxa were in <0.4% of the communities on average, suggesting that rare microbes play an important role in distinguishing sediment microbial communities.

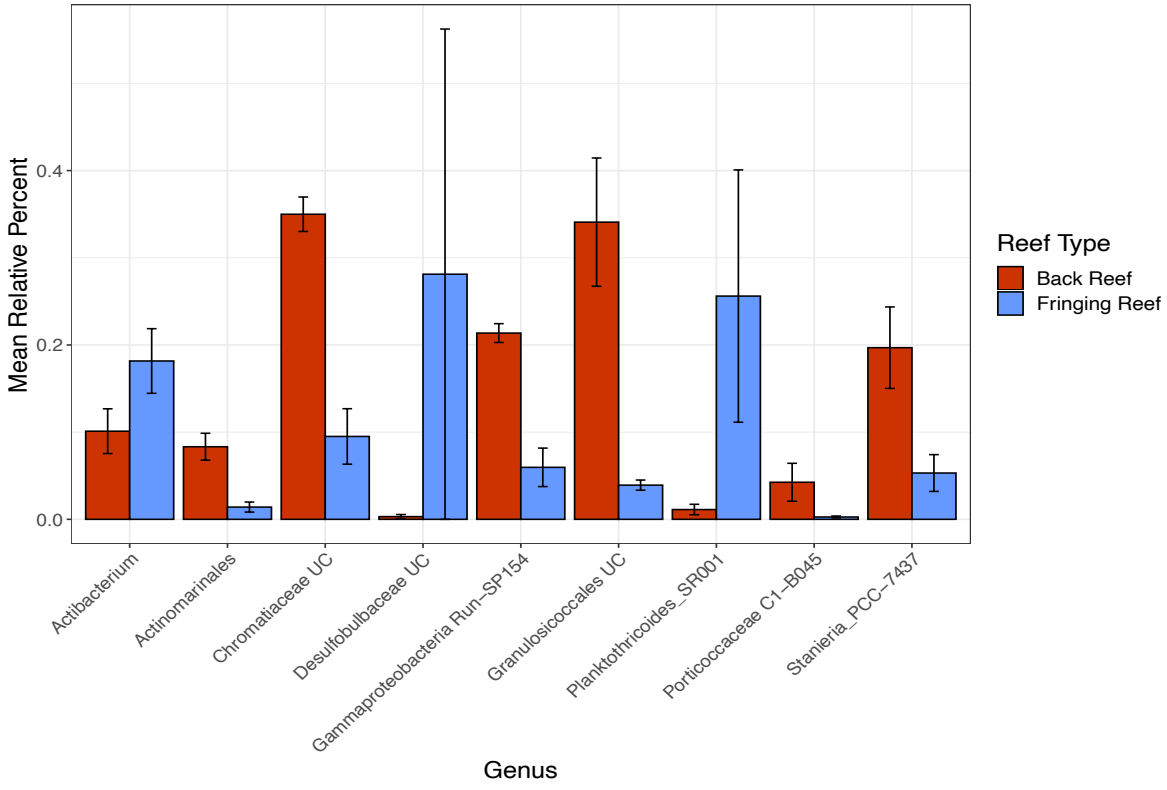

Figure 4.7: Mean relative percent of significant genera determined through ANCOM analysis. Error bars denote standard error. "UC" represents genera identified as "uncultured."

4.4.2 Functional Profiles of Microbial Communities

While microbial communities separated based on site and reef type, functional profiling of the communities based on metagenomics revealed few differences in metabolism (Figure 4.8). Some samples that stood out as slightly different in their metabolism were a site 6 sample (MO18_180) which had higher nitrogen assimilation and oxidative phosphorylation. Two site 6 samples (MO18_180 and MO18_170) also appeared to have increased nitrogen mineralization-assimilation, Krebs cycling, respiration and assimilatory sulfate reduction. In contrast, the site 6 samples had reduced nitrogen fixation and nitrogen and sulfur assimilation (Figure 4.8). Interestingly, different sequencing runs of the same samples (denoted by the suffix letter) revealed inconsistencies. For instance, while one site 6 sample (MO18_180B) showed reduction in methanogenesis and reverse methanogenesis, as well as dissimilatory sulfate-nitrate reduction, the other two technical replicates from that same sample (MO18_180A & C) were more consistent with the other sites (Figure 4.8). This suggests to me, that in complex heterogeneous samples such as sediments, there can be considerable variation in sequencing results if the samples are not sequenced to saturation. One other metabolic pathway that showed slight variation among samples was aerobic anoxygenic phototrophy, which appeared in greater abundance in samples from site 3 (MO18_071) and samples from site 6 (MO18_188). In the case of site 6, the samples were from different transects, thus highlighting the patchiness of function within sediment communities.

There were also a few differences in the relative amount of glycoside hydrolase enzymes with GH9, GH13 and GH29 seen in higher abundances for some samples. GH9, which are mainly cellulases, were higher in site 3 (MO18_071A-C) and site 6 (MO18_167A&B, MO18_184C, MO18_188A-C, MO18_192A). Site 3 (MO18_071A-C) also had increased amounts of GH13 which act on carbohydrates with alpha-glucoside linkages. Site 6 (MO18_180 B&C) results also

indicated increased GH29, which are exo-acting alpha-fucosidase enzymes (Figure 4.8). Thus, while the functional profiles of the sediment communities largely overlapped, there were slight differences that could aid in explanation of community differences. Site 6 which stood out the most in terms of functional differences, as well as in beta diversity, was also the most distinct in terms of the benthos, having noticeably more algae and a more turbid water column, as well as being near a riverine input.

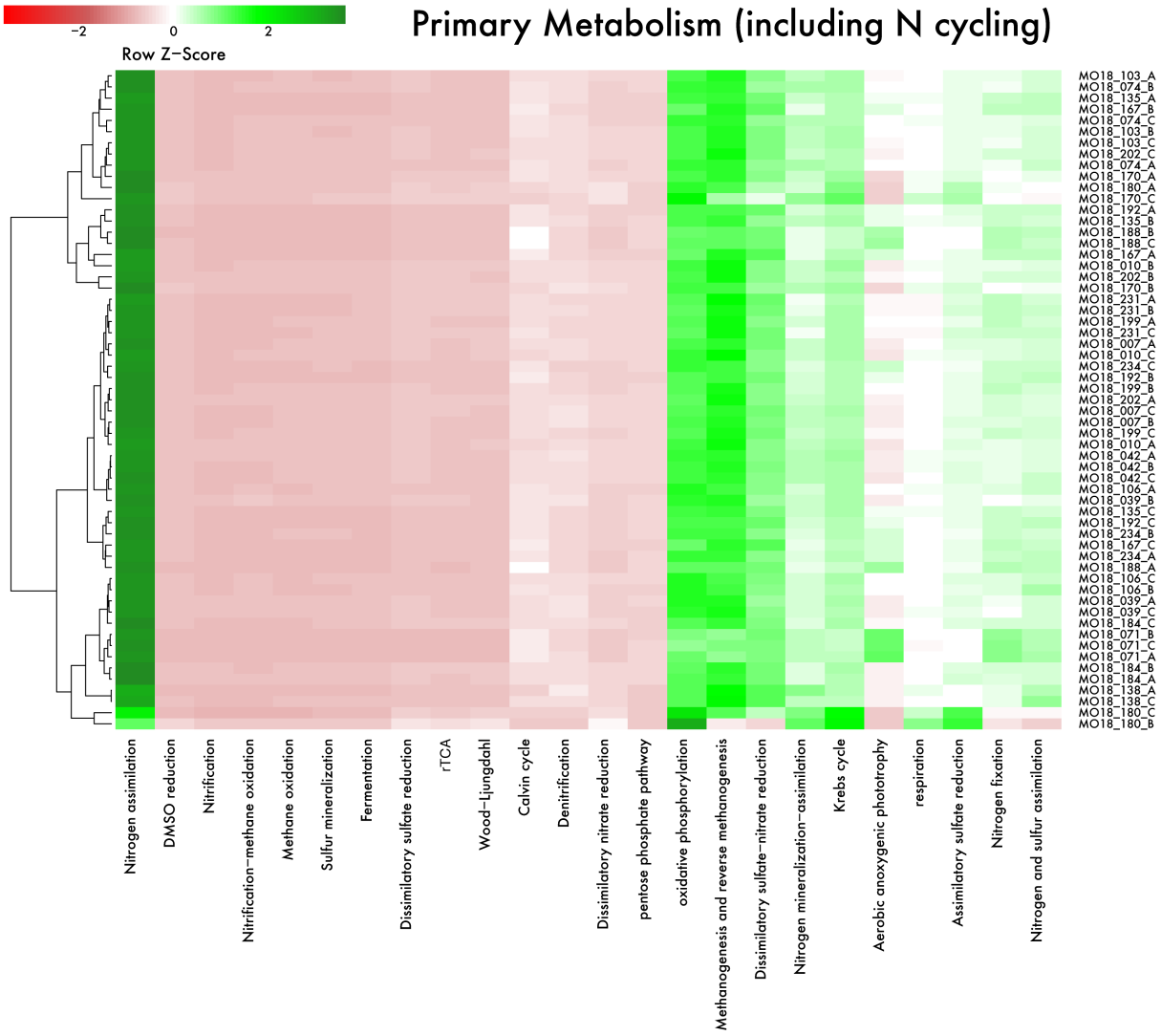

Figure 4.8: Metabolic profiles of sediment communities in Mo'orea, French Polynesia based on metagenomic sequencing. Each row represents a specific sample, see Table 4.1 for site information. Suffix letters denote the technical replicate from sequencing.

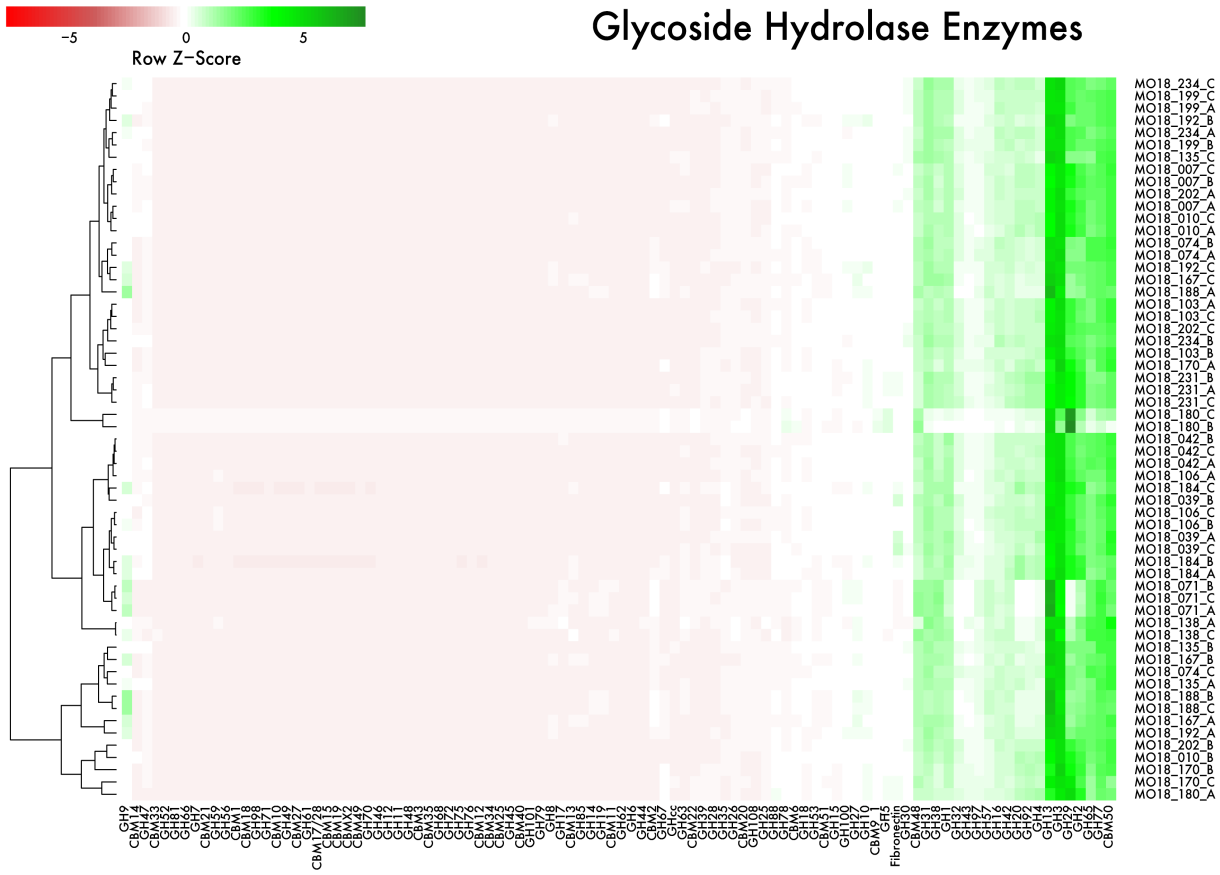

Figure 4.9: Metabolic profiles of sediment communities in Mo'orea, French Polynesia, Continued.

4.4.3 Sediment Metabolomes of Mo'orea

In an effort to have a fully paired dataset, I aimed to deploy 256 resin bags within the sediments of Mo'orea. Unfortunately, between the field and the extraction process, 15 samples were lost. Of the 241 sediment samples analyzed, 2,370 MS1 features were detected after subtracting any features associated with the control samples. The number of observed metabolites in the sediment, or metabolomic richness, was significantly different by site (Figure 4.9a; Kruskal-Wallis $H=56.048$, $p<0.001$) with the average number of metabolites per site ranging from 128.06 ± 6.67 at site 1 to 297.77 ± 31.9 at site 8. The number of metabolites also significantly differed

when comparing reef types, with the back reef having significantly less metabolites than the fringing reef (Figure 4.9b; Wilcoxon test W=56.048, p<0.001).

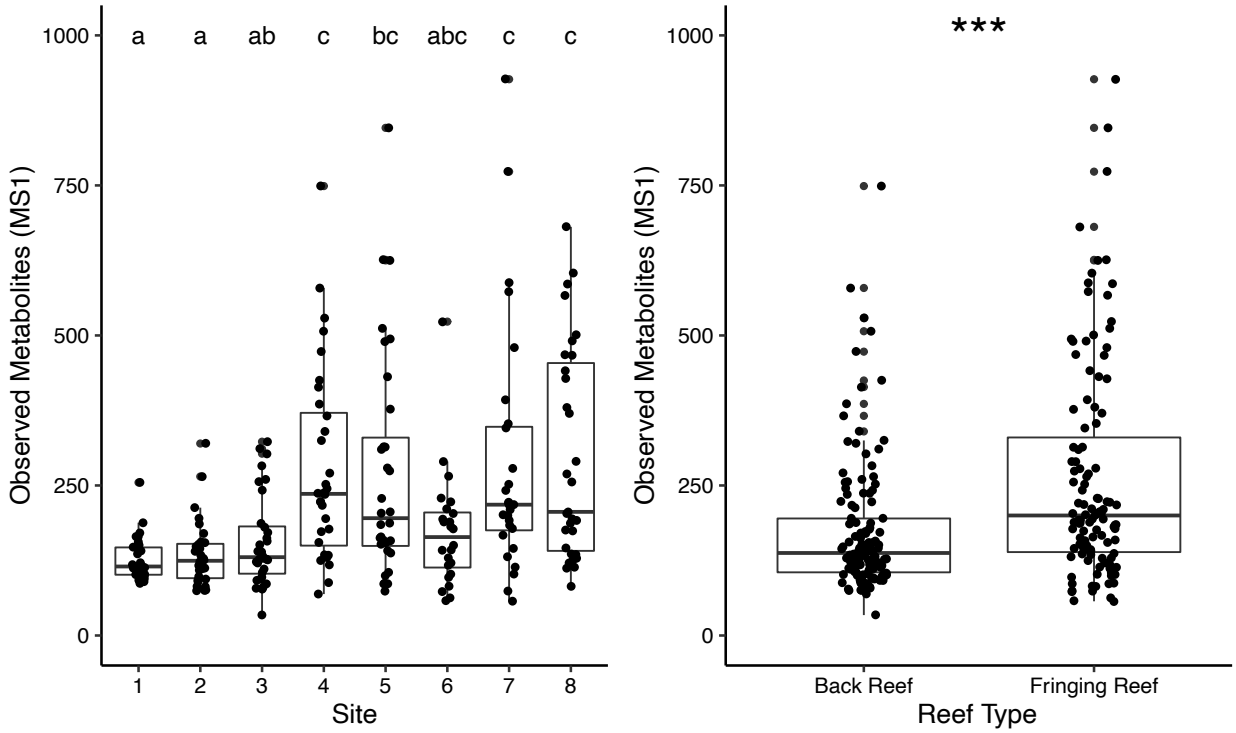

Figure 4.10: Metabolomic richness in sediments from a) eight sites in Mo'orea and by b) reef type. Letters and asterisks denote significance (alpha=0.05).

The metabolomic profile of sediments based on reef type was also found to be significantly different based on principle components analysis with MS1 data (Figure 4.10; Adonis PERMANOVA F=12.883, p=0.001). PCA results indicate greater variation within the metabolomic profiles of sediments from fringing reefs compared to those from back reefs (Figure 4.10). This result aligns with the MetaboAnalyst analysis which identified 442 metabolites significantly higher in fringing reefs while only 145 metabolites were found in greater abundance in back reefs. MS1 profiles were also analyzed in relation to the microbial communities detected and the environmental characteristics. Metabolomic profiles were correlated with distance from

shore and the percent *Acropora* cover, but not correlated with depth, *Pocillopora* cover, nutrient estimates or microbial communities (Table 4.4).

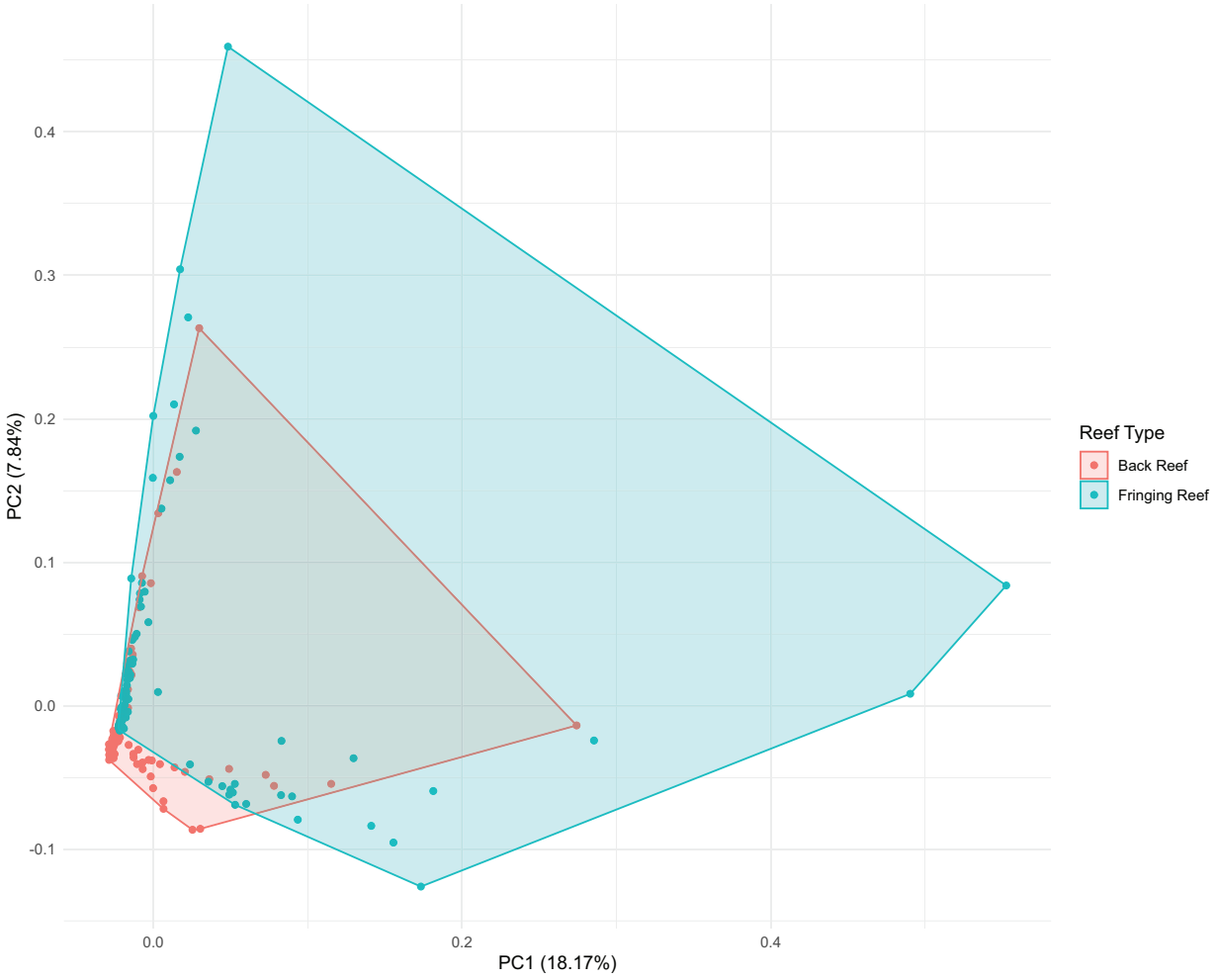

Figure 4.11: Metabolomic beta diversity in sediments from back and fringing reefs in Mo'orea based on MS1 data. Reef types were significantly different from each other (Adonis F=12.883, p=0.001).

Table 4.4: Correlations between metabolomic beta diversity, microbial beta diversity and environmental traits. Mantel tests with Spearman's correlation were used to assess significance. (*) = p<0.05, (**) = p<0.005, (***) = p<0.001.

Beta Diversity Metric	Trait	Spearman's rho	P-value
Metabolites	Weighted UniFrac	-0.045	0.199
	Unweighted UniFrac	-0.028	0.434
	Depth	-0.0062	0.854
	Distance to shore	0.1635	0.001 (**)
	% *Acropora*	0.0637	0.019 (*)
	% *Pocillopora*	0.0263	0.206
	N:C estimates	0.0363	0.289

I also ran the metabolomics dataset through GNPS in an attempt to identify compounds based on MS2 fragmentation patterns. Classical molecular networking resulted in a network with 1,844 nodes after the removal of compounds associated with the control samples. Of those nodes, only 12 had putative matches with the GNPS compound library, thus only 0.6% of nodes were annotated (Appendix B). Unfortunately, the few putative compound hits were mainly fatty acids, such as pinolenic acid, and potential contaminates, such as hexanedioic acid (precursor to nylon). Interestingly, when the MS2 profiles were analyzed via PCoA in GNPS, the pattern in variation is opposite to what was observed with MS1 profiles. While samples generally cluster by reef type, the back reef appears more variable in the metabolomic profile compared to the fringing reef (Figure 4.11a). By visualizing the samples based on site, I was able to determine that site 4 is the back reef that seems to be clustering with the fringing reef samples in MS2 analysis (Figure 4.11b).

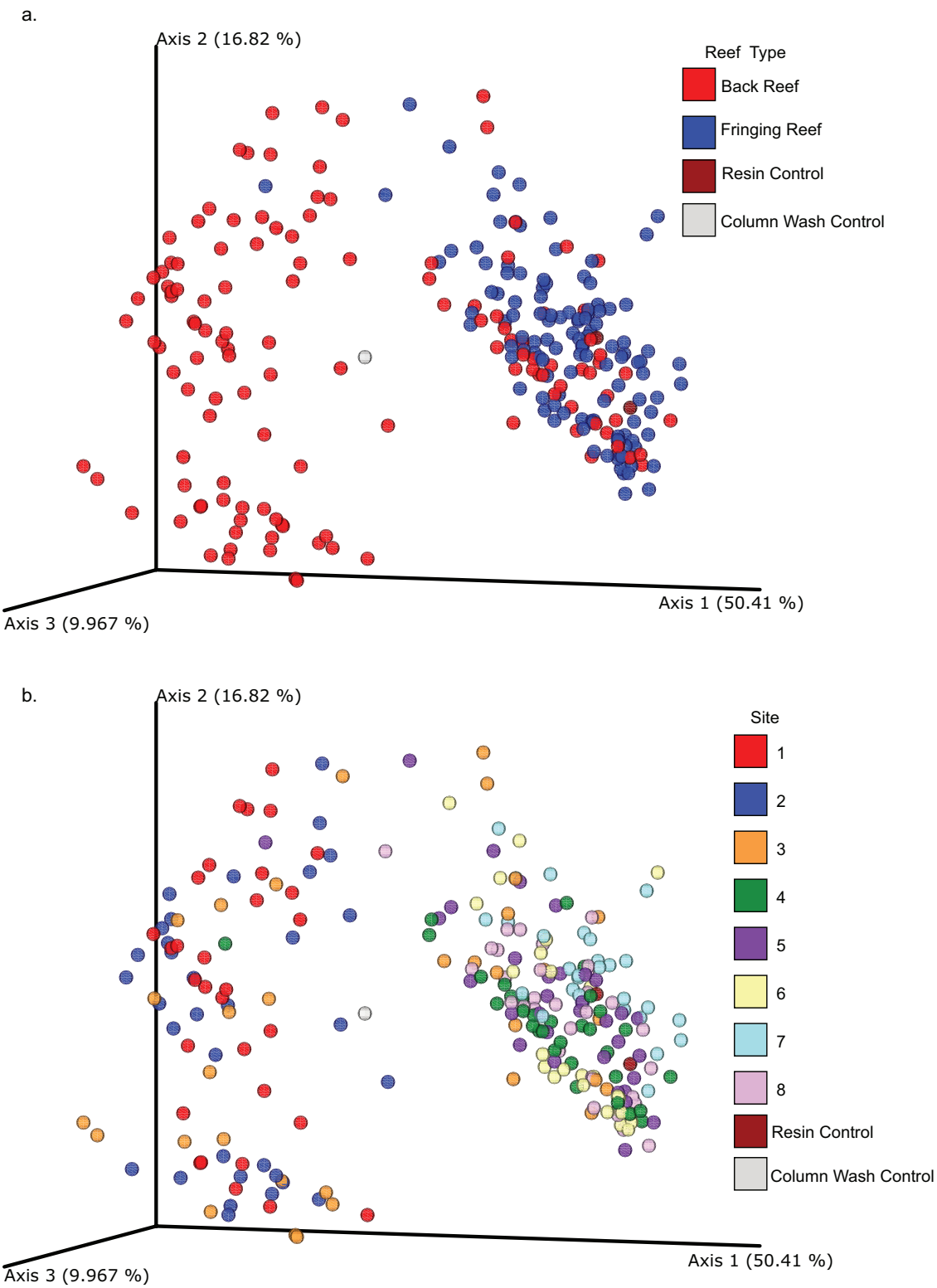

Figure 4.12: Metabolomic beta diversity in sediments from Mo'orea based on MS2 data. Samples are colored based on a) reef types and b) site. Sites 1 – 4 represented the back reef while sites 5 – 8 represented the fringing reef.

4.4.4 Correlations between microbes and metabolites

Finally, in an attempt to draw linkages between the datasets, DIABLO analysis was done using the MixOmics package in R with reef type as the distinguishing factor. Microbial genera were most correlated with environmental traits (r = 0.74) and metabolites (r = 0.65), while metabolites and traits were not well correlated (r = 0.38) (Appendix C). This supervised result follows the pattern seen with the Mantel tests which also saw better correlations between the microbial communities and environmental characteristics. Network analysis identified eight genera that were both positively correlated with nutrient concentrations and negatively correlated with depth (Figure 4.12a). An additional genus, *Desulfuromonadia* C8S 102, was negatively correlated with depth but had no correlation with nutrient concentration. Three metabolites were found to be positively correlated with nutrient concentration while negatively correlated with depth and another three metabolites were only correlated with depth (Figure 4.12a). To better visualize the strongest connections between the >100 metabolites that were linked with the percent *Acropora* cover, I raised the base cutoff from 0.7 to 0.8, which reduced the connections to 21 metabolites (Figure 4.12b). All of the metabolites linked with percent *Acropora* cover showed negative correlations, indicative of reduction or loss of compounds with the loss of this coral genus. Surprisingly, no microbe-metabolite correlations were found with >0.7 correlation.

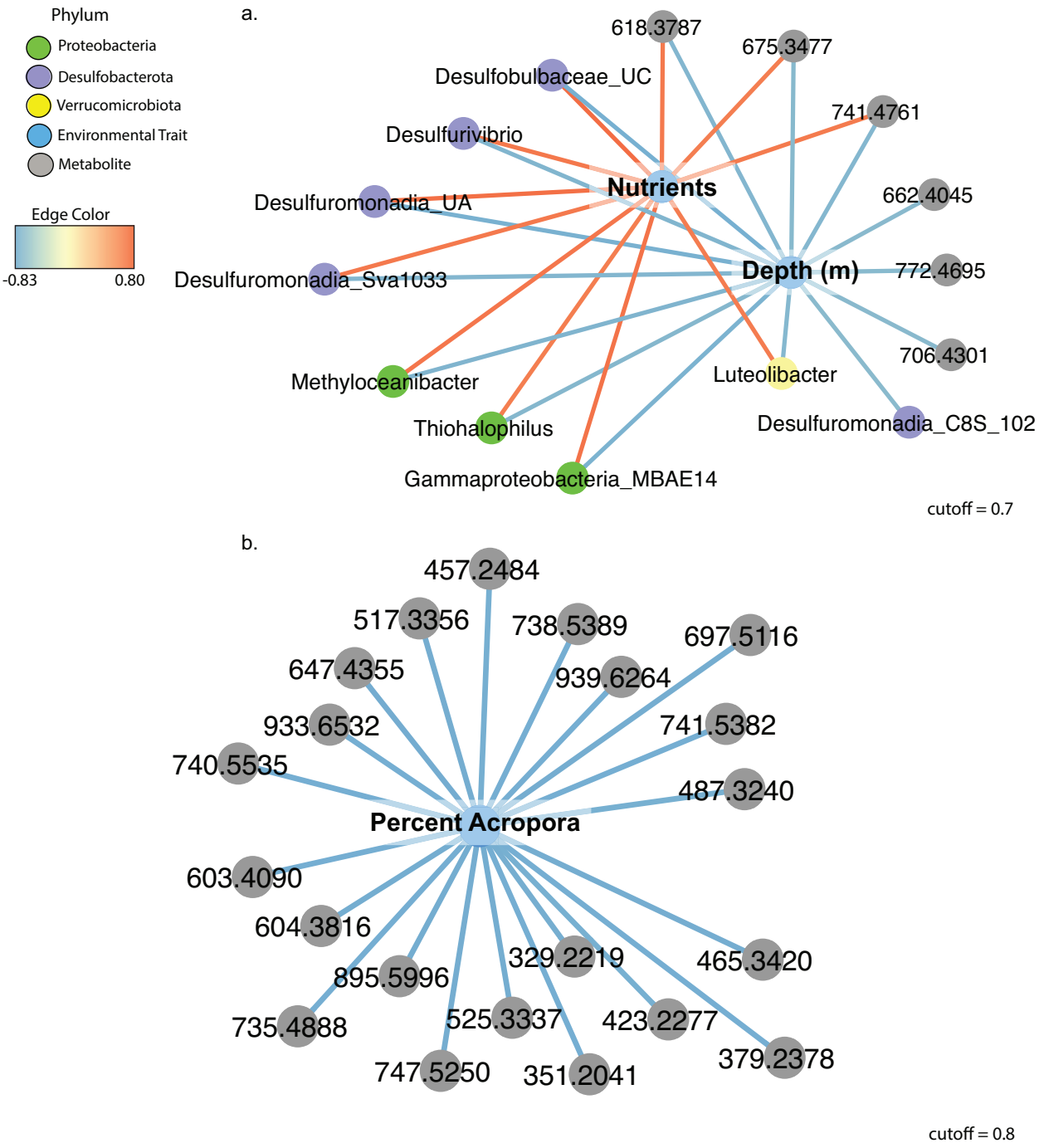

Figure 4.13: DIABLO correlation network of microbial genera, metabolites and environmental traits. a) Network connections associated with nutrients (N:C) and depth and b) network associated with percent *Acropora* cover. "UC" = uncultured while "UA" = unannotated genus. Edges are colored based on correlation value.

In an effort to assess microbe-metabolite linkages, I ran a series of networks with a cutoff of 0.6 (Figure 4.13). Depending on which factor was used to supervise the analysis, the resulting networks varied. For instance, when only pairing samples based on their sample identification, seven genera were identified and linked with over 100 metabolites (Figure 4.13a). Based on site, the number of genera identified is reduced to four with 67 corresponding metabolites (Figure 4.13b). Finally, looking at reef type patterns, only an uncultured genus of Desulfobulbaceae was identified and correlated with two metabolites (Figure 4.13c). The reduction in significant network connections as the factor becomes broader (sample→site→reef type) highlights that despite a large sample size, the ability to identify connections is limited by the strength of the overall correlation and how well the supervising factor explains community variance. In other words, the Mantel test did not find a correlation between microbes and metabolites and when supervised analyses were employed, the correlation was still fairly weak. As the supervising factor becomes broader, and thus encompasses more samples, the sPLS-DA is weaker and fewer connections can be detected. Since the uncultured Desulfobulbaceae genus is consistently identified in the networks, including in the network based on reef type, it appears to be the strongest connection and is thus worth further exploring the role that taxon might be playing in the sediments of Mo'orea. The uncultured Desulfobulbaceae genus was also identified in ANCOM analysis as being in higher abundance in fringing reefs (Figure 4.7). Interestingly, when assessing this taxon based on site, the uncultured Desulfobulbaceae genus was identified as a rare member of the site 1 (0.010%) and site 4 communities (0.002%) but was 1.125% of the site 6 community. Therefore, it seems that in this case it is not the presence or absence of a taxon that is driving the pattern, but its relative abundance across sites.

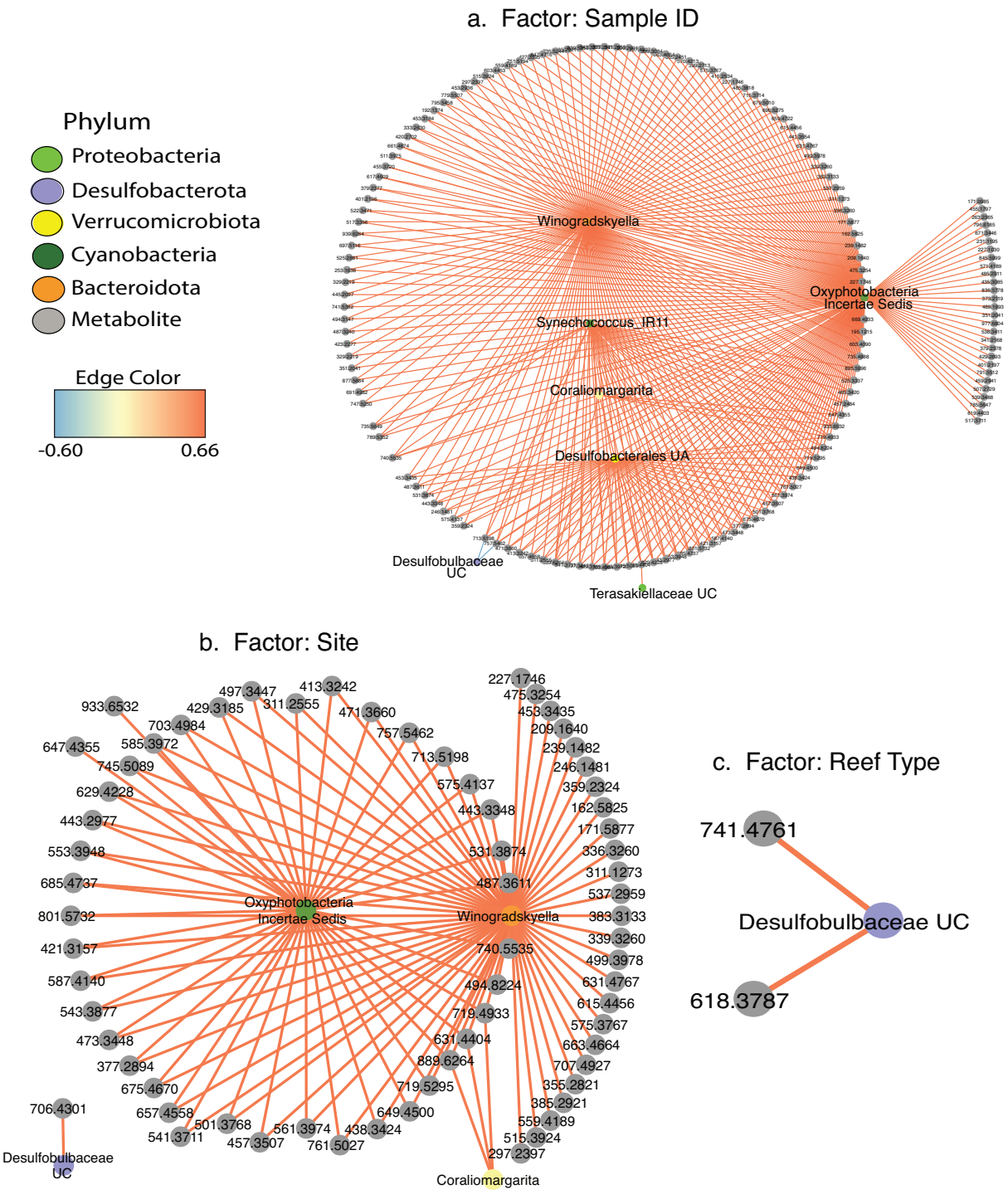

Figure 4.14: DIABLO correlation network generated from sPLS-DA analysis of microbial genera and metabolites when factored by a) sample ID, b) site and c) reef type. "UC" = uncultured while "UA" = unannotated genus. Edges are colored based on correlation value.

4.5 Discussion

In this study, I sampled fringing and back reefs in Mo'orea as a proxy for algal- and coral-dominated environments. I hypothesized that microbial communities associated with coral-dominated environments would exhibit greater bacterial diversity when compared to sediment communities from algal-dominated sites. My results indicate that the back reef was richer and more diverse than the fringing reef, supporting my hypothesis. While the differences observed were significant, they were admittedly not as striking as I thought they would be, and that is likely due to the minimal differences between the fringing reef and back reef in terms of macroecology. Due to logistical constraints, the sites sampled were all within 1 km^2 of the LTER, specifically in and around LTER 2 (Figure 4.1). Island-wide surveys indicate a much greater range of algal and coral cover around the island. Thus, it would be beneficial to expand the sampling and compare sediment microbial community richness and diversity across the other LTER sections as well.

Beta diversity of the microbial communities was also significantly different by site and reef type. Interestingly, of the parameters tested, none seemed to explain the variance observed in the first axis of the weighted UniFrac. It would be interesting to gather more metadata associated with these communities (e.g., oxygen levels, organic content, etc.) to see what might help explain the observed variation in these communities. While weighted analyses indicated significant differences in reef type and based on site, the clustering pattern observed in the PCoAs contained overlap. This is indicative of overlap in dominant taxa such as Proteobacteria and Cyanobacteria. In contrast, the clustering pattern observed in unweighted UniFrac was more distinct between the back and fringing reefs. When visualizing the unweighted PCoA by site, the primary difference seemed to be driven by site 6. Of all the sites sampled, it was the most distinct in terms of the observed macroecology. Site 6 was next to a riverine input, it was dominated by macroalgae, the

sediment consistency was less calcareous than the other sites and the water column was noticeably turbid. Thus, it was not surprising that site 6 seemed to have such a distinct microbial community. Since the separation of site 6 from the other sites is more distinct in the unweighted PCoA, this suggests there are likely more unique community members present at this site and/or some of the taxa consistent across the other sites may be absent at site 6.

While I hoped to incorporate more metadata from the Mo'orea LTER to understand community variation, it was difficult to obtain at the small spatial resolution of our samples. The environmental data I was able to assess did correlate with the microbial communities in many cases. Interestingly, the percent *Acropora* cover seemed to impact the weighted microbial community but not the unweighted composition while *Pocillopora* coral cover was correlated with both beta diversity metrics. *Pocillopora* corals tend to be more resilient to disturbance and thus have increased in their abundance relative to *Acropora* in Mo'orea over the years and were present at every site I sampled (Pratchett *et al.*, 2011). Despite the overall correlation pattern between beta diversity and *Pocillipora* cover, network analysis did not link any genera with *Pocillipora*. It would therefore be interesting to do network analyses at different taxonomic levels, down to ASVs, and see if specific correlations between microbes and *Pocillopora* are detected. In contrast to *Pocillopora*, *Acropora* cover was only associated with sites 2 and 3, and site 2 had the highest coral cover of all sites sampled. Additionally, of the six genera identified via ANCOM in higher abundance in the back reef, three of them were in the greatest relative proportion at site 2. This included an uncultured Granulosicoccales genus, Gammaproteobacteria Run-SP154 and Porticoccaceae C1-B045. The type strain of Granulosicoccales was isolated from Antarctic seawater (Lee *et al.*, 2007) and additional representatives were recently identified in association with *Amphisorus* forams in Australia, but more work is needed to understand the role of this taxon

in marine environments (Macher *et al.*, 2021). The Porticoccaceae family is also associated with the marine environment, with cultures from the Yellow Sea (Oh *et al.*, 2010) and Loch Creran (Gutierrez *et al.*, 2012) and culture-independent detection in Norwegian seas (Ribicic *et al.*, 2018) and mangrove sediments (Araujo *et al.*, 2020). Porticoccaceae are best known for their ability to degrade hydrocarbons, so their presence may suggest some type of hydrocarbon contamination in the environment or, given how variable microbial families can be, their presence in Mo'orea sediments could be indicative of a related clade with unknown function.

An Actinomarinales genus, *Stanieria* and an uncultured Chromatiaceae were also identified in association with the back reef. Actinomarinales are a subclass of marine Actinobacteria that exhibit streamlined genomes and are aerobic photoheterotrophs typically associated with surface waters (López-Pérez *et al.*, 2020). They also possess heliorhodopsins, which may play a role in nitrate or nitrate reduction, but that has not been confirmed (Kovalev *et al.*, 2020; López-Pérez *et al.*, 2020). *Stanieria* are cyanobacteria that typically grow in shallow waters attached to calcareous rocks, shells or on the surfaces of algae (Wu *et al.*, 2016), while the Chromatiaceae family are phototrophic purple sulfur bacteria involved in carbon, sulfur and nitrogen cycling (Imhoff, 2014). In contrast, *Actibacterium*, *Planktothricoides* (SR001) and an uncultured Desulfobulbaceae were seen in association with the fringing reefs. Members of Rhodobacteraceae, including *Actibacterium*, are commonly associated with marine sediments and frequently correlated with redox condition (Pohlner *et al.*, 2019), so it would be interesting in future work to compare the redox state of sites surveyed. The filamentous cyanobacterium *Planktothricoides* was seen in the highest abundance at site 6. This genus has previously been associated with blooms in fresh and brackish waters (Te *et al.*, 2017) and site 6 was subject to fresh riverine inputs. Genome analysis indicates that *Planktothricoides* (SR001) can grow under a

variety of nitrogen sources and has the ability to produce a putative cyanobactin compound (Te *et al.*, 2017). Thus, *Planktothricoides* may be able to opportunistically grow in the fringing reefs where there are additional nutrient and freshwater inputs. While the ecological role of the putatively identified cyanobactin from *Planktothricoides* SR001 has not been identified (Te *et al.*, 2017), related compounds possess anti-microbial properties which could indicate an additional competitive edge for this cyanobacterium under the right conditions. Cyanobactin was not identified in the GNPS analysis, but only one cyanobactin compound is present in the GNPS library. Therefore, it would be good to do a more targeted search for cyanobactin-like compounds in future work with these samples.

In relation to this study, the uncultured Desulfobulbaceae genus identified from ANCOM is one of the most interesting taxa distinguishing fringing and back reefs because it was also the one consistent genus linked in the network analyses. While it was identified with ANCOM in association with fringing reefs, of the fringing reefs surveyed it was actually only present at site 6. However, this one uncultured genus was found in >1% of the community at site 6, a noticeably high percent of the community. Network analyses also found a significant correlation between this genus and nutrient levels and depth; site 6 had the highest nutrient levels and was the shallowest of all sites surveyed. Desulfobulbaceae, commonly called cable bacteria, are known for their ability to generate long filaments that are used to couple the reduction of nitrate or oxygen at the sediment surface with the oxidation of sulfur in deeper anoxic layers of sediment, thus making them an important contributor to sulfur cycling (Geelhoed *et al.*, 2020). Considering the spatial scheme employed, it would be interesting to assess how this genus of cable bacteria was distributed among site 6 and subsequently, whether the population of Desulfobulbaceae at this site is clonal.

Site 6 also stood out in terms of metabolic profiles, with increased nitrogen mineralization-assimilation possibly related to the increased nutrient inputs at that site, and assimilatory sulfate reduction coinciding with sulfur cycling bacteria such as Desulfobulbaceae. Krebs cycling was also seen in higher abundance at site 6 and has previously been found to be associated with algal-dominated sites (Haas *et al.*, 2016). Aerobic anoxygenic phototrophy was seen in the highest abundance at site 3, which corresponds with the increased presence of purple sulfur bacteria in back reef sites relative to the fringing reef. It would be helpful in future analyses to do paired networks that incorporate functional traits to determine if specific taxa correspond with the few functional differences detected across samples. Sediments are patchy in their microbial communities though, so it is also possible that the two metagenomic samples from within a small area of the quadrats at each site will have less observed differences than if we had performed metagenomics with more disparate samples. Thus, more work is needed to assess functional differences in the microbial communities associated with fringing and back reefs.

While microbial richness was greater in the back reef, metabolomic richness showed the opposite pattern, with greater richness observed in the fringing reef. Approximately four times as many metabolites were identified as significantly increased in fringing reef samples compared to the back reef, indicating that there is not only greater diversity of metabolites in the fringing reef, but the compounds observed are in higher abundances. This result contradicts my original hypothesis that metabolomic complexity would be reduced in algal-dominated habitats. It is interesting to note though, that site 6, which was the most algal-dominated of all sites surveyed, showed less variation and a lower median in the number of observed metabolites when compared to the other fringing reef sites. Another intriguing pattern was the clustering in metabolomic beta diversity. While there was a significant split between back and fringing reefs in the PCoA (Figure

4.11), site 4 (categorized as a back reef) clustered more with the fringing reefs. Site 4 also exhibited significantly more metabolomic richness than the other back reefs. Site 4 was the deepest of the sites sampled and was also the closest to Cook's Bay, so it is possible that the microbial communities and associated metabolites were divergent based on the reef location relative to the other sites. While metabolomic analyses identified a number of significant metabolites, both through the MetaboAnalyst pipeline and in network analysis, the annotation rate for the compounds detected was <1%. Unfortunately, sediment extracts are difficult to annotate because most compounds are not in available databases. Considering that the mass spectrometry dataset in this study was higher quality than that of Chapter 3, it was disappointing to see so few putative compounds. However, since the instrumentation in this study was more accurate, I used stricter thresholds in my GNPS analysis. Therefore, it is likely that the <1% annotation rate observed in this study is more reflective of annotation rates than the previous chapter. Much of the poor annotation rates comes down to the databases used, in this case the GNPS library. GNPS is a relatively new tool, so it will be interesting to reevaluate these samples as more library spectra are added to the database and as more associated tools are developed. Until then, additional chemical analysis is needed to interrogate the metabolomic spectra, but the high data quality can support the putative identification of natural products from these resin bags. For instance, one node from the back reef had a mass (316.2846) and MS2 fragmentation pattern that corresponded with Halisphingosine A (Bogdanov et al. unpublished), a sphingosine previously identified from sponges (Molinski *et al.*, 2013). While the use of a library standard and NMR would be needed to confirm this match, this result indicates that resin metabolomics can successfully capture natural products present in marine sediments.

Another important consideration when thinking about the sediment metabolome is whether mass spectrometry results would actually be indicative of the chemical dynamics at play in these environments. For instance, the compounds detected are going to change based on the method used and how well those compounds ionize. Interestingly, I ran MetaboAnalyst with my full list of MS1 spectra (presented in the results section), but I also ran it with the MS1 data that had corresponding MS2 fragmentation spectra. While approximately half of the compounds had MS2 data, the results from a random forest analysis indicate that one of the top compounds explaining variance between the back reef and fringing reefs did not generate MS2 spectra. Therefore, it is important to consider the caveats and limitations of methods employed in studies such as this one. Additionally, the abundance of a compound is not necessarily indicative of its role in the community. Bacteria can produce potent antimicrobial compounds; in which case they do not necessarily need to be in high abundance to be effective in their ecological role. Of course, more research is needed to explore the role of chemistry in complex microbial communities, but I think focusing on compounds linked to microbes through community data or metagenomics is a good place to start.

Marine sediments host rich and diverse microbial communities that are woefully understudied relative to other systems (Lloyd *et al.*, 2018; Baker *et al.*, 2021). Just recently, a global survey assessed marine sediments, providing important baseline information on microbial richness and abundance estimates for bacteria and archaea (Hoshino *et al.*, 2020). Results from Hoshino et al. (2020) highlight the role of oxygen in sediment communities and how communities change with sediment depth. However, no study to date has focused on establishing baseline information on sediment microbial communities associated with coral reef ecosystems. The use of microbial indicators for coral reef health and environmental perturbations has been documented and, in some cases, has incorporated sediment sampling. For instance, sediments were sampled in

conjunction with corals, seaweeds, sponges and seawater in a study assessing the diagnostic value of microbes in inferring the environmental state of coral reefs in Australia (Glasl *et al.*, 2019). Seawater was ultimately deemed a better predictor of environmental fluctuations while sediment communities were more influenced by their habitat characteristics (Glasl *et al.*, 2019). Therefore, depending on the goal of indicator taxa, if one wants to catch an acute environmental change like a temperature fluctuation, versus a more chronic stressor, it may be beneficial to sample different microbial communities. Interestingly, one of the indicator taxa identified in seawater associated with total suspended solids (or turbidity) was Verrucomicrobia. In my study, one of the genera positively correlated with nutrient concentrations was *Luteolibacter,* a genus of Verrucomicrobia. Indeed, *Luteolibacter* was in the highest abundance at site 6, where the water column was noticeably turbid.

My original hypothesis for this study was that the more algal-dominated fringing reefs would have a reduction in microbial and chemical diversity. While microbial diversity was reduced, chemical diversity was actually higher in fringing reefs. One potential explanation for this pattern is the Anna Karenina principle, based on the quote: “all happy families are all alike; each unhappy family is unhappy in its own way.” The Anna Karenina principle was recently discussed in relation to patterns observed in dysbiotic coral and other animal microbiomes, where one of the outcomes of a dysbiotic community is increased variability (Zaneveld *et al.*, 2017). Thus, stressors may not reduce microbial diversity, rather they can increase it through the destabilization of the community which results in increased dispersion. In the case of these sediment communities, something similar could be occurring where a healthy complex community is kept in balance through its chemically mediated interactions. However, when the community shifts and becomes less stable, a greater variety of chemistry may be produced as it works towards

its new community equilibrium. Additionally, while I broadly grouped my sites based on whether they were back or fringing reefs, the site patterns observed highlight that the reefs are not consistent in their communities, thus sites are "unhappy" in their own ways. Furthermore, while I sought to limit confounding factors by focusing on one geographic area, the delineation of site 6 in many of the analyses highlights how important abiotic factors can be to microbial communities. The riverine input likely played a large role in why site 6 was so distinct; it would be interesting in future research to further compare reefs based on their proximity to freshwater inputs.

In conclusion, the microbial communities and metabolomes of sediment communities in the fringing and back reefs of Mo'orea were found to be significantly different from each other. Further, site differences were observed in communities and metabolomes despite all sites being within a 1 km^2 area. This result highlights the complexity of marine sediment communities. While I was able to extract significant taxa and metabolites driving differences in these communities, my conclusions are limited based on the available databases and associated information. Many of the microbial taxa could not be identified to the genus level and the lack of knowledge on even dominant marine lineages adds to the difficulty in extrapolating the role these microbes may fulfil in their communities (Baker *et al.*, 2021). Likewise, high-throughput environmental metabolomics and the use of public MS/MS spectral libraries is in its infancy. My results highlight the large metabolomic richness present in marine environments and the potential for novel chemistry. Finally, I was able to make linkages between microbes, metabolites and environmental parameters associated with the environmental state of the reef. Microbial communities showed significant correlations with the environmental parameters tested while much of the variance in metabolomics remains unexplained. The resin bags proved capable of capturing natural products, but more work is needed to establish a high-throughput method for compound identification. In the future, I would

like to survey the other LTER sites around Mo'orea while simultaneously collecting data on coral and algal cover to ensure a clearer gradient in community comparisons. It would also be valuable to do manipulative experiments to assess how sediment communities in Mo'orea are impacted by environmental factors and metabolites. Ultimately, this study provides baseline information on sediment communities in Mo'orea and how they vary. Hopefully future research in Mo'orea will also consider the role these diverse communities may be playing in this important ecosystem.

4.6 Acknowledgments

Thanks to the members of the Jensen Laboratory who assisted in field preparations including Krystle Chavarría, Kaitlin Creamer, Dulce Guillén Matus, Douglas Sweeney, Fisher Price and Victoria Vasilat. Additional thanks to Rebecca Vega Thurber, Andrew Thurber, Deron Burkepile, Tom Adams, Neil Davies, Hinano Teavai-Murphy, Keith Seydel and Tony You Sing for helpful insight and assistance in the field. I would also like to thank Reiko Cullum, Min Kim, Bill Fenical, Christopher Leber and the Gerwick lab for use of their equipment during chemical extractions.

This work was possible through a Marine Science Research Award from the UCSD Academic Senate, a UCSD Center for Microbiome (CMI) Seed Grant, and a National Science Foundation Graduate Research Fellowship under Grant No. DGE-1650112 to A.M.D. Thanks to members of CMI including Austin Swafford, Julia Gauglitz, Kelly Weldon, Gregory Humphrey Daniel Freed, Sandrine Miller-Montgomery, Pieter Dorrestein and Rob Knight for their support and assistance with sequencing and metabolomics.

Chapter 4 is coauthored with Leesa Klau, Alexander Bogdanov, Alexander Chase and Paul R. Jensen. The dissertation author was the primary investigator and author of this chapter.

Chapter 5

Final Remarks

The last two centuries have brought about fundamental shifts in our understanding of microbial communities. From the initial culturing work of the 1800s to culture-independent methods today, we now know that bacteria are everywhere and encompass an unprecedented amount of diversity relative to the rest of life on earth (Hug *et al.*, 2016; Thompson *et al.*, 2017). Marine sediments in particular, host some of the most rich and diverse microbial communities, with more microbial cells estimated to be contained in sediments than in any other environment or system (Thompson *et al.*, 2017; Lloyd *et al.*, 2018; Hoshino *et al.*, 2020; Baker *et al.*, 2021). Sediment microbial communities play a fundamental role in the marine food web, are integral to global nutrient cycles including carbon, nitrogen and iron and changes to these communities can have important ramifications for ecosystem functioning (Snelgrove *et al.*, 1997; Madigan *et al.*, 2003; Gribben *et al.*, 2017; Rios-Del Toro *et al.*, 2018; Baker *et al.*, 2021). However, much remains unknown about these complex communities. It is thought that sediment microbes engage in complex interactions, much of which is likely mediated by small molecule production. For instance, the secretion of iron-scavenging siderophores can provide a growth advantage (Butler, 2005; D'Onofrio *et al.*, 2010; Roberts *et al.*, 2012; Kummerli *et al.*, 2014) while the production of antimicrobial compounds can act as chemical warfare (Riley and Gordon, 1999; Griffin *et al.*, 2004; Burkepile *et al.*, 2006; Ghoul and Mitri, 2016). As a result, marine sediments have been a rich source for natural product discovery (Fenical and Jensen, 2006; Dalisay *et al.*, 2013; Jensen *et al.*, 2015; Bech *et al.*, 2020). Culturing of marine microbes has provided important insight into these communities. Research into the marine obligate actinomycete genus *Salinispora* for instance,

has identified evolutionary trade-offs in competitive strategy (Patin *et al.*, 2015). Additionally, extracts from *Salinispora arenicola* have been shown to alter sediment communities, indicative of the important role small molecules play in community dynamics (Patin *et al.*, 2017). While case-studies on specific microbes continues to shed light onto these understudied systems, much remains unknown about many marine sediment lineages and the communities as a whole (Baker *et al.*, 2021).

The goal of this dissertation was to use next-generation sequencing and newly developed omic techniques to gain ecological insight into marine sediment communities. I hoped to not only establish baseline information about the microbial diversity present in sediments, but aimed to unlock connections between microbes, the metabolites they produce and the surrounding environmental characteristics. I was in the fortunate position of being in a laboratory that works at the interface of microbial ecology and natural product chemistry. This facilitated my ability to ask questions about microbial diversity detected using culture versus culture-independent methods (Chapter 2) and to develop methods for sediment metabolomics using the small molecule *in situ* resin capture (SMIRC) technique. Rather than directly extracting sediments for metabolomics, the use of adsorbent resins allows for a targeted approach that should capture molecules actively being secreted in the sediment. Thus, I was able to use SMIRC and incorporate metabolomics as a variable when investigating microbial communities across varying spatial scales (Chapter 3). This method was further improved upon for a large-scale experiment that explicitly paired 256 microbial community samples with sediment metabolites in an effort to assess how microbial communities compare in relation to the state of the surrounding benthic environment (Chapter 4).

The findings of Chapter 2 show that with just two basic media, >1% of the sediment microbial community can be identified in culture. Culture-dependent methods also detected

microbial taxa that were absent in culture-independent analysis, including taxa that are divergent from known cultured strains. The detection of taxa with culturing methods only, highlights two important factors in studying marine sediments, 1- marine sediment communities are inherently very patchy, so no two samples are exactly alike and 2- culture-independent methods have limitations and biases. Another important takeaway from Chapter 2 was the sequencing depth required to reach saturation and the associated implications of assessing diversity when missing part of the community. Studies using 16S rRNA in marine sediments frequently do not sequence to saturation, thus missing rarer taxa (Santelli *et al.*, 2008; Thompson *et al.*, 2017; Hoshino *et al.*, 2020). While there are tradeoffs between the number of samples and sequencing depth, a lack of sequencing depth results in underestimates of diversity. This can in turn, have important effects on our understanding of complex microbial communities. In comparing sites from Belize, unweighted beta diversity analysis which accounts for absence/presence of taxa rather than proportion, showed clearer community clustering patterns when compared to weighted analysis. Dominant community members may overlap between sites, so it is likely the rare microbes drive site differentiation. Consequently, without fully sequencing the community, the impact of rare microbes can be missed.

For Chapter 3, I wanted to address how community diversity changes with spatial scale. I also wanted to incorporate the role that sedimentary characteristics play in distinguishing communities. Additionally, Chapter 3 utilized SMIRC methods to take a snapshot of the sediment metabolome present at five different sites around Nacula Island, Fiji, allowing for the incorporation of metabolomics into this study. While I did not do surveys on the macroecology of sites around Fiji, the majority of them were representative of healthy coral reefs. It was therefore surprising that the microbial communities associated with the five sites were so divergent in their beta

diversity. When comparing microbial diversity across sites to the diversity detected within a 1 m^2 and across 30 m of transect samples from one site, microbial diversity significantly increased with sampling scale. This result once again highlighted the inherent patchiness of marine sediments and how sampling, even within one site, can be highly variable. Of note, ASV richness across 12 km^2 in Fiji rivaled the richness detected in a recent global survey of marine sediments (Hoshino *et al.*, 2020). This result further emphasizes the impact of sequencing depth on community richness estimates. Of the sediment traits evaluated, nitrate, iron and sediment grain size were found to be correlated with microbial communities. Results from sediment metabolomics were promising in that we could putatively identify some natural products, however the lack of explicitly paired samples and the noise present in the metabolomics dataset limited the insights gained. Future work is needed to isolate and identify compounds present from these sites.

The goal of Chapter 4 was to expand on the idea of sediment microbial communities in thinking about how these communities relate to the surrounding state of the ecosystem. Terrestrial work has revealed the unique properties of suppressive soils (Van Bruggen and Semenov, 2000; Schlatter *et al.*, 2017), but that phenomenon has not been tested in marine environments. I was curious if changes to the microbial community associated with increased nutrient input from algae would result in a loss of diversity and associated metabolites, which could then have impacts on the surrounding ecosystem. To test this, I compared fringing and back reef sediment communities and metabolomes. The fringing reef is subjected to greater nutrient input than the back reef and therefore is frequently representative of a more algal-dominated environment compared to higher coral cover in the back reef. In an effort to gain better resolution in community and metabolomic patterns, eight sites were surveyed in the same spatial scheme with a 1 m^2 quadrat and four 4 m transects. For every sediment sample that was taken, a corresponding metabolomics sample was

gathered using an improved SMIRC method with smaller resin bags. This experimental design provided additional power and spatial resolution needed to assess paired-omic interactions. Interestingly, while I determined that the fringing reef was less diverse in microbes when compared to the back reef, the opposite pattern was observed in metabolite richness. I think this result provides an important perspective to think about moving forward with environmental metabolomics. While I was initially thinking about explicit correlations between microbes and metabolites, there are a variety of microbes that can produce the same or similar metabolites. For instance, staurosporine has been detected in sediments but has multiple potential producers (Tuttle *et al.*, 2019). Thus, it can be difficult to identify correlative patterns in metabolomics when different producers can alternatively fluctuate in the community and when we cannot confidently identify compounds and/or producers. I was able to use supervised multi-omics analyses to extract direct connections between microbes, metabolites and environmental factors. Considering that the area we sampled was not as representative of algal versus coral-dominated as I had hoped, the identification of differences in these communities provides a meaningful place to start. In particular, the identification of positive correlations between Desulfobacteria and Proteobacteria with nutrients suggests the important role of these taxa in reefs facing an influx of nutrients from shore and/or from algae. Over twenty metabolites were negatively correlated with percent *Acropora* cover. While this may partially be a function of the few sites with *Acropora*, I think this result also suggests the loss of metabolic diversity with the loss of corals. In future work, I would like to explore a greater gradient in algal to coral-dominated systems and to track sediment communities through time. I think the NSF-funded Long Term Ecological Research Monitoring Network around Mo'orea provides a unique opportunity to assess sediment communities in conjunction with a host of abiotic and biotic factors. While I was limited in the region I could

survey, adding sediment sampling and metabolomics to the monitoring program would provide important baseline information and allow for the continued tracking of how communities change through time and in relation to the benthos. It would also be great to do manipulative studies that involve spiking sediments with nutrients as a proxy for eutrophication and increased algal abundance, to see the impact on microbial community structure.

While improvements in metabolomics from Chapter 2 to Chapter 3 resulted in higher quality metabolomics data, the annotation rate for compounds remained low. I think the low annotation rate speaks to the novel chemical space of marine sediments, but also highlights a limitation to current metabolomic databases which are heavily skewed toward industrial compounds. As databases expand, more compounds will be identified, adding increased insight into the correlations I identified in this study. Additionally, many microbial lineages from marine sediments remain poorly studied (Baker *et al.*, 2021). The use of metagenomics to gain insight into these taxonomic groups will provide additional insight into how community changes affect function. Our metagenomic analysis did not identify dramatic changes between our sites in Mo'orea, but it will be valuable to add metagenomic analyses to more disparate community studies including my study from Fiji, which revealed clearer site delineation. Metagenomics will also facilitate connecting microbes to metabolites through the identification of biosynthetic gene clusters, thus my datasets provide a unique opportunity to ask more questions and re-evaluate the results as databases and multi-omic methods improve.

In conclusion, marine sediments contain rich and complex microbial communities that are worth greater exploration. Our oceans are facing a slew of threats including climate change, ocean acidification, overfishing, eutrophication and disease. Sediments are the foundation of these ecosystems and host a greater diversity of microbes than the biota above and the surrounding water

column (Thompson *et al.*, 2017; Lloyd *et al.*, 2018). The microbialization of coral reefs has already been documented (Haas *et al.*, 2016), but we do not know the role sediment communities are playing in our changing oceans. This dissertation highlights the diversity of sediment microbial communities, the rich chemical landscapes in which they reside, and connections between microbes, metabolites and the environment. These studies provide baseline information that I hope future research can build off of to better understand how connections identified here relate to functional changes in the environment.

References

Achberger, A.M., Christner, B.C., Michaud, A.B., Priscu, J.C., Skidmore, M.L., Vick-Majors, T.J., Adkins, W., Anandakrishnan, S., Barbante, C., Barcheck, G., Beem, L., Behar, A., Beitch, M., Bolsey, R., Branecky, C., Carter, S., Christianson, K., Edwards, R., Fisher, A., Fricker, H., Foley, N., Guthrie, B., Hodson, T., Jacobel, R., Kelley, S., Mankoff, K., McBryan, E., Mikucki, J., Mitchell, A., Powell, R., Purcell, A., Sampson, D., Scherer, R., Sherve, J., Siegfried, M., and Tulaczyk, S. (2016) Microbial community structure of subglacial Lake Whillans, West Antarctica. *Front Microbiol* **7**: 1–13.

Ahmerkamp, S., Marchant, H.K., Peng, C., Probandt, D., Littmann, S., Kuypers, M.M.M., and Holtappels, M. (2020) The effect of sediment grain properties and porewater flow on microbial abundance and respiration in permeable sediments. *Sci Rep* **10**: 1–12.

Amir, A., Daniel, M., Navas-Molina, J., Kopylova, E., Morton, J., Xu, Z.Z., Eric, K., Thompson, L., Hyde, E., Gonzalez, A., and Knight, R. (2017) Deblur rapidly resolves single-nucleotide community sequence patterns. *mSystems* **2**: e00191-16.

Amor, M., Mathon, F.P., Monteil, C.L., Busigny, V., and Lefevre, C.T. (2020) Iron-biomineralizing organelle in magnetotactic bacteria: function, synthesis and preservation in ancient rock samples. *Environ Microbiol* **22**: 3611–3632.

Anderson, M.J. (2001) A new method for non-parametric multivariate analysis of variance. *Austral Ecol* **26**: 32–46.

Apprill, A., Mcnally, S., Parsons, R., and Weber, L. (2015) Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquat Microb Ecol* **75**: 129–137.

Arahal, D.R., Lekunberri, I., González, J.M., Pascual, J., Pujalte, M.J., Pedrós-Alió, C., and Pinhassi, J. (2007) Neptuniibacter caesariensis gen. nov., sp. nov., a novel marine genome-sequenced gammaproteobacterium. *Int J Syst Evol Microbiol* **57**: 1000–1006.

Araujo, F.J., Hissa, D.C., Silva, G.O., Antunes, A.S.L.M., Nogueira, V.L.R., Gonçalves, L.R.B., and Melo, V.M.M. (2020) A novel bacterial carboxylesterase identified in a metagenome derived-clone from Brazilian mangrove sediments. *Mol Biol Rep* **47**: 3919–3928.

Armour, C.R., Nayfach, S., Pollard, K.S., and Sharpton, T.J. (2019) A Metagenomic Meta-analysis Reveals Functional Signatures of Health and Disease in the Human Gut Microbiome. *mSystems* **4**: 1–15.

Arora, P.K. (2013) Staphylococcus lipolyticus sp. nov., a new cold-adapted lipase producing marine species. *Ann Microbiol* **63**: 913–922.

Asker, D., Beppu, T., and Ueda, K. (2007) Mesoflavibacter zeaxanthinifaciens gen. nov., sp. nov., a novel zeaxanthin-producing marine bacterium of the family Flavobacteriaceae. *Syst*

Appl Microbiol **30**: 291–296.

Baker, B.J., Appler, K.E., and Gong, X. (2021) New Microbial Biodiversity in Marine Sediments. *Ann Rev Mar Sci* **13**: 1–15.

Barreto, C.R., Morrissey, E.M., Wykoff, D.D., and Chapman, S.K. (2018) Co-occurring Mangroves and Salt Marshes Differ in Microbial Community Composition. *Wetlands* **38**: 497–508.

Bayer, T., Neave, M.J., Alsheikh-Hussain, A., Aranda, M., Yum, L.K., Mincer, T., Hughen, K., Apprill, A., and Voolstra, C.R. (2013) The microbiome of the red sea coral stylophora pistillata is dominated by tissue-associated endozoicomonas bacteria. *Appl Environ Microbiol* **79**: 4759–4762.

Bech, P.K., Lysdal, K.L., Gram, L., Bentzon-Tilia, M., and Strube, M.L. (2020) Marine Sediments Hold an Untapped Potential for Novel Taxonomic and Bioactive Bacterial Diversity. *mSystems* **5**:.

Ben-Dov, E., Yosef, D.Z. Ben, Pavlov, V., and Kushamro, A. (2009) Corynebacterium maris sp. nov., a marine bacterium isolated from the mucus of the coral Fungia granulosa. *Int J Syst Evol Microbiol* **59**: 2458–2463.

Berdy, B., Spoering, A.L., Ling, L.L., and Epstein, S.S. (2017) In situ cultivation of previously uncultivable microorganisms using the ichip. *Nat Protoc* **12**: 2232–2242.

Biebl, H., Tindall, B.J., Pukall, R., Lünsdorf, H., Allgaier, M., and Wagner-Döbler, I. (2006) Hoeflea phototrophica sp. nov., a novel marine aerobic alphaproteobacterium that forms bacteriochlorophyll a. *Int J Syst Evol Microbiol* **56**: 821–826.

Black, C.A. (1965) Methods of soil analysis, Madison, WI: Am. Soc. Agronomy, Mono. 9.

Bokulich, N.A., Collins, T., Masarweh, C., Allen, G., Heymann, H., Ebeler, S.E., and Mills, D.A. (2016) Fermentation Behavior Suggest Microbial Contribution to Regional. *MBio* **7**: 1–12.

Bollmann, A., Lewis, K., and Epstein, S.S. (2007) Incubation of environmental samples in a diffusion chamber increases the diversity of recovered isolates. *Appl Environ Microbiol* **73**: 6386–6390.

Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C., Ghalith, G.A. Al, Alexander, H., Alm, E.J., Arumugam, M., Bai, Y., Bisanz, J.E., Bittinger, K., Brejnrod, A., Colin, J., Brown, C.T., Callahan, B.J., Mauricio, A., Rodríguez, C., Chase, J., Cope, E., Silva, R. Da, Dorrestein, P.C., Douglas, G.M., Duvallet, C., Edwardson, C.F., Ernst, M., Fouquier, J., Gauglitz, J.M., Gibson, D.L., Gonzalez, A., Huttley, G.A., Janssen, S., Jarmusch, A.K., Kaehler, B.D., Kang, K. Bin, Keefe, C.R., Keim, P., Kelley, S.T., Ley, R., Loftfield, E., Marotz, C., Martin, B., Mcdonald, D., Mciver, L.J., Alexey, V., Metcalf, J.L., Morgan, S.C., Morton, J.T., and Naimey, A.T. (2018) QIIME 2 : Reproducible , interactive , scalable , and extensible microbiome data science. *PeerJ Prepr*.

Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C.C., Al-Ghalith, G.A., Alexander, H., Alm, E.J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J.E., Bittinger, K., Brejnrod, A., Brislawn, C.J., Brown, C.T., Callahan, B.J., Caraballo-Rodríguez, A.M., Chase, J., Cope, E.K., Da Silva, R., Diener, C., Dorrestein, P.C., Douglas, G.M., Durall, D.M., Duvallet, C., Edwardson, C.F., Ernst, M., Estaki, M., Fouquier, J., Gauglitz, J.M., Gibbons, S.M., Gibson, D.L., Gonzalez, A., Gorlick, K., Guo, J., Hillmann, B., Holmes, S., Holste, H., Huttenhower, C., Huttley, G.A., Janssen, S., Jarmusch, A.K., Jiang, L., Kaehler, B.D., Kang, K. Bin, Keefe, C.R., Keim, P., Kelley, S.T., Knights, D., Koester, I., Kosciolek, T., Kreps, J., Langille, M.G.I., Lee, J., Ley, R., Liu, Y.X., Loftfield, E., Lozupone, C., Maher, M., Marotz, C., Martin, B.D., McDonald, D., McIver, L.J., Melnik, A. V., Metcalf, J.L., Morgan, S.C., Morton, J.T., Naimey, A.T., Navas-Molina, J.A., Nothias, L.F., Orchanian, S.B., Pearson, T., Peoples, S.L., Petras, D., Preuss, M.L., Pruesse, E., Rasmussen, L.B., Rivers, A., Robeson, M.S., Rosenthal, P., Segata, N., Shaffer, M., Shiffer, A., Sinha, R., Song, S.J., Spear, J.R., Swafford, A.D., Thompson, L.R., Torres, P.J., Trinh, P., Tripathi, A., Turnbaugh, P.J., Ul-Hasan, S., van der Hooft, J.J.J., Vargas, F., Vázquez-Baeza, Y., Vogtmann, E., von Hippel, M., Walters, W., Wan, Y., Wang, M., Warren, J., Weber, K.C., Williamson, C.H.D., Willis, A.D., Xu, Z.Z., Zaneveld, J.R., Zhang, Y., Zhu, Q., Knight, R., and Caporaso, J.G. (2019) Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* **37**: 852–857.

Bourne, D.G., Garren, M., Work, T.M., Rosenberg, E., Smith, G.W., and Harvell, C.D. (2009) Microbial disease and the coral holobiont. *Trends Microbiol* **17**: 554–562.

Bowman, J.S. and Ducklow, H.W. (2015) Microbial Communities Can Be Described by Metabolic Structure : A General Framework and Application to a Seasonally Variable , Depth-Stratified Microbial Community from the Coastal West Antarctic Peninsula. 1–18.

Brooks, J.P., Edwards, D.J., Harwich, M.D., Rivera, M.C., Fettweis, J.M., Serrano, M.G., Reris, R.A., Sheth, N.U., Huang, B., Girerd, P., Strauss, J.F., Jefferson, K.K., and Buck, G.A. (2015) The truth about metagenomics: Quantifying and counteracting bias in 16S rRNA studies Ecological and evolutionary microbiology. *BMC Microbiol* **15**: 1–14.

Van Bruggen, A.H.C. and Semenov, A.M. (2000) In search of biological indicators for soil health and disease suppression. *Appl Soil Ecol* **15**: 13–24.

Burkepile, D. and Adam, T. (2019) MCR LTER: Coral Reef: Coral bleaching with nitrogen and heat stress: 2016 data in support of Donovan et al. submitted to PNAS. Environmental Data Initiative. Deposited 25 October 2019.

Burkepile, D.E., Parker, J.D., Woodson, C.B., Mills, H.J., Kubanek, J., Sobecky, P.A., and Hay, M.E. (2006) Chemically mediated competition between microbes and animal: Microbes as consumers in food webs. *Ecology* **87**: 2821–2831.

Butler, A. (2005) Marine siderophores and microbial iron mobilization. *BioMetals* **18**: 369–374.

Callahan, B.J., McMurdie, P.J., and Holmes, S.P. (2017) Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME J* **11**: 2639–2643.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., and Holmes, S.P. (2016) DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* **13**: 581–583.

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

Cha, J.Y., Han, S., Hong, H.J., Cho, H., Kim, D., Kwon, Y., Kwon, S.K., Crusemann, M., Bok Lee, Y., Kim, J.F., Giaever, G., Nislow, C., Moore, B.S., Thomashow, L.S., Weller, D.M., and Kwak, Y.S. (2016) Microbial and biochemical basis of a Fusarium wilt-suppressive soil. *ISME J* **10**: 119–129.

Chen, H.C., Wang, S.Y., and Chen, M.J. (2008) Microbiological study of lactic acid bacteria in kefir grains by culture-dependent and culture-independent methods. *Food Microbiol* **25**: 492–501.

Chen, M.L., Becraft, E.D., Pachiadaki, M., Brown, J.M., Jarett, J.K., Gasol, J.M., Ravin, N. V., Moser, D.P., Nunoura, T., Herndl, G.J., Woyke, T., and Stepanauskas, R. (2020) Hiding in Plain Sight: The Globally Distributed Bacterial Candidate Phylum PAUC34f. *Front Microbiol* **11**: 1–12.

Chisholm, S.W., Olson, R.J., Zettler, E.R., Goericke, R., Waterbury, J.B., and Welschmeyer, N.A. (1988) A novel free-living prochlorophyte abundant in the oceanic euphotic zone. *Nature* **334**: 340–343.

Cho, I. and Blaser, M.J. (2012) The human microbiome: At the interface of health and disease. *Nat Rev Genet* **13**: 260–270.

Choi, A., Baek, K., Lee, H., and Cho, J.C. (2015) Mesonia aquimarina sp. nov., a marine bacterium isolated from coastal seawater. *Int J Syst Evol Microbiol* **65**: 135–140.

Chong, J., Wishart, D.S., and Xia, J. (2019) Using MetaboAnalyst 4.0 for Comprehensive and Integrative Metabolomics Data Analysis. *Curr Protoc Bioinforma* **68**: 1–128.

Cordero, O.X., Wildschutte, H., Kirkup, B., Proehl, S., Ngo, L., Hussain, F., Le Roux, F., Mincer, T., Polz, M.F., and Roux, F. Le (2012) Ecological populations of bacteria act as socially cohesive units of antibiotic production and resistance. *Science* **337**: 1228–1231.

Crespo, B.G., Wallhead, P.J., Logares, R., and Pedrós-Alió, C. (2016) Probing the rare biosphere of the north-west mediterranean sea: An experiment with high sequencing effort. *PLoS One* **11**: 1–16.

Cross, K.L., Campbell, J.H., Balachandran, M., Campbell, A.G., Cooper, S.J., Griffen, A., Heaton, M., Joshi, S., Klingeman, D., Leys, E., Yang, Z., Parks, J.M., and Podar, M. (2019) Targeted isolation and cultivation of uncultivated bacteria by reverse genomics. *Nat Biotechnol* **37**: 1314–1321.

D'Onofrio, A., Crawford, J.M., Stewart, E.J., Witt, K., Gavrish, E., Epstein, S., Clardy, J., and Lewis, K. (2010) Siderophores from Neighboring Organisms Promote the Growth of Uncultured Bacteria. *Chem Biol* **17**: 254–264.

Dai, D., Rhoads, W.J., Edwards, M.A., and Pruden, A. (2018) Shotgun Metagenomics Reveals Taxonomic and Functional Shifts in Hot water microbiome due to temperature setting and stagnation. *Front Microbiol* **9**: 1–17.

Daims, H., Lebedeva, E. V., Pjevac, P., Han, P., Herbold, C., Albertsen, M., Jehmlich, N., Palatinszky, M., Vierheilig, J., Bulaev, A., Kirkegaard, R.H., Von Bergen, M., Rattei, T., Bendinger, B., Nielsen, P.H., and Wagner, M. (2015) Complete nitrification by Nitrospira bacteria. *Nature* **528**: 504–509.

Dale, N.G. (1974) Bacteria in intertidal sediments: Factors related to their distribution. *Limnol Oceanogr* **19**: 509–518.

Dalisay, D.S., Williams, D.E., Wang, X.L., Centko, R., Chen, J., and Andersen, R.J. (2013) Marine Sediment-Derived Streptomyces Bacteria from British Columbia, Canada Are a Promising Microbiota Resource for the Discovery of Antimicrobial Natural Products. *PLoS One* **8**: 1–14.

Dastager, S.G., Mawlankar, R., Srinivasan, K., Tang, S.K., Lee, J.C., Ramana, V.V., and Shouche, Y.S. (2014) Fictibacillus enclensis sp. nov., isolated from marine sediment. *Antonie van Leeuwenhoek, Int J Gen Mol Microbiol* **105**: 461–469.

Dean Jr, W.E. (1974) Determination of carbonate and organic matter in calcareous sediments and sedimentary rocks by loss-on- ignition: comparison with other methods. *J Sediment Petrol* **44**: 242–248.

DeBruyn, J.M., Nixon, L.T., Fawaz, M.N., Johnson, A.M., and Radosevich, M. (2011) Global Biogeography and Quantitative Seasonal Dynamics of Gemmatimonadetes in Soil. *Appl Environ Microbiol* **77**: 6295–6300.

Deng, L., Bölsterli, D., Kristensen, E., Meile, C., Su, C.C., Bernasconi, S.M., Seidenkrantz, M.S., Glombitza, C., Lagostina, L., Han, X., Jørgensen, B.B., Røy, H., and Lever, M.A. (2020) Macrofaunal control of microbial community structure in continental margin sediments. *Proc Natl Acad Sci U S A* **117**: 15911–15922.

Dickson, R.P., Erb-Downward, J.R., Prescott, H.C., Martinez, F.J., Curtis, J.L., Lama, V.N., and Huffnagle, G.B. (2014) Analysis of culture-dependent versus culture-independent techniques for identification of bacteria in clinically obtained bronchoalveolar lavage fluid. *J Clin Microbiol* **52**: 3605–3613.

Dinno, A. (2017) dunn.test: Dunn's Test of Multiple Comparisons Using Rank Sums.

Donovan, M.K., Adam, T.C., Shantz, A.A., Speare, K.E., Munsterman, K.S., Rice, M.M., Schmitt, R.J., Holbrook, S.J., and Burkepile, D.E. (2020) Nitrogen pollution interacts with heat stress to increase coral bleaching across the seascape. *Proc Natl Acad Sci U S A* **117**:

5351–5357.

Du, Z.J., Miao, T.T., Rooney, A.P., Liu, Q.Q., and Chen, G.J. (2013) Neiella marina gen. nov., sp. nov., isolated from the sea cucumber Apostichopus japonicus. *Int J Syst Evol Microbiol* **63**: 1597–1601.

Dunn, O.J. (1964) Multiple comparisons using rank sums. *Technometrics* **6**: 241–252.

Edgar, R.C. (2017) Accuracy of microbial community diversity estimated by closed- and open-reference OTUs. *PeerJ* **2017**:.

Edgar, R.C. (2018) Updating the 97% identity threshold for 16S ribosomal RNA OTUs. *Bioinformatics* **34**: 2371–2375.

Elfeki, M., Alanjary, M., Green, S.J., Ziemert, N., and Murphy, B.T. (2018) Assessing the Efficiency of Cultivation Techniques to Recover Natural Product Biosynthetic Gene Populations from Sediment. *ACS Chem Biol* **13**: 2074–2081.

Eloe-Fadrosh, E.A., Ivanova, N.N., Woyke, T., and Kyrpides, N.C. (2016) Metagenomics uncovers gaps in amplicon-based detection of microbial diversity. *Nat Microbiol* **1**: 15032.

Evans, P.N., Parks, D.H., Chadwick, G.L., Robbins, S.J., Orphan, V.J., Golding, S.D., and Tyson, G.W. (2015) Methane metabolism in the archaeal phylum Bathyarchaeota revealed by genome-centric metagenomics. *Science (80-)* **350**: 434–438.

Faith, D.P. (1992) Conservation evaluation and phylogenetic diversity. *Biol Conserv* **61**: 1–10.

Farag, I.F., Youssef, N.H., and Elshahed, M.S. (2017) Global Distribution Patterns and Pangenomic Diversity of the Candidate Phylum "Latescibacteria" (WS3). *Appl Environ Microbiol* **83**: 1–21.

Fenical, W. and Jensen, P.R. (2006) Developing a new resource for drug discovery: Marine actinomycete bacteria. *Nat Chem Biol* **2**: 666–673.

Fernández-Gómez, B., Richter, M., Schüler, M., Pinhassi, J., Acinas, S.G., González, J.M., and Pedrós-Alió, C. (2013) Ecology of marine bacteroidetes: A comparative genomics approach. *ISME J* **7**: 1026–1037.

Fischer, M.A., Güllert, S., Neulinger, S.C., Streit, W.R., and Schmitz, R.A. (2016) Evaluation of 16S rRNA gene primer pairs for monitoring microbial community structures showed high reproducibility within and low comparability between datasets generated with multiple archaeal and bacterial primer pairs. *Front Microbiol* **7**: 1–15.

Fox, G.E., Wisotzkey, J.D., and Jurtshuk, P. (1992) How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. *Int J Syst Bacteriol* **42**: 166–170.

Fritz, I., Strömpl, C., Nikitin, D.I., Lysenko, A.M., and Abraham, W.R. (2005) Brevundimonas

mediterranea sp. nov., a non-stalked species from the Mediterranean Sea. *Int J Syst Evol Microbiol* **55**: 479–486.

Gallagher, K.A. and Jensen, P.R. (2015) Genomic insights into the evolution of hybrid isoprenoid biosynthetic gene clusters in the MAR4 marine streptomycete clade. *BMC Genomics* **16**: 1–13.

Garcia, R., La Clair, J.J., and Müller, R. (2018) Future directions of marine myxobacterial natural product discovery inferred from metagenomics. *Mar Drugs* **16**: 1–11.

Geelhoed, J.S., van de Velde, S.J., and Meysman, F.J.R. (2020) Quantification of Cable Bacteria in Marine Sediments via qPCR. *Front Microbiol* **11**: 1–13.

Ghoul, M. and Mitri, S. (2016) The Ecology and Evolution of Microbial Competition. *Trends Microbiol* **24**: 833–845.

Gil-Turnes, M., Hay, M., and Fenical, W. (1989) Symbiotic marine bacteria chemically defend crustacean embryos from a pathogenic fungus. *Science (80-)* **246**: 116–118.

Gilbert, J.A., Blaser, M.J., Caporaso, J.G., Jansson, J.K., Lynch, S. V., and Knight, R. (2018) Current understanding of the human microbiome. *Nat Med* **24**: 392–400.

Glasl, B., Bourne, D.G., Frade, P.R., Thomas, T., Schaffelke, B., and Webster, N.S. (2019) Microbial indicators of environmental perturbations in coral reef ecosystems. *Microbiome* **7**: 1–13.

Gloor, G.B., Macklaim, J.M., Pawlowsky-Glahn, V., and Egozcue, J.J. (2017) Microbiome datasets are compositional: And this is not optional. *Front Microbiol* **8**: 1–6.

Goldberg, S.R., Correa, H., Haltli, B.A., and Kerr, R.G. (2020) Fulvivirga aurantia sp. Nov. and xanthovirga aplysinae gen. nov., sp. nov., marine bacteria isolated from the sponge aplysina fistularis, and emended description of the genus fulvivirga. *Int J Syst Evol Microbiol* **70**: 2766–2781.

Gonzalez, J.M., Portillo, M.C., Belda-Ferre, P., and Mira, A. (2012) Amplification by PCR artificially reduces the proportion of the rare biosphere in microbial communities. *PLoS One* **7**:.

Gribben, P.E., Nielsen, S., Seymour, J.R., Bradley, D.J., West, M.N., and Thomas, T. (2017) Microbial communities in marine sediments modify success of an invasive macrophyte. *Sci Rep* **7**: 1–8.

Griffin, A.S., West, S.A., and Buckling, A. (2004) Cooperation and competition in pathogenic bacteria. *Nature* **430**: 1024–1027.

Gutierrez, T., Nichols, P.D., Whitman, W.B., and Aitken, M.D. (2012) Porticoccus hydrocarbonoclasticus sp. nov., an aromatic hydrocarbon-degrading bacterium identified in laboratory cultures of marine phytoplankton. *Appl Environ Microbiol* **78**: 628–637.

Haas, A.F., Fairoz, M.F.M., Kelly, L.W., Nelson, C.E., Dinsdale, E.A., Edwards, R.A., Giles, S., Hatay, M., Hisakawa, N., Knowles, B., Lim, Y.W., Maughan, H., Pantos, O., Roach, T.N.F., Sanchez, S.E., Silveira, C.B., Sandin, S., Smith, J.E., and Rohwer, F. (2016) Global microbialization of coral reefs. *Nat Microbiol* **1**: 1–7.

Haber, M., Shefer, S., Giordano, A., Orlando, P., Gambacorta, A., and Ilan, M. (2013) Luteivirga sdotyamensis gen. nov., sp. nov., a novel bacterium of the phylum Bacteroidetes isolated from the Mediterranean sponge Axinella polypoides. *Int J Syst Evol Microbiol* **63**: 939–945.

Halfvarson, J., Brislawn, C.J., Lamendella, R., Vázquez-Baeza, Y., Walters, W.A., Bramer, L.M., D'Amato, M., Bonfiglio, F., McDonald, D., Gonzalez, A., McClure, E.E., Dunklebarger, M.F., Knight, R., and Jansson, J.K. (2017) Dynamics of the human gut microbiome in inflammatory bowel disease. *Nat Microbiol* **2**: 1–7.

Hameed, A., Shahina, M., Lin, S.Y., Lai, W.A., Liu, Y.C., Hsu, Y.H., Cheng, I.C., and Young, C.C. (2014) Robertkochia marina gen. nov., sp. nov., of the family Flavobacteriaceae, isolated from surface seawater, and emended descriptions of the genera Joostella and Galbibacter. *Int J Syst Evol Microbiol* **64**: 533–539.

Hemkemeyer, M., Dohrmann, A.B., Christensen, B.T., and Tebbe, C.C. (2018) Bacterial preferences for specific soil particle size fractions revealed by community analyses. *Front Microbiol* **9**: 1–13.

Henson, M.W., Lanclos, V.C., Faircloth, B.C., and Thrash, J.C. (2018) Cultivation and genomics of the first freshwater SAR11 (LD12) isolate. *ISME J* **12**: 1846–1860.

Hillmann, B., Al-ghalith, G.A., Shields-cutler, R.R., Zhu, Q., Gohl, D.M., Beckman, K.B., Knight, R., and Knights, D. (2018) Evaluating the Information Content of Shallow Shotgun Metagenomics. *mSystems* **3**: 1–12.

Hofer, S. (2003) Determination of Ammonia (Salicylate) in 2M KCl soil extracts by Flow Injection Analysis. QuikChem Method 12-107-06-2-A., Loveland, CO.

Hoshino, T., Doi, H., Uramoto, G.I., Wörmer, L., Adhikari, R.R., Xiao, N., Morono, Y., D'Hondt, S., Hinrichs, K.U., and Inagaki, F. (2020) Global diversity of microbial communities in marine sediment. *Proc Natl Acad Sci U S A* **117**: 27587–27597.

Hug, L.A., Baker, B.J., Anantharaman, K., Brown, C.T., Probst, A.J., Castelle, C.J., Butterfield, C.N., Hernsdorf, A.W., Amano, Y., Ise, K., Suzuki, Y., Dudek, N., Relman, D.A., Finstad, K.M., Amundson, R., Thomas, B.C., and Banfield, J.F. (2016) A new view of the tree of life. *Nat Microbiol* **1**: 1–6.

Hugenholtz, P., Tyson, G.W., Webb, R.I., Wagner, A.M., and Blackall, L.L. (2001) Investigation of candidate division TM7, a recently recognized major lineage of the domain Bacteria, with no known pure-culture representatives. *Appl Environ Microbiol* **67**: 411–419.

Imachi, H., Nobu, M.K., Nakahara, N., Morono, Y., Ogawara, M., Takaki, Y., Takano, Y., Uematsu, K., Ikuta, T., Ito, M., Matsui, Y., Miyazaki, M., Murata, K., Saito, Y., Sakai, S.,

Song, C., Tasumi, E., Yamanaka, Y., Yamaguchi, T., Kamagata, Y., Tamaki, H., and Takai, K. (2020) Isolation of an archaeon at the prokaryote–eukaryote interface. *Nature* **577**: 519–525.

Imhoff, J.F. (2014) The Family Chromatiaceae. In *The Prokaryotes: Gammaproteobacteria*. Rosenberg, E., DeLong, E.F., Lory, S., Stackebrandt, E., and Thompson, F. (eds). Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 151–178.

Jackson, C.R. and Weeks, A.Q. (2008) Influence of particle size on bacterial community structure in aquatic sediments as revealed by 16S rRNA gene sequence analysis. *Appl Environ Microbiol* **74**: 5237–5240.

Jean, W.D., Shieh, W.Y., and Liu, T.Y. (2006) Thalassomonas agarivorans sp. nov., a marine agarolytic bacterium isolated from shallow coastal water of An-Ping Harbour, Taiwan, and emended description of the genus Thalassomonas. *Int J Syst Evol Microbiol* **56**: 1245–1250.

Jensen, P.R. and Mafnas, C. (2006) Biogeography of the marine actinomycete Salinispora. *Environ Microbiol* **8**: 1881–1888.

Jensen, P.R., Moore, B.S., and Fenical, W. (2015) The marine actinomycete genus Salinispora: a model organism for secondary metabolite discovery. *Nat Prod Rep* **32**: 738–751.

Johnson, J.S., Spakowicz, D.J., Hong, B.Y., Petersen, L.M., Demkowicz, P., Chen, L., Leopold, S.R., Hanson, B.M., Agresta, H.O., Gerstein, M., Sodergren, E., and Weinstock, G.M. (2019) Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nat Commun* **10**: 1–11.

Jousset, A., Bienhold, C., Chatzinotas, A., Gallien, L., Gobet, A., Kurm, V., Küsel, K., Rillig, M.C., Rivett, D.W., Salles, J.F., Van Der Heijden, M.G.A., Youssef, N.H., Zhang, X., Wei, Z., and Hol, G.W.H. (2017) Where less may be more: How the rare biosphere pulls ecosystems strings. *ISME J* **11**: 853–862.

Kaeberlein, T., Lewis, K., and Epstein, S.S. (2002) Isolating "Uncultivable" Microorganisms in Pure Culture in a Simulated Natural Environment. *Science (80-)* **296**: 1127–1129.

Kalisky, T. and Quake, S.R. (2011) Single-cell genomics. **8**: 311–314.

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

Keeney, D.R. and Nelson, D.W. (1982) Nitrogen-inorganic forms. In *Methods of soil analysis: Part 2: Chemical and microbiological properties*. Page, A.L. (ed). ASA Monograph Number 9, pp. 643–698.

Kelly, L.W., Williams, G.J., Barott, K.L., Carlson, C.A., Dinsdale, E.A., Edwards, R.A., Haas, A.F., Haynes, M., Lim, Y.W., McDole, T., Nelson, C.E., Sala, E., Sandin, S.A., Smith, J.E., Vermeij, M.J.A., Youle, M., and Rohwer, F. (2014) Local genomic adaptation of coral reef-

associated microbiomes to gradients of natural variability and anthropogenic stressors. *Proc Natl Acad Sci* **111**: 10227–10232.

Kennedy, K., Hall, M.W., Lynch, M.D.J., Moreno-Hagelsieb, G., and Neufeld, J.D. (2014) Evaluating bias of Illumina-based bacterial 16S rRNA gene profiles. *Appl Environ Microbiol* **80**: 5717–5722.

Kim, S.B., Nedashkovskaya, O.I., Mikhailov, V. V., Han, S.K., Kim, K.O., Rhee, M.S., and Bae, K.S. (2004) Kocuria marina sp. nov., a novel actinobacterium isolated from marine sediment. *Int J Syst Evol Microbiol* **54**: 1617–1620.

Kim, Y.O., Park, S., Nam, B.H., Lee, C., Park, J.M., Kim, D.G., and Yoon, J.H. (2014) Ascidiaceihabitans donghaensis gen. nov., sp. nov., isolated from the golden sea squirt Halocynthia aurantium. *Int J Syst Evol Microbiol* **64**: 3970–3975.

Knepel, K. (2003) Determination of Nitrate in 2M KCl soil extracts by Flow Injection Analysis.QuikChem Method 12-107-04-1-B, Loveland, CO.

Kovalev, K., Volkov, D., Astashkin, R., Alekseev, A., Gushchin, I., Haro-Moreno, J.M., Chizhov, I., Siletsky, S., Mamedov, M., Rogachev, A., Balandin, T., Borshchevskiy, V., Popov, A., Bourenkov, G., Bamberg, E., Rodriguez-Valera, F., Büldt, G., and Gordeliy, V. (2020) High-resolution structural insights into the heliorhodopsin family. *Proc Natl Acad Sci* **117**: 4131–4141.

Kramer, J., Özkaya, Ö., and Kümmerli, R. (2020) Bacterial siderophores in community and host interactions. *Nat Rev Microbiol* **18**: 152–163.

Krediet, C.J., Ritchie, K.B., Paul, V.J., and Teplitski, M. (2013) Coral-associated micro-organisms and their roles in promoting coral health and thwarting diseases. *Proc R Soc B Biol Sci* **280**: 20122328–20122328.

Kreeger, D., Padeletti, Angela, Citatino, S., and Padeletti, A (2010) Method for Loss-On-Ignition (LOI) Weight Analyses. 3.

Kruskal, W.H. and Wallis, W.A. (1952) Use of Ranks in One-Criterion Variance Analysis. *J Am Stat Assoc* **47**: 583–621.

Kumar, A.G., Mathew, N.C., Sujitha, K., Kirubagaran, R., and Dharani, G. (2019) Genome analysis of deep sea piezotolerant Nesiotobacter exalbescens COD22 and toluene degradation studies under high pressure condition. *Sci Rep* **9**: 1–14.

Kummerli, R., Schiessl, K.T., and Waldvogel, T.; McNeil, K.; and Ackerman, M. (2014) Habitat structure and the evolution of diffusible siderophores in bacteria. *Ecol Lett*.

Kurahashi, M. and Yokota, A. (2007) Endozoicomonils elysicola gen. nov., sp. nov., a y-proteobacterium isolated from the sea slug Elysia ornata. *Syst Appl Microbiol* **30**: 202–206.

Lahti, L. and Shetty, S. (2019) microbiome R package.

Lane, J.Q., Roddam, C.M., Langlois, G.W., and Kudela, R.M. (2010) Application of Solid Phase Adsorption Toxin Tracking (SPATT) for field detection of the hydrophilic phycotoxins domoic acid and saxitoxin in coastal California. *Limnol Oceanogr Methods* **8**: 645–660.

Lang, J.M., Coil, D.A., Neches, R.Y., Brown, W.E., Cavalier, D., Severance, M., Hampton-Marcell, J.T., Gilbert, J.A., and Eisen, J.A. (2017) A microbial survey of the International Space Station (ISS). *PeerJ* **2017**: 1–20.

Langille, M.G.I., Zaneveld, J., Caporaso, J.G., Mcdonald, D., Knights, D., Reyes, J.A., Clemente, J.C., Burkepile, D.E., Thurber, R.L.V., Knight, R., Beiko, R.G., and Huttenhower, C. (2013) Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotechnol* **31**: 814–821.

Lau, K.W.K., Ren, J., Fung, M.C., Woo, P.C.Y., Yuen, K.Y., Chan, K.K.M., Qian, P.Y., Wong, P.K., and Wu, M. (2007) Fangia hongkongensis gen. nov., sp. nov., a novel gammaproteobacterium of the order Thiotrichales isolated from coastal seawater of Hong Kong. *Int J Syst Evol Microbiol* **57**: 2665–2669.

Laursen, M.F., Dalgaard, M.D., and Bahl, M.I. (2017) Genomic GC-content affects the accuracy of 16S rRNA gene sequencing bsed microbial profiling due to PCR bias. *Front Microbiol* **8**: 1–8.

Lee, J., Shin, N.R., Jung, M.J., Roh, S.W., Kim, M.S., Lee, J.S., Lee, K.C., Kim, Y.O., and Bae, J.W. (2013) Paenibacillus oceanisediminis sp. nov. isolated from marine sediment. *Int J Syst Evol Microbiol* **63**: 428–434.

Lee, K., Lee, H.K., Choi, T.H., Kim, K.M., and Cho, J.C. (2007) Granulosicoccaceae fam. nov., to include granulosicoccus antarcticus gen. nov., sp. nov., a non-phototrophic, obligately aerobic chemoheterotroph in the order chromatiales, isolated from antarctic seawater. *J Microbiol Biotechnol* **17**: 1483–1490.

Lee, S.D. (2008) Agrococcus jejuensis sp. nov., isolated from dried seaweed. *Int J Syst Evol Microbiol* **58**: 2297–2300.

Lee, S.Y., Park, S., Oh, T.K., and Yoon, J.H. (2012) Celeribacter baekdonensis sp. nov., isolated from seawater, and emended description of the genus Celeribacter Ivanova et al. 2010. *Int J Syst Evol Microbiol* **62**: 1359–1364.

Lemos, L.N., Medeiros, J.D., Dini-Andreote, F., Fernandes, G.R., Varani, A.M., Oliveira, G., and Pylro, V.S. (2019) Genomic signatures and co-occurrence patterns of the ultra-small Saccharimonadia (phylum CPR/Patescibacteria) suggest a symbiotic lifestyle. *Mol Ecol* **28**: 4259–4271.

Lewis, W.H., Tahon, G., Geesink, P., Sousa, D.Z., and Ettema, T.J.G. (2020) Innovations to culturing the uncultured microbial majority. *Nat Rev Microbiol*.

Li, H. (2015) Microbiome, Metagenomics, and High-Dimensional Compositional Data Analysis. *Annu Rev Stat Its Appl* **2**: 73–94.

Li, S., Chen, M., Chen, Y., Tong, J., Wang, L., Xu, Y., Hu, Z., and Chen, H. (2019) Epibiotic bacterial community composition in red-tide dinoflagellate Akashiwo sanguinea culture under various growth conditions. *FEMS Microbiol Ecol* **95**: 1–12.

Li, Y., Chan, Y., Fu, Y., Zhang, R., and Chiu, J.M.Y. (2013) Coralslurrinella hongkonensis gen. nov., sp. nov., a novel bacterium in the family Psychromonadaceae, isolated from the coral Platygyra carnosus. *Antonie van Leeuwenhoek, Int J Gen Mol Microbiol* **104**: 983–991.

Lin, W., Zhang, W., Zhao, X., Roberts, A.P., Paterson, G.A., Bazylinski, D.A., and Pan, Y. (2018) Genomic expansion of magnetotactic bacteria reveals an early common origin of magnetotaxis with lineage-specific evolution. *ISME J* **12**: 1508–1519.

Liu, Z., Desantis, T.Z., Andersen, G.L., and Knight, R. (2008) Accurate taxonomy assignments from 16S rRNA sequences produced by highly parallel pyrosequencers. *Nucleic Acids Res* **36**: 1–11.

Liu, Z.P., Wang, B.J., Liu, X.Y., Dai, X., Liu, Y.H., and Liu, S.J. (2008) Paracoccus halophilus sp. nov., isolated from marine sediment of the South China Sea, China and emended description of genus Paracocccus Davis 1969. *Int J Syst Evol Microbiol* **58**: 257–261.

Lloyd, K.G., Steen, A.D., Ladau, J., Yin, J., and Crosby, L. (2018) Phylogenetically Novel Uncultured Microbial Cells Dominate Earth Microbiomes. *mSystems* **3**: 1–12.

López-Pérez, M., Haro-Moreno, J.M., Iranzo, J., and Rodriguez-Valera, F. (2020) Genomes of the "Candidatus Actinomarinales" Order: Highly Streamlined Marine Epipelagic Actinobacteria. *mSystems* **5**: 1–19.

Lozupone, C.A. and Knight, R. (2005) UniFrac : a New Phylogenetic Method for Comparing Microbial Communities UniFrac : a New Phylogenetic Method for Comparing Microbial Communities [see notes, compare to Bray-Curtis]. *Appl Environ Microbiol* **71**: 8228–8235.

Luzzatto-Knaan, T., Garg, N., Wang, M., Glukhov, E., Peng, Y., Ackermann, G., Amir, A., Duggan, B.M., Ryazanov, S., Gerwick, L., Knight, R., Alexandrov, T., Bandeira, N., Gerwick, W.H., and Dorrestein, P.C. (2017) Digitizing mass spectrometry data to explore the chemical diversity and distribution of marine cyanobacteria and algae. *Elife* **6**: 1–20.

Lynch, M.D.J. and Neufeld, J.D. (2015) Ecology and exploration of the rare biosphere. *Nat Rev Microbiol* **13**: 217–229.

Lynch, S. V. and Bruce, K.D. (2013) The cystic fibrosis airway microbiome. *Cold Spring Harb Perspect Med* **3**: 1–7.

Macher, J.-N., Prazeres, M., Taudien, S., Jompa, J., Sadekov, A., and Renema, W. (2021) Integrating morphology and metagenomics to understand taxonomic variability of Amphisorus (Foraminifera, Miliolida) from Western Australia and Indonesia. *PLoS One* **16**: e0244616.

Madigan, M.T., Martinko, J.M., and Parker, J. (2003) Brock Biology of Microorganisms, 10th

ed. Upper Saddle River, NJ: Prentice Hall.

Maldonado, L.A., Fragoso-Yáñez, D., Pérez-García, A., Rosellón-Druker, J., and Quintana, E.T. (2009) Actinobacterial diversity from marine sediments collected in Mexico. *Antonie van Leeuwenhoek, Int J Gen Mol Microbiol* **95**: 111–120.

Mandal, S., Van Treuren, W., White, R.A., Eggesbø, M., Knight, R., and Peddada, S.D. (2015) Analysis of composition of microbiomes: a novel method for studying microbial composition. *Microb Ecol Heal Dis* **26**: 1–7.

Marietou, A., Richardson, D., Cole, J., and Mohan, S. (2005) Nitrate reduction by Desulfovibrio desulfuricans: A periplasmic nitrate reductase system that lacks NapB, but includes a unique tetraheme c-type cytochrome, NapM. *FEMS Microbiol Lett* **248**: 217–225.

Martinez Arbizu, P. (2017) pairwiseAdonis: Pairwise Multilevel Comparison using Adonis.

Martiny, A.C. (2019) High proportions of bacteria are culturable across major biomes. *ISME J* 3–6.

Melton, E.D., Sorokin, D.Y., Overmars, L., Chertkov, O., Clum, A., Pillay, M., Ivanova, N., Shapiro, N., Kyrpides, N.C., Woyke, T., Lapidus, A.L., and Muyzer, G. (2016) Complete genome sequence of Desulfurivibrio alkaliphilus strain AHT2T, a haloalkaliphilic sulfidogen from Egyptian hypersaline alkaline lakes. *Stand Genomic Sci* **11**: 1–9.

Meng, Y.C., Liu, H.C., Kang, Y.Q., Zhou, Y.G., and Cai, M. (2017) Blastomonas marina sp. nov., a bacteriochlorophyll-containing bacterium isolated from seawater. *Int J Syst Evol Microbiol* **67**: 3015–3019.

Mermillod-Blondin, F., Rosenberg, R., François-Carcaillet, F., Norling, K., and Mauclaire, L. (2004) Influence of bioturbation by three benthic infaunal species on microbial communities and biogeochemical processes in marine sediment. *Aquat Microb Ecol* **36**: 271–284.

Meyer, J.L., Castellanos-Gell, J., Aeby, G.S., Häse, C.C., Ushijima, B., and Paul, V.J. (2019) Microbial community shifts associated with the ongoing stony coral tissue loss disease outbreak on the Florida reef tract. *Front Microbiol* **10**: 1–12.

Mohr, K.I., Garcia, R.O., Gerth, K., Irschik, H., and Müller, R. (2012) Sandaracinus amylolyticus gen. nov., sp. nov., a starch-degrading soil myxobacterium, and description of Sandaracinaceae fam. nov. *Int J Syst Evol Microbiol* **62**: 1191–1198.

Molinski, T.F., Biegelmeyer, R., Stout, E.P., Wang, X., Frota, M.L.C., and Henriques, A.T. (2013) Halisphingosines A and B, modified sphingoid bases from Haliclona tubifera. Assignment of configuration by circular dichroism and van't Hoff's principle of optical superposition. *J Nat Prod* **76**: 374–381.

Morton, J.T., Aksenov, A.A., Nothias, L.F., Foulds, J.R., Quinn, R.A., Badri, M.H., Swenson, T.L., Van Goethem, M.W., Northen, T.R., Vazquez-Baeza, Y., Wang, M., Bokulich, N.A.,

Watters, A., Song, S.J., Bonneau, R., Dorrestein, P.C., and Knight, R. (2019) Learning representations of microbe–metabolite interactions. *Nat Methods* **16**: 1306–1314.

Musat, N., Werner, U., Knittel, K., Kolb, S., Dodenhof, T., van Beusekom, J.E.E., de Beer, D., Dubilier, N., and Amann, R. (2006) Microbial community structure of sandy intertidal sediments in the North Sea, Sylt-Rømø Basin, Wadden Sea. *Syst Appl Microbiol* **29**: 333–348.

Nearing, J.T., Douglas, G.M., Comeau, A.M., and Langille, M.G.I. (2018) Denoising the Denoisers: An independent evaluation of microbiome sequence error- correction approaches. *PeerJ* **2018**: 1–22.

Neave, M.J., Michell, C.T., Apprill, A., and Voolstra, C.R. (2017) Endozoicomonas genomes reveal functional adaptation and plasticity in bacterial strains symbiotically associated with diverse marine hosts. *Sci Rep* **7**: 1–12.

Nedashkovskaya, O.I., Kim, S.B., Suzuki, M., Shevchenko, L.S., Lee, M.S., Lee, K.H., Park, M.S., Frolova, G.M., Oh, H.W., Bae, K.S., Park, H.Y., and Mikhailov, V. V. (2005) Pontibacter actiniarum gen. nov., sp. nov., a novel member of the phylum "Bacteroidetes", and proposal of Reichenbachiella gen. nov. as a replacement for the illegitimate prokaryotic generic name Reichenbachia Nedashkovskaya et al. 2003. *Int J Syst Evol Microbiol* **55**: 2583–2588.

Nedashkovskaya, O.I., Kim, S.B., Vancanneyt, M., Shin, D.S., Lysenko, A.M., Shevchenko, L.S., Krasokhin, V.B., Mikhailov, V. V., Swings, J., and Bae, K.S. (2006) Salegentibacter agarivorans sp. nov., a novel marine bacterium of the family Flavobacteriaceae isolated from the sponge Artemisina sp. *Int J Syst Evol Microbiol* **56**: 883–887.

Nunoura, T., Chikaraishi, Y., Izaki, R., Suwa, T., Sato, T., Harada, T., Mori, K., Kato, Y., Miyazaki, M., Shimamura, S., Yanagawa, K., Shuto, A., Ohkouchi, N., Fujita, N., Takaki, Y., Atomi, H., and Takai, K. (2018) A primordial and reversible TCA cycle in a facultatively chemolithoautotrophic thermophile. *Science (80-)* **359**: 559–563.

Ogle, D., Wheeler, P., and Dinno, A. (2020) FSA: Fisheries Stock Analysis.

Oh, H.M., Kim, H., Kim, K.M., Min, G.S., and Cho, J.C. (2010) Porticoccus litoralis gen. nov., sp. nov., a gammaproteobacterium isolated from the Yellow Sea. *Int J Syst Evol Microbiol* **60**: 727–732.

Oksanen, A.J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., Mcglinn, D., Minchin, P.R., Hara, R.B.O., Simpson, G.L., Solymos, P., Stevens, M.H.H., and Szoecs, E. (2020) vegan: Community Ecology Package.

van Oppen, M.J.H. and Blackall, L.L. (2019) Coral microbiome dynamics, functions and design in a changing world. *Nat Rev Microbiol* **17**: 557–567.

Orphan, V.J., Taylor, L.T., Hafenbradl, D., and Delong, E.F. (2000) Culture-dependent and culture-independent characterization of microbial assemblages associated with high-

temperature petroleum reservoirs. *Appl Environ Microbiol* **66**: 700–711.

Parada, A.E., Needham, D.M., and Fuhrman, J.A. (2016) Every base matters: Assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ Microbiol* **18**: 1403–1414.

Park, J.R., Bae, J.W., Nam, Y. Do, Chang, H.W., Kwon, H.Y., Quan, Z.X., and Park, Y.H. (2007) Sulfitobacter litoralis sp. nov., a marine bacterium isolated from the East Sea, Korea. *Int J Syst Evol Microbiol* **57**: 692–695.

Parkes, R.J., 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. *Mar Geol* **352**: 409–425.

Parks, D.H., Chuvochina, M., Chaumeil, P.A., Rinke, C., Mussig, A.J., and Hugenholtz, P. (2020) A complete domain-to-species taxonomy for Bacteria and Archaea. *Nat Biotechnol* **38**: 1079–1086.

Parks, D.H., Chuvochina, M., Waite, D.W., Rinke, C., Skarshewski, A., Chaumeil, P.A., and Hugenholtz, P. (2018) A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* **36**: 996.

Parks, D.H., Rinke, C., Chuvochina, M., Chaumeil, P.A., Woodcroft, B.J., Evans, P.N., Hugenholtz, P., and Tyson, G.W. (2017) Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat Microbiol* **2**: 1533–1542.

Parks, R.J. and Sass, H. (2009) Deep Sub-Surface. In *Encyclopedia of Microbiology*. Schaechter, M. (ed). Cambridge, MA: Academic Press, pp. 64–79.

Parte, A.C. (2018) LPSN - List of prokaryotic names with standing in nomenclature (Bacterio.net), 20 years on. *Int J Syst Evol Microbiol* **68**: 1825–1829.

Parte, A.C., Carbasse, J.S., Meier-Kolthoff, J.P., Reimer, L.C., and Göker, M. (2020) List of prokaryotic names with standing in nomenclature (LPSN) moves to the DSMZ. *Int J Syst Evol Microbiol* **70**: 5607–5612.

Partensky, F., Hess, W.R., and Vaulot, D. (1999) Prochlorococcus, a marine photosynthetic prokaryote of global significance. *MicrobiolMol BiolRev* **63**: 106–127.

Patin, N. V., Kunin, V., Lidström, U., and Ashby, M.N. (2013) Effects of OTU Clustering and PCR Artifacts on Microbial Diversity Estimates. *Microb Ecol* **65**: 709–719.

Patin, N. V., Schorn, M., Aguinaldo, K., Lincecum, T., Moore, B.S., and Jensen, P.R. (2017) Effects of actinomycete secondary metabolites on sediment microbial communities. *Appl Environ Microbiol* **83**: 1–16.

Patin, N. V, Duncan, K.R., Dorrestein, P.C., and Jensen, P.R. (2015) Competitive strategies differentiate closely related species of marine actinobacteria. *ISME J* 1–13.

Paul, C., Mausz, M.A., and Pohnert, G. (2013) A co-culturing/metabolomics approach to investigate chemically mediated interactions of planktonic organisms reveals influence of bacteria on diatom metabolism. *Metabolomics* **9**: 349–359.

Pédron, J., Guyon, L., Lecomte, A., Blottière, L., Chandeysson, C., Rochelle-Newall, E., Raynaud, X., Berge, O., and Barny, M.-A. (2020) Comparison of Environmental and Culture-Derived Bacterial Communities through 16S Metabarcoding: A Powerful Tool to Assess Media Selectivity and Detect Rare Taxa. *Microorganisms* **8**: 1129.

Penn, K., Jenkins, C., Nett, M., Udwary, D.W., Gontang, E.A., McGlinchey, R.P., Foster, B., Lapidus, A., Podell, S., Allen, E.E., Moore, B.S., and Jensen, P.R. (2009) Genomic islands link secondary metabolism to functional adaptation in marine Actinobacteria. *ISME J* **3**: 1193–1203.

Peoples, L.M., Grammatopoulou, E., Pombrol, M., Xu, X., Osuntokun, O., Blanton, J., Allen, E.E., Nunnally, C.C., Drazen, J.C., Mayor, D.J., and Bartlett, D.H. (2019) Microbial community diversity within sediments from two geographically separated hadal trenches. *Front Microbiol* **10**:.

Perin, L.M., Savo Sardaro, M.L., Nero, L.A., Neviani, E., and Gatti, M. (2017) Bacterial ecology of artisanal Minas cheeses assessed by culture-dependent and -independent methods. *Food Microbiol* **65**: 160–169.

Peters, E.C. (2015) Diseases of Coral Reef Organisms. In *Coral Reefs in the Anthropocene*. Birkeland, C. (ed). Dordrecht: Springer Netherlands, pp. 147–178.

Petras, D., Minich, J.J., Cancelada, L.B., Torres, R.R., Kunselman, E., Wang, M., White, M.E., Allen, E.E., Prather, K.A., Aluwihare, L.I., and Dorrestein, P.C. (2021) Non-targeted tandem mass spectrometry enables the visualization of organic matter chemotype shifts in coastal seawater. *Chemosphere* **271**: 129450.

Pielou, E.C. (1966) The measurement of diversity in different types of biological collections. *J Theor Biol* **13**: 131–144.

Pluskal, T., Castillo, S., Villar-Briones, A., and Orešič, M. (2010) MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **11**:.

Pohlner, M., Dlugosch, L., Wemheuer, B., Mills, H., Engelen, B., and Reese, B.K. (2019) The majority of active Rhodobacteraceae in marine sediments belong to uncultured genera: A molecular approach to link their distribution to environmental conditions. *Front Microbiol* **10**: 1–16.

Pollock, J., Laura, G., Trong, W., and Mick, W. (2018) The Madness of Microbiome : Attempting To Find Consensus. *Appl Environ Microbiol* **84**: 1–12.

Poulin, R.X., Poulson-Ellestad, K.L., Roy, J.S., and Kubanek, J. (2018) Variable allelopathy among phytoplankton reflected in red tide metabolome. *Harmful Algae* **71**: 50–56.

Pratchett, M.S., Trapon, M., Berumen, M.L., and Chong-Seng, K. (2011) Recent disturbances augment community shifts in coral assemblages in Moorea, French Polynesia. *Coral Reefs* **30**: 183–193.

Probandt, D., Eickhorst, T., Ellrott, A., Amann, R., and Knittel, K. (2018) Microbial life on a sand grain: From bulk sediment to single grains. *ISME J* **12**: 623–633.

Prodan, A., Tremaroli, V., Brolin, H., Zwinderman, A.H., Nieuwdorp, M., and Levin, E. (2020) Comparing bioinformatic pipelines for microbial 16S rRNA amplicon sequencing. *PLoS One* **15**: 1–19.

Prokopenko, M.G., Hirst, M.B., De Brabandere, L., Lawrence, D.J.P., Berelson, W.M., Granger, J., Chang, B.X., Dawson, S., Crane, E.J., Chong, L., Thamdrup, B., Townsend-Small, A., and Sigman, D.M. (2013) Nitrogen losses in anoxic marine sediments driven by Thioploca-anammox bacterial consortia. *Nature* **500**: 194–198.

Prosser, J.I., Head, I.M., and Stein, L.Y. (2014) The family nitrosomonadaceae. In *The Prokaryotes*. Rosenberg, E., DeLong, E., Lory, S., Stackebrandt, E., and Thompson, F. (eds). Springer Berlin Heidelberg, pp. 901–918.

Qiu, Z., Egidi, E., Liu, H., Kaur, S., and Singh, B.K. (2019) New frontiers in agriculture productivity: Optimised microbial inoculants and in situ microbiome engineering. *Biotechnol Adv* **37**: 107371.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glöckner, F.O. (2013) The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res* **41**: 590–596.

Quesada, V. (2020) nVennR: Create n-Dimensional, Quasi-Proportional Venn Diagrams.

R Core Team (2019) R: A language and environment for statistical computing.

Raaijmakers, J.M. and Mazzola, M. (2016) Soil immune response. *Science (80-)* **352**: 1392–1394.

Raina, J.-B., Tapiolas, D., Motti, C.A., Foret, S., Seemann, T., Tebben, J., Willis, B.L., and Bourne, D.G. (2016) Isolation of an antimicrobial compound produced by bacteria associated with reef-building corals. *PeerJ* **4**: e2275.

Rappé, M.S., Connon, S.A., Vergin, K.L., and Giovannoni, S.J. (2002) Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. *Nature* **418**: 630–633.

Rathsack, K., Reitner, J., Stackebrandt, E., and Tindall, B.J. (2011) Reclassification of Aurantimonas altamirensis (Jurado et al. 2006), Aurantimonas ureilytica (Weon et al. 2007) and Aurantimonas frigidaquae (Kim et al. 2008) as members of a new genus, Aureimonas gen. nov., as Aureimonas altamirensis gen. nov., comb. nov. *Int J Syst Evol Microbiol* **61**: 2722–2728.

Razumov, A.S. (1932) The direct method of calculation of bacteria in water: comparison with the Koch method. *Mikrobiologija* **1**: 131–146.

Reverter, M., Sasal, P., Suzuki, M.T., Raviglione, D., Inguimbert, N., Pare, A., Banaigs, B., Voisin, S.N., Bulet, P., and Tapissier-Bontemps, N. (2020) Insights into the natural defenses of a coral reef fish against gill ectoparasites: Integrated metabolome and microbiome approach. *Metabolites* **10**: 1–18.

Ribicic, D., Netzer, R., Hazen, T.C., Techtmann, S.M., Drabløs, F., and Brakstad, O.G. (2018) Microbial community and metagenome dynamics during biodegradation of dispersed oil reveals potential key-players in cold Norwegian seawater. *Mar Pollut Bull* **129**: 370–378.

Riley, M.A. and Gordon, D.M. (1999) The ecological role of bacteriocins in bacterial competition. *Trends Microbiol* **7**: 129–133.

Rios-Del Toro, E.E., Valenzuela, E.I., López-Lozano, N.E., Cortés-Martínez, M.G., Sánchez-Rodríguez, M.A., Calvario-Martínez, O., Sánchez-Carrillo, S., and Cervantes, F.J. (2018) Anaerobic ammonium oxidation linked to sulfate and ferric iron reduction fuels nitrogen loss in marine sediments. *Biodegradation* **29**: 429–442.

Ritchie, K.B. (2006) Regulation of microbial populations by coral surface mucus and mucus-associated bacteria. *Mar Ecol Prog Ser* **322**: 1–14.

Roberts, A.A., Schultz, A.W., Kersten, R.D., Dorrestein, P.C., and Moore, B.S. (2012) Iron acquisition in the marine actinomycete genus Salinispora is controlled by the desferrioxamine family of siderophores. *FEMS Microbiol Lett* **335**: 95–103.

Robeson II, M.S., O'Rourke, D.R., Kaehler, B.D., Ziemski, M., Dillon, M.R.., Foster, J.T., and Bokulich, N.A. (2020) RESCRIPt: Reproducible sequence taxonomy reference database management for the masses تتتت. *bioRxiv*.

Rohart, F., Gautier, B., Singh, A., and Lê Cao, K.A. (2017) mixOmics: An R package for 'omics feature selection and multiple data integration. *PLoS Comput Biol* **13**: 1–19.

Romanenko, L. a, Schumann, P., Rohde, M., Lysenko, A.M., Mikhailov, V. V, and Stackebrandt, E. (2002) Halophiles From Marine Environments. *Int J Syst Evol Microbiol* **52**: 1291–1297.

Romanenko, L.A., Schumann, P., Rohde, M., and Mikhailov, V. V (2002) Halomonas halocynthiae sp. nov., isolated from the marine ascidian Halocynthia aurantium. *Int J Syst Evol Microbiol* 1767–1772.

Romanenko, L.A., Tanaka, N., and Frolova, G.M. (2009) Marinomonas arenicola sp. nov., isolated from marine sediment. *Int J Syst Evol Microbiol* **59**: 2834–2838.

Rosselló-Mora, R. and Amann, R. (2001) The species concept for prokaryotes. *FEMS Microbiol Rev* **25**: 39–67.

Rusch, A., Huettel, M., Reimers, C.E., Taghon, G.L., and Fuller, C.M. (2003) Activity and

distribution of bacterial populations in Middle Atlantic Bight shelf sands. *Fems Microbiol Ecol* **44**: 89–100.

Rygaard, A. Mac, Thøgersen, M.S., Nielsen, K.F., Gram, L., and Bentzon-Tilia, M. (2017) Effects of gelling agent and extracellular signaling molecules on the culturability of marine bacteria. *Appl Environ Microbiol* **83**: 1–16.

Sah, R.N. and Miller, R.O. (1992) Spontaneous Reaction for Acid Dissolution of Biological Tissues in Closed Vessels. *Anal Chem* **64**: 230–233.

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

Santelli, C.M., Orcutt, B.N., Banning, E., Bach, W., Moyer, C.L., Sogin, M.L., Staudigel, H., and Edwards, K.J. (2008) Abundance and diversity of microbial life in ocean crust. *Nature* **453**: 653–656.

Santmire, J.A. and Leff, L.G. (2007) The effect of sediment grain size on bacterial communities in streams. *J North Am Benthol Soc* **26**: 601–610.

Schelli, K., Zhong, F., and Zhu, J. (2017) Comparative metabolomics revealing Staphylococcus aureus metabolic response to different antibiotics. *Microb Biotechnol* **10**: 1764–1774.

Schlatter, D., Kinkel, L., Thomashow, L., Weller, D., and Paulitz, T. (2017) Disease suppressive soils: New insights from the soil microbiome. *Phytopathology* **107**: 1284–1297.

Schloss, P.D. and Westcott, S.L. (2011) Assessing and improving methods used in OTU-based approaches for 16S rRNA gene sequence analysis. *Appl Environ Microbiol* **77**: 3219.

Schreiber, L., Kjeldsen, K.U., Funch, P., Jensen, J., Obst, M., López-Legentil, S., and Schramm, A. (2016) Endozoicomonas are specific, facultative symbionts of sea squirts. *Front Microbiol* **7**: 1–15.

Schut, F., Gottschal, J.C., and Prins, R.A. (1997) Isolation and characterisation of the marine ultramicrobacterium Sphingomonas sp. strain RB2256. *FEMS Microbiol Rev* **20**: 363–369.

Sessitsch, A., Weilharter, A., Gerzabek, M.H., Kirchmann, H., and Kandeler, E. (2001) Microbial Population Structures in Soil Particle Size Fractions of a Long-Term Fertilizer Field Experiment. *Appl Environ Microbiol* **67**: 4215–4224.

Shannon, C.E. and Weaver, W. (1964) The Relationship Between Distinctive Capabilities , Innovativeness , Strategy T ... *Int Bus* **8**: 21–33.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003) Cytoscape: A Software Environment for Integrated Models. *Genome Res* **13**: 2498–2504.

Shao, R., Lai, Q., Liu, X., Sun, F., Du, Y., Li, G., and Shao, Z. (2014) Zunongwangia atlantica sp. nov., isolated from deep-sea water. *Int J Syst Evol Microbiol* **64**: 16–20.

Shivaji, S., Kumari, K., Kishore, K.H., Pindi, P.K., Rao, P.S., Radha Srinivas, T.N., Asthana, R., and Ravindra, R. (2011) Vertical distribution of bacteria in a lake sediment from Antarctica by culture-independent and culture-dependent approaches. *Res Microbiol* **162**: 191–203.

Shreiner, A.B., Kao, J.Y., and Young, V.B. (2015) The gut microbiome in health and in disease. *Curr Opin Gastroenterol* **31**: 69–75.

Smith, A.L. and Nicolaou, K.C. (1996) The enediyne antibiotics. *J Med Chem* **39**: 2103–2117.

Snelgrove, P.V.R. (1997) The Importance of Marine Sediment Biodiversity in Ecosystem Processes The Importance of Marine Sediment Biodiversity in Ecosystem Processes. *Ambio* **26**: 578–583.

Snelgrove, P.V.R., Henry Blackburn, T., Hutchings, P.A., Alongi, D.M., Frederick Grassle, J., Hummel, H., King, G., Koike, I., Lambshead, P.J.D., Ramsing, N.B., and Solis-Weiss, V. (1997) The importance of marine sediment biodiversity in ecosystem processes. *Ambio* **26**: 578–583.

Soliman, T., Reimer, J.D., Yang, S.Y., Villar-Briones, A., Roy, M.C., and Jenke-Kodama, H. (2017) Diversity of microbial communities and quantitative chemodiversity in layers of marine sediment cores from a causeway (Kaichu-Doro) in Okinawa Island, Japan. *Front Microbiol* **8**:.

Soro, V., Dutton, L.C., Sprague, S. V., Nobbs, A.H., Ireland, A.J., Sandy, J.R., Jepson, M.A., Micaroni, M., Splatt, P.R., Dymock, D., and Jenkinson, H.F. (2014) Axenic Culture of a Candidate Division TM7 Bacterium from the Human Oral Cavity and Biofilm Interactions with Other Oral Bacteria. *Appl Environ Microbiol* **80**: 6480–6489.

Sorokin, D.Y., Merkel, A.Y., and Muyzer, G. (2015) Thiohalophilus. *Bergey's Man Syst Archaea Bact* 1–7.

de Souza, J.T., Weller, D.M., and Raaijmakers, J.M. (2003) Frequency, Diversity, and Activity of 2,4-Diacetylphloroglucinol-Producing Fluorescent Pseudomonas spp. in Dutch Take-all Decline Soils. *Phytopathology* **93**: 54–63.

Stackebrandt, E. and Goebel, B.M. (1994) Taxonomic note: A place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *Int J Syst Bacteriol* **44**: 846–849.

Staley, J.T. (2006) The bacterial species dilemma and the genomic-phylogenetic species concept. *Philos Trans R Soc B Biol Sci* **361**: 1899–1909.

Staley, J.T. and Konopka, A. (1985) Microorganisms in Aquatic and Terrestrial Habitats. *Annu Rev Microbiol* **39**: 321–346.

Steen, A.D., Crits-Christoph, A., Carini, P., DeAngelis, K.M., Fierer, N., Lloyd, K.G., and Cameron Thrash, J. (2019) High proportions of bacteria and archaea across most biomes remain uncultured. *ISME J*.

Straub, D., Blackwell, N., Fuentes, A.L., Peltzer, A., Nahnsen, S., and Kleindienst, S. (2019) Interpretations of microbial community studies are biased by the selected 16S rRNA gene amplicon sequencing pipeline. **11**: 1–18.

Tamaki, H., Hanada, S., Sekiguchi, Y., Tanaka, Y., and Kamagata, Y. (2009) Effect of gelling agent on colony formation in solid cultivation of microbial community in lake sediment. *Environ Microbiol* **11**: 1827–1834.

Tanaka, T., Kawasaki, K., Daimon, S., Kitagawa, W., Yamamoto, K., Tamaki, H., Tanaka, M., Nakatsu, C.H., and Kamagata, Y. (2014) A Hidden Pitfall in the Preparation of Agar Media Undermines. *Appl Environ Microbiol* **80**: 7659–7666.

Tang, Z.Z., Chen, G., Hong, Q., Huang, S., Smith, H.M., Shah, R.D., Scholz, M., and Ferguson, J.F. (2019) Multi-omic analysis of the microbiome and metabolome in healthy subjects reveals microbiome-dependent relationships between diet and metabolites. *Front Genet* **10**:.

Te, S.H., Tan, B.F., Boo, C.Y., Thompson, J.R., and Gin, K.Y.H. (2017) Genomics insights into production of 2-methylisoborneol and a putative cyanobactin by Planktothricoides sp. SR001. *Stand Genomic Sci* **12**: 1–9.

Teasdale, M.E., Liu, J., Wallace, J., Akhlaghi, F., and Rowley, D.C. (2009) Secondary metabolites produced by the marine bacterium halobacillus salinus that Inhibit quorum sensing-controlled phenotypes in gram-negative bacteria. *Appl Environ Microbiol* **75**: 567–572.

Teramoto, M., Ohuchi, M., Hatmanti, A., Darmayati, Y., Widyastuti, Y., Harayama, S., and Fukunaga, Y. (2011) Oleibacter marinus gen. nov., sp. nov., a bacterium that degrades petroleum aliphatic hydrocarbons in a tropical marine environment. *Int J Syst Evol Microbiol* **61**: 375–380.

Thajudeen, J., Yousuf, J., Veetil, V.P., Varghese, S., Singh, A., and Abdulla, M.H. (2017) Nitrogen fixing bacterial diversity in a tropical estuarine sediments. *World J Microbiol Biotechnol* **33**: 1–11.

Thomas, T., Easson, C., Moitinho-silva, L., Lurgi, M., Bjo, J.R., Luter, H., Olson, J.B., Erwin, P.M., Lo, S., Astudillo-garcı, C., Chaves-fonnegra, A., Costa, R., Schupp, P.J., Steindler, L., and Erpenbeck, D. (2016) Diversity, structure and convergent evolution of the global sponge microbiome. *Nat Commun*.

Thompson, L.R., Sanders, J.G., McDonald, D., Amir, A., Ladau, J., Locey, K.J., Prill, R.J., Tripathi, A., Gibbons, S.M., Ackermann, G., Navas-Molina, J.A., Janssen, S., Kopylova, E., Vázquez-Baeza, Y., González, A., Morton, J.T., Mirarab, S., Xu, Z.Z., Jiang, L., Haroon, M.F., Kanbar, J., Zhu, Q., Song, S.J., Kosciolek, T., Bokulich, N.A., Lefler, J., Brislawn, C.J., Humphrey, G., Owens, S.M., Hampton-Marcell, J., Berg-Lyons, D., McKenzie, V.,

Fierer, N., Fuhrman, J.A., Clauset, A., Stevens, R.L., Shade, A., Pollard, K.S., Goodwin, K.D., Jansson, J.K., Gilbert, J.A., Knight, R., Agosto Rivera, J.L., Al-Moosawi, L., Alverdy, J., Amato, K.R., Andras, J., Angenent, L.T., Antonopoulos, D.A., Apprill, A., Armitage, D., Ballantine, K., Bárta, J., Baum, J.K., Berry, A., Bhatnagar, A., Bhatnagar, M., Biddle, J.F., Bittner, L., Boldgiv, B., Bottos, E., Boyer, D.M., Braun, J., Brazelton, W., Brearley, F.Q., Campbell, A.H., Caporaso, J.G., Cardona, C., Carroll, J.L., Cary, S.C., Casper, B.B., Charles, T.C., Chu, H., Claar, D.C., Clark, R.G., Clayton, J.B., Clemente, J.C., Cochran, A., Coleman, M.L., Collins, G., Colwell, R.R., Contreras, M., Crary, B.B., Creer, S., Cristol, D.A., Crump, B.C., Cui, D., Daly, S.E., Davalos, L., Dawson, R.D., Defazio, J., Delsuc, F., Dionisi, H.M., Dominguez-Bello, M.G., Dowell, R., Dubinsky, E.A., Dunn, P.O., Ercolini, D., Espinoza, R.E., Ezenwa, V., Fenner, N., Findlay, H.S., Fleming, I.D., Fogliano, V., Forsman, A., Freeman, C., Friedman, E.S., Galindo, G., Garcia, L., Garcia-Amado, M.A., Garshelis, D., Gasser, R.B., Gerdts, G., Gibson, M.K., Gifford, I., Gill, R.T., Giray, T., Gittel, A., Golyshin, P., Gong, D., Grossart, H.P., Guyton, K., Haig, S.J., Hale, V., Hall, R.S., Hallam, S.J., Handley, K.M., Hasan, N.A., Haydon, S.R., Hickman, J.E., Hidalgo, G., Hofmockel, K.S., Hooker, J., Hulth, S., Hultman, J., Hyde, E., Ibáñez-Álamo, J.D., Jastrow, J.D., Jex, A.R., Johnson, L.S., Johnston, E.R., Joseph, S., Jurburg, S.D., Jurelevicius, D., Karlsson, A., Karlsson, R., Kauppinen, S., Kellogg, C.T.E., Kennedy, S.J., Kerkhof, L.J., King, G.M., Kling, G.W., Koehler, A. V., Krezalek, M., Kueneman, J., Lamendella, R., Landon, E.M., Lanede Graaf, K., LaRoche, J., Larsen, P., Laverock, B., Lax, S., Lentino, M., Levin, I.I., Liancourt, P., Liang, W., Linz, A.M., Lipson, D.A., Liu, Y., Lladser, M.E., Lozada, M., Spirito, C.M., MacCormack, W.P., MacRae-Crerar, A., Magris, M., Martín-Platero, A.M., Martín-Vivaldi, M., Martínez, L.M., Martínez-Bueno, M., Marzinelli, E.M., Mason, O.U., Mayer, G.D., McDevitt-Irwin, J.M., McDonald, J.E., McGuire, K.L., McMahon, K.D., McMinds, R., Medina, M., Mendelson, J.R., Metcalf, J.L., Meyer, F., Michelangeli, F., Miller, K., Mills, D.A., Minich, J., Mocali, S., Moitinho-Silva, L., Moore, A., Morgan-Kiss, R.M., Munroe, P., Myrold, D., Neufeld, J.D., Ni, Y., Nicol, G.W., Nielsen, S., Nissimov, J.I., Niu, K., Nolan, M.J., Noyce, K., O'Brien, S.L., Okamoto, N., Orlando, L., Castellano, Y.O., Osuolale, O., Oswald, W., Parnell, J., Peralta-Sánchez, J.M., Petraitis, P., Pfister, C., Pilon-Smits, E., Piombino, P., Pointing, S.B., Pollock, F.J., Potter, C., Prithiviraj, B., Quince, C., Rani, A., Ranjan, R., Rao, S., Rees, A.P., Richardson, M., Riebesell, U., Robinson, C., Rockne, K.J., Rodriguezl, S.M., Rohwer, F., Roundstone, W., Safran, R.J., Sangwan, N., Sanz, V., Schrenk, M., Schrenzel, M.D., Scott, N.M., Seger, R.L., Seguinorlando, A., Seldin, L., Seyler, L.M., Shakhsheer, B., Sheets, G.M., Shen, C., Shi, Y., Shin, H., Shogan, B.D., Shutler, D., Siegel, J., Simmons, S., Sjöling, S., Smith, D.P., Soler, J.J., Sperling, M., Steinberg, P.D., Stephens, B., Stevens, M.A., Taghavi, S., Tai, V., Tait, K., Tan, C.L., Taş, N., Taylor, D.L., Thomas, T., Timling, I., Turner, B.L., Urich, T., Ursell, L.K., Van Der Lelie, D., Van Treuren, W., Van Zwieten, L., Vargas-Robles, D., Thurber, R.V., Vitaglione, P., Walker, D.A., Walters, W.A., Wang, S., Wang, T., Weaver, T., Webster, N.S., Wehrle, B., Weisenhorn, P., Weiss, S., Werner, J.J., West, K., Whitehead, A., Whitehead, S.R., Whittingham, L.A., Willerslev, E., Williams, A.E., Wood, S.A., Woodhams, D.C., Yang, Y., Zaneveld, J., Zarraonaindia, I., Zhang, Q., and Zhao, H. (2017) A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* **551**: 457–463.

Thorup, C., Schramm, A., Findlay, A.J., Finster, K.W., and Schreiber, L. (2017) Disguised as a sulfate reducer: Growth of the deltaproteobacterium Desulfurivibrio alkaliphilus by Sulfide

Oxidation with Nitrate. *MBio* **8**: 1–9.

Tripp, H.J., Kitner, J.B., Schwalbach, M.S., Dacey, J.W.H., Wilhelm, L.J., and Giovannoni, S.J. (2008) SAR11 marine bacteria require exogenous reduced sulphur for growth. *Nature* **452**: 741–744.

Tuttle, R.N., Demko, A.M., Jensen, P.R., Patin, N. V, Kapono, C.A., Donia, M.S., and Dorrestein, P. (2019) Detection of natural products and their producers in ocean sediments. *Appl Environ Microbiol* **85**: 1–15.

Vaz-Moreira, I., Egas, C., Nunes, O.C., and Manaia, C.M. (2011) Culture-dependent and culture-independent diversity surveys target different bacteria : a case study in a freshwater sample. *Antonie van Leeuwenhoek* **100**: 245–257.

Vázquez-Baeza, Y., Pirrung, M., Gonzalez, A., and Knight, R. (2013) EMPeror: A tool for visualizing high-throughput microbial community data. *Gigascience* **2**: 2–5.

Větrovský, T. and Baldrian, P. (2013) The Variability of the 16S rRNA Gene in Bacterial Genomes and Its Consequences for Bacterial Community Analyses. *PLoS One* **8**: 1–10.

Wang, M., Carver, J.J., Phelan, V. V., Sanchez, L.M., Garg, N., Peng, Y., Nguyen, D.D., Watrous, J., Kapono, C.A., Luzzatto-Knaan, T., Porto, C., Bouslimani, A., Melnik, A. V., Meehan, M.J., Liu, W.T., Crüsemann, M., Boudreau, P.D., Esquenazi, E., Sandoval-Calderón, M., Kersten, R.D., Pace, L.A., Quinn, R.A., Duncan, K.R., Hsu, C.C., Floros, D.J., Gavilan, R.G., Kleigrewe, K., Northen, T., Dutton, R.J., Parrot, D., Carlson, E.E., Aigle, B., Michelsen, C.F., Jelsbak, L., Sohlenkamp, C., Pevzner, P., Edlund, A., McLean, J., Piel, J., Murphy, B.T., Gerwick, L., Liaw, C.C., Yang, Y.L., Humpf, H.U., Maansson, M., Keyzers, R.A., Sims, A.C., Johnson, A.R., Sidebottom, A.M., Sedio, B.E., Klitgaard, A., Larson, C.B., Boya, C.A.P., Torres-Mendoza, D., Gonzalez, D.J., Silva, D.B., Marques, L.M., Demarque, D.P., Pociute, E., O'Neill, E.C., Briand, E., Helfrich, E.J.N., Granatosky, E.A., Glukhov, E., Ryffel, F., Houson, H., Mohimani, H., Kharbush, J.J., Zeng, Y., Vorholt, J.A., Kurita, K.L., Charusanti, P., McPhail, K.L., Nielsen, K.F., Vuong, L., Elfeki, M., Traxler, M.F., Engene, N., Koyama, N., Vining, O.B., Baric, R., Silva, R.R., Mascuch, S.J., Tomasi, S., Jenkins, S., Macherla, V., Hoffman, T., Agarwal, V., Williams, P.G., Dai, J., Neupane, R., Gurr, J., Rodríguez, A.M.C., Lamsa, A., Zhang, C., Dorrestein, K., Duggan, B.M., Almaliti, J., Allard, P.M., Phapale, P., Nothias, L.F., Alexandrov, T., Litaudon, M., Wolfender, J.L., Kyle, J.E., Metz, T.O., Peryea, T., Nguyen, D.T., VanLeer, D., Shinn, P., Jadhav, A., Müller, R., Waters, K.M., Shi, W., Liu, X., Zhang, L., Knight, R., Jensen, P.R., Palsson, B., Pogliano, K., Linington, R.G., Gutiérrez, M., Lopes, N.P., Gerwick, W.H., Moore, B.S., Dorrestein, P.C., and Bandeira, N. (2016) Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat Biotechnol* **34**: 828–837.

Ward, N.L., Challacombe, J.F., Janssen, P.H., Henrissat, B., Coutinho, P.M., Wu, M., Xie, G., Haft, D.H., Sait, M., Badger, J., Barabote, R.D., Bradley, B., Brettin, T.S., Brinkac, L.M., Bruce, D., Creasy, T., Daugherty, S.C., Davidsen, T.M., DeBoy, R.T., Detter, J.C., Dodson, R.J., Durkin, A.S., Ganapathy, A., Gwinn-Giglio, M., Han, C.S., Khouri, H., Kiss, H.,

Kothari, S.P., Madupu, R., Nelson, K.E., Nelson, W.C., Paulsen, I., Penn, K., Ren, Q., Rosovitz, M.J., Selengut, J.D., Shrivastava, S., Sullivan, S.A., Tapia, R., Thompson, S., Watkins, K.L., Yang, Q., Yu, C., Zafar, N., Zhou, L., and Kuske, C.R. (2009) Three genomes from the phylum Acidobacteria provide insight into the lifestyles of these microorganisms in soils. *Appl Environ Microbiol* **75**: 2046–2056.

Waters, C.M. and Bassler, B.L. (2005) Quorum sensing : Cell-to-cell communication in bacteria. *Annu Rev Cell Dev Biol* **21**: 319–346.

Watve, M., Shejval, V., Sonawane, C., and Rahalkar, M. (2020) Current Science Association The ' K ' selected oligophilic bacteria : A key to uncultured diversity ? Matapurkar , Yogesh Shouche , Milind Patole , Nitin Phadnis , Ameya Champhenkar , Kalyani Damle , Sulabha Karandikar , Vasudev Kshirsagar and Maithili Jo.

Wear, E.K., Wilbanks, E.G., Nelson, C.E., and Carlson, C.A. (2018) Primer selection impacts specific population abundances but not community dynamics in a monthly time-series 16S rRNA gene amplicon analysis of coastal marine bacterioplankton. *Environ Microbiol* **20**: 2709–2726.

Weingarden, A., González, A., Vázquez-Baeza, Y., Weiss, S., Humphry, G., Berg-Lyons, D., Knights, D., Unno, T., Bobr, A., Kang, J., Khoruts, A., Knight, R., and Sadowsky, M.J. (2015) Dynamic changes in short- and long-term bacterial composition following fecal microbiota transplantation for recurrent Clostridium difficile infection. *Microbiome* **3**: 1–8.

Wickham, H. (2016) ggplot2: Elegant Graphics for Data Analysis, New York: Springer-Verlag.

Wickham, H., Hester, J., and Chang, W. (2020) devtools: Tools to Make Developing R Packages Easier.

Wietz, M., Duncan, K., Patin, N. V., and Jensen, P.R. (2013) Antagonistic Interactions Mediated by Marine Bacteria: The Role of Small Molecules. *J Chem Ecol* **39**: 879–891.

Willis, C., Desai, D., and Laroche, J. (2019) Influence of 16S rRNA variable region on perceived diversity of marine microbial communities of the Northern North Atlantic. *FEMS Microbiol Lett* **366**: 1–9.

Wu, Y.S., Yu, G.L., Jiang, H.X., Liu, L.J., and Zhao, R. (2016) Role and lifestyle of calcified cyanobacteria (Stanieria) in Permian-Triassic boundary microbialites. *Palaeogeogr Palaeoclimatol Palaeoecol* **448**: 39–47.

Xia, Y. and Sun, J. (2017) Hypothesis testing and statistical analysis of microbiome. *Genes Dis* **4**: 138–148.

Yang, B., Wang, Y., and Qian, P.Y. (2016) Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. *BMC Bioinformatics* **17**: 1–8.

Yang, S.H., Seo, H.S., Lee, J.H., Kim, S.J., and Kwon, K.K. (2016) Pseudofulvibacter gastropodicola sp. Nov., isolated from a marine conch and emended descriptions of the

genus pseudofulvibacter yoon et al. 2013 and pseudofulvibacter geojedonensis. *Int J Syst Evol Microbiol* **66**: 430–434.

Yi, H. and Chun, J. (2006) Thalassobius aestuarii sp. nov., isolated from tidal flat sediment. *J Microbiol* **44**: 171–176.

Yim, C.Y., Le, T.C., Lee, T.G., Yang, I., Choi, H., Lee, J., Kang, K.Y., Lee, J.S., Lim, K.M., Yee, S.T., Kang, H., Nam, S.J., and Fenical, W. (2017) Saccharomonopyrones A–C, new α-pyrones from a marine sediment-derived bacterium Saccharomonospora sp. CNQ-490. *Mar Drugs* **15**: 4–11.

Youssef, N.H., Farag, I.F., Rinke, C., Hallam, S.J., Woyke, T., and Elshahed, M.S. (2015) In silico analysis of the metabolic potential and niche specialization of candidate phylum "Latescibacteria" (WS3). *PLoS One* **10**: 1–21.

Zak, D.R. and Kling, G.W. (2006) Microbial community composition and function across an arctic tundra landscape. *Ecology* **87**: 1659–1670.

Zaneveld, J.R., McMinds, R., and Thurber, R.V. (2017) Stress and stability: Applying the Anna Karenina principle to animal microbiomes. *Nat Microbiol* **2**: 1–8.

Zeng, Y.X., Yu, Y., Li, H.R., and Luo, W. (2017) Prokaryotic community composition in arctic kongsfjorden and sub-arctic northern bering sea sediments as revealed by 454 pyrosequencing. *Front Microbiol* **8**:.

Zhang, Q., Chen, H., Huang, D., Xu, C., Zhu, H., and Zhu, Q. (2019) Water managements limit heavy metal accumulation in rice: Dual effects of iron-plaque formation and microbial communities. *Sci Total Environ* **687**: 790–799.

Zheng, B., Wang, L., and Liu, L. (2014) Bacterial community structure and its regulating factors in the intertidal sediment along the Liaodong Bay of Bohai Sea, China. *Microbiol Res* **169**: 585–592.

Ziemert, N., Lechner, A., Wietz, M., Millán-Aguiñaga, N., Chavarria, K.L., and Jensen, P.R. (2014) Diversity and evolution of secondary metabolism in the marine actinomycete genus Salinispora. *Proc Natl Acad Sci U S A* **111**: E1130-9.

Zierer, J., Jackson, M.A., Kastenmüller, G., Mangino, M., Long, T., Telenti, A., Mohney, R.P., Small, K.S., Bell, J.T., Steves, C.J., Valdes, A.M., Spector, T.D., and Menni, C. (2018) The fecal metabolome as a functional readout of the gut microbiome. *Nat Genet* **50**: 790–795.

Appendix A

Amplicon sequence variants (ASVs) from Chapter 2 controls

Table A.1 Amplicon sequence variants (ASVs) detected in control samples. Blank controls were marine agar (MA) or seawater agar (SWA) plates spiked with *Vibrio* and extracted. Inoculum controls were MA or SWA plates spiked with *Vibrio* and fresh sediment inoculum. All controls were done in triplicate, but DNA was pooled prior to 16S sequencing resulting in one replicate per control sample type. Taxonomy was assigned with SILVA v132 after denoising with DADA2. ASVs associated with the *Vibrio* spike-in are omitted from the table. See methods section in Chapter 2 for further details.

Control Sample Type	**ASV ID**	**ASV Frequency**	**Taxonomy (SILVA v132)**
MA Inoculum Control	e404670ea4fca621d9f2756b247d9175	67	D_0__Bacteria
MA Inoculum Control	60d5e57286a1514bec6769b7053f176e	47	D_0__Bacteria
MA Inoculum Control	cee310843c63ba44a8bfd3ad880cf100	47	Unassigned
SWA Blank	e3b2eebceee48d4957957ffa3c7a9b77	40	Unassigned
MA Inoculum Control	786af38bd06a643d424015418485259f	38	Unassigned
SWA Inoculum Control	715816a6597a6ace8baf53c84d866264	31	D_0__Bacteria;D_1__Deinococcus-Thermus;D_2__Deinococci;D_3__Thermales;D_4__Thermaceae;D_5__Thermus
MA Inoculum Control	40708c9a8539ffbfb24e9a1cda919dd1	30	D_0__Bacteria
MA Inoculum Control	c8edde6a5e3f5f6a7ee86aee147446d4	25	Unassigned

Table A.1 Amplicon sequence variants (ASVs) detected in control samples, Continued.

Control Sample Type	ASV ID	ASV Frequency	Taxonomy (SILVA v132)
MA Inoculum Control	f3e48ce7e77a3fd1ffb5dbe49d1801c3	20	Unassigned
MA Inoculum Control	ffc60c2bab515b700acae3ce5fff026e	20	Unassigned
MA Inoculum Control	b390ffb1f55e5e59a9131bb7c3e3aaa9	17	Unassigned
SWA Blank	d4d8aaf721a16d4a1021ac0719a3ab80	15	D_0__Bacteria;D_1__Proteobacteria;D_2__Gammaproteobacteria; D_3__Xanthomonadales;D_4__Xanthomonadaceae;D_5__Stenotrophomonas; D_6__Stenotrophomonas rhizophila
MA Inoculum Control	77962eb5decfa77b2e60ae9aa845f61d	10	Unassigned
SWA Inoculum Control	e1c9956bad83dc6aaffd55873b0c28ac	10	D_0__Bacteria;D_1__Proteobacteria;D_2__Gammaproteobacteria; D_3__Pseudomonadales;D_4__Pseudomonadaceae; D_5__Pseudomonas
MA Inoculum Control	de36abbf7f16b4b8e210d8b17de2cfd4	9	Unassigned
SWA Inoculum Control	d514209c2927e2919df1347dd2f42fe4	9	D_0__Bacteria;D_1__Proteobacteria;D_2__Gammaproteobacteria; D_3__Pseudomonadales;D_4__Moraxellaceae;D_5__Acinetobacter
MA Inoculum Control	445e32f6ef6dd17439a6afc3d029b43d	8	Unassigned

Table A.1 Amplicon sequence variants (ASVs) detected in control samples, Continued.

Control Sample Type	ASV ID	ASV Frequency	Taxonomy (SILVA v132)
MA Inoculum Control	83b562bcc8ad656537f2d2d02e9cc4f6	8	Unassigned
SWA Inoculum Control	1d2947fc7b331bf28a4f0f1ef68e783e	8	D_0__Bacteria;D_1__Firmicutes;D_2__Bacilli;D_3__Lactobacillales; D_4__Lactobacillaceae;D_5__Lactobacillus; D_6__Lactobacillus iners AB-1
MA Inoculum Control	2413a34b3fa758167e898a0af1ec640f	7	D_0__Bacteria
MA Inoculum Control	3ba6a3fbab18bf492bcdf6d9a0bb4273	7	Unassigned
MA Inoculum Control	693e3170491aa62e5d552b372b323a2f	7	Unassigned
MA Inoculum Control	c69538568a8a9418d11c9d9d1ec3b900	6	Unassigned
MA Inoculum Control	96ccc3efa8b2d674220c5e74b18c976e	5	D_0__Bacteria
MA Inoculum Control	bfadd2679ef710f658ec65ec8b54f7ed	5	Unassigned
MA Inoculum Control	da119c120ad996f30af4e1594de0155a	5	Unassigned
SWA Blank	9da99a5d6393fd1e85203ff8d2e6e53a	5	D_0__Bacteria;D_1__Proteobacteria;D_2__Alphaproteobacteria; D_3__Rhodobacterales;D_4__Rhodobacteraceae

Table A.1 Amplicon sequence variants (ASVs) detected in control samples, Continued.

Control Sample Type	ASV ID	ASV Frequency	Taxonomy (SILVA v132)
SWA Inoculum Control	721d7a0077bff1887da3d179f87a901e	5	D_0__Bacteria;D_1__Cyanobacteria;D_2__Oxyphotobacteria; D_3__Chloroplast
MA Blank	48a5041c9bb7c914ed6394a47a02b1a8	4	Unassigned
MA Blank	c82e52e30fe2a40eacac9de50db1f930	4	D_0__Bacteria;D_1__Proteobacteria;D_2__Gammaproteobacteria; D_3__Pseudomonadales;D_4__Pseudomonadaceae; D_5__Pseudomonas
MA Inoculum Control	822a084890ea650a2fcda5da525cc9b6	4	Unassigned
MA Inoculum Control	86fc8b757d981481f8ebbae180a8a9d2	4	Unassigned
SWA Inoculum Control	fc5f950b9ef328a6e431375c57a9e923	4	D_0__Bacteria;D_1__Bacteroidetes;D_2__Bacteroidia; D_3__Bacteroidales;D_4__Muribaculaceae; D_5__uncultured bacterium;D_6__uncultured bacterium
SWA Inoculum Control	ff52c94b8a4aadb880409dee2c3555b7	4	D_0__Bacteria;D_1__Proteobacteria;D_2__Alphaproteobacteria; D_3__Sphingomonadales;D_4__Sphingomonadaceae;D_5__Ellin6055; D_6__uncultured bacterium
SWA Inoculum Control	dd53650d9f68a7953e28279097c3fffc	3	D_0__Bacteria;D_1__Proteobacteria;D_2__Gammaproteobacteria; D_3__Xanthomonadales;D_4__Xanthomonadaceae; D_5__Stenotrophomonas;D_6__Stenotrophomonas rhizophila

Appendix B

Putative GNPS hits

The following tables contain library matches from GNPS after removal of control-associated compounds. All hits are putative and need further verification.

Table B.1: Putative GNPS hits for metabolites detected from environmental metabolomics of Fijian sediments from Chapter 3.

Compound Name	Parent Mass	Mass Difference	Cosine Score	MZErrorPPM	Site
Cilostazol	762.833	1.409	0.941728	1850.48	Site 4, Site 2, Site 1, Site 5
Spectral Match to Gln-Glu-Lys from NIST14	402.458	1.74899	0.935188	4326.97	Site 4, Site 3, Site 2, Site 1, Site 5
Spectral Match to Dioctyl sulfosuccinate from NIST14	467.513	0.30899	0.90752	661.361	Site 4, Site 3, Site 2, Site 5
Spectral Match to Arg-Asn from NIST14	288.275	0.875	0.857616	3026.11	Site 4, Site 3, Site 2, Site 1, Site 5
Pesticide5_Propargite_C19H26O4 S_Sulfurous acid, 2-[4-(1,1-dimethylethyl)phenoxy]cyclohexyl 2-propyn-1-yl ester	387.527	0.399994	0.842709	1033.24	Site 2
Spectral Match to Dioctyl phthalate from NIST14	391.57	0.287018	0.824687	733.53	Site 2
MLS002153841-01!Iobenguane sulfate	275.311	0.691986	0.809076	2507.17	Site 1
Bastimolide C	761.871	1.52905	0.801329	2002.95	Site 4
Methoxyhaemoventosine	691.849	0.721008	0.791631	1043.23	Site 4, Site 2
Benzalkonium chloride (C12)	304.533	0.242981	0.776506	798.518	Site 2
Massbank:AU247101 Hexa(methoxymethyl)melamine|2-N,2-N,4-N,4-N,6-N,6-N-hexakis(methoxymethyl)-1,3,5-triazine-2,4,6-triamine	391.455	0.224976	0.775782	575.047	Site 4, Site 3, Site 2, Site 1, Site 5
Benzalkonium chloride (C12)	304.502	0.212006	0.764798	696.722	Site 2

Table B.1: Putative GNPS hits for metabolites detected from environmental metabolomics of Fijian sediments from Chapter 3, Continued.

Compound Name	Parent Mass	Mass Difference	Cosine Score	MZErrorPPM	Site
Cer(d18:0/16:0); [M+H]+ C34H70N1O3	540.674	1.146	0.758374	2124.07	Site 1
"""MLS000860019-01!R()-2,10,11-Trihydroxyaporphine hydrobromide"""	283.132	1.00101	0.741831	3523.02	Site 3, Site 2, Site 1
Spectral Match to Inosine from NIST14	267.915	1.173	0.73805	4359.18	Site 4, Site 2
Spectral Match to Glycerol 1-myristate from NIST14	627.675	0.216003	0.737718	344.251	Site 2
Spectral Match to Dodecylbenzenesulfonic acid from NIST14	719.779	0.47699	0.735842	663.129	Site 4, Site 3
Massbank:PB002101 Oxycodone|Morphinan-6-one, 4,5alpha-epoxy-14-hydroxy-3-methoxy-17-methyl-	316.17	0.0150146	0.71559	47.4914	Site 3, Site 1, Site 5
Methoxyhaemoventosine	692.043	0.915039	0.704858	1323.98	Site 4, Site 3, Site 5

Table B.2: Putative GNPS hits for metabolites detected from environmental metabolomics of Mo'orean sediments from Chapter 4.

Compound Name	Parent Mass	Mass Difference	Cosine	MZ Error PPM	Site
Spectral Match to Undecaethylene glycol from NIST14	503.307	0.00701904	0.988005	13.946	Site_4, Site_2, Site_8, Site_3, Site_7, Site_1, Site_5, Site_6
Spectral Match to Tris(2-butoxyethyl) phosphate from NIST14	421.233	0.00100708	0.957642	2.3908	Site_2, Site_3, Site_1
Spectral Match to Decaethylene glycol from NIST14	459.28	0	0.950603	0	Site_4, Site_2, Site_8, Site_3, Site_7, Site_1, Site_5, Site_6
MS_Contaminant_Sodium_Formate_Cluster	838.84	0.00500488	0.928987	5.96647	Site_4, Site_2, Site_8, Site_3, Site_7, Site_1, Site_5, Site_6
Massbank:LU071403 N-[3-(Dimethylamino)propyl]dodecanamide	285.291	0.00097656	0.923025	3.42305	Site_2, Site_1, Site_5
Spectral Match to Lauric acid leelamide from NIST14	200.201	0	0.892685	0	Site_4, Site_7, Site_5
Spectral Match to Conjugated linoleic Acid (10E,12Z) from NIST14	281.247	0.00097656	0.853342	3.47225	Site_4, Site_8, Site_3, Site_1, Site_5, Site_6
Tryptophan conjugated cholic acid	617.351	0.00500488	0.843438	8.10696	Site_4, Site_8, Site_3, Site_7, Site_5
Spectral Match to Pinolenic acid from NIST14	279.232	0.00198364	0.799359	7.10397	Site_2, Site_8, Site_7, Site_1
C17-Sphinganine	288.29	0.00097656	0.776443	3.38742	Site_4, Site_2, Site_3, Site_1, Site_5, Site_6
Spectral Match to Hexanedioic acid, bis(2-ethylhexyl) ester from NIST14	371.319	0.00299072	0.754817	8.05439	Site_4, Site_2, Site_8, Site_7, Site_5
Spectral Match to cis-9-Hexadecenoic acid from NIST14	237.221	0.00100708	0.749348	4.24531	Site_3, Site_7

Appendix C

DIABLO Plots

DIABLO plots generated with the MixOmics package in R. Plots visualize results from sparse partial least squares discriminant analysis (sPLS-DA). Values in the lower triangular panels denote the Pearson's correlation coefficient while the upper triangular panel is the corresponding scatter plot.

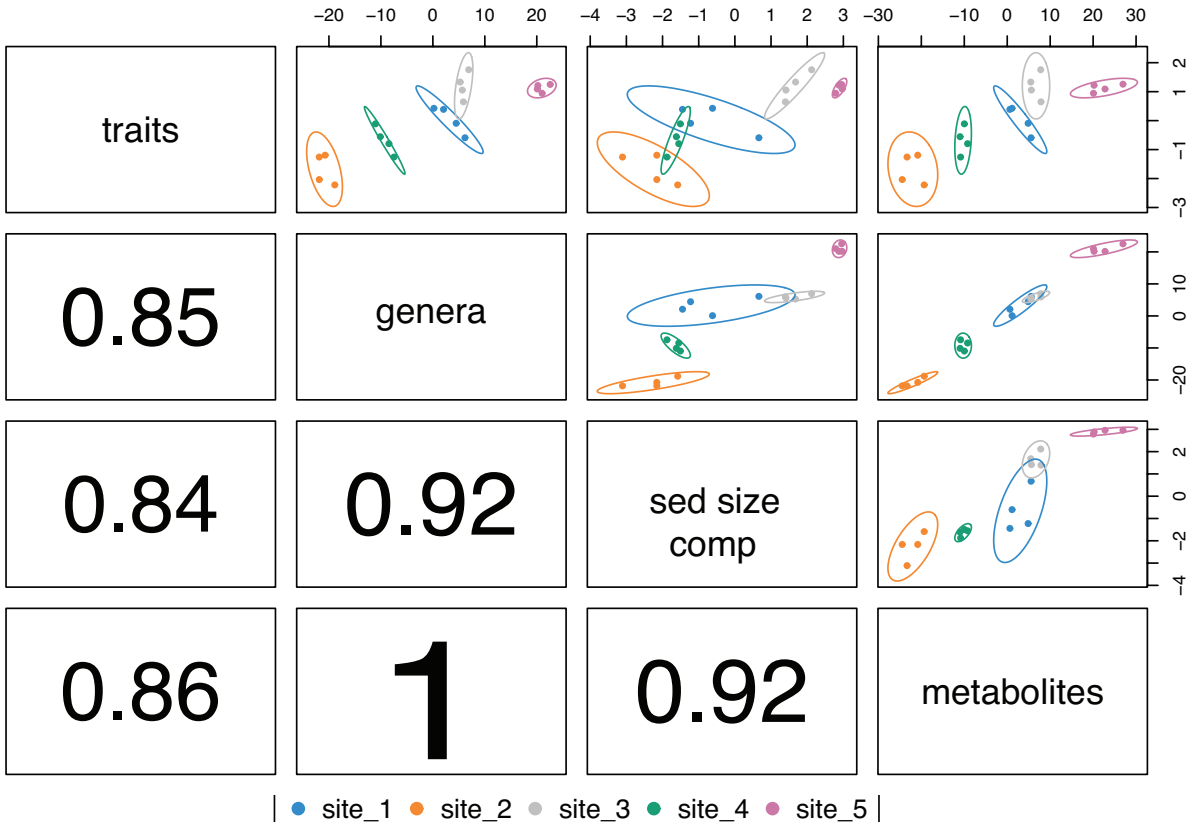

Figure C.1: Chapter 3 DIABLO plot with site as factor.

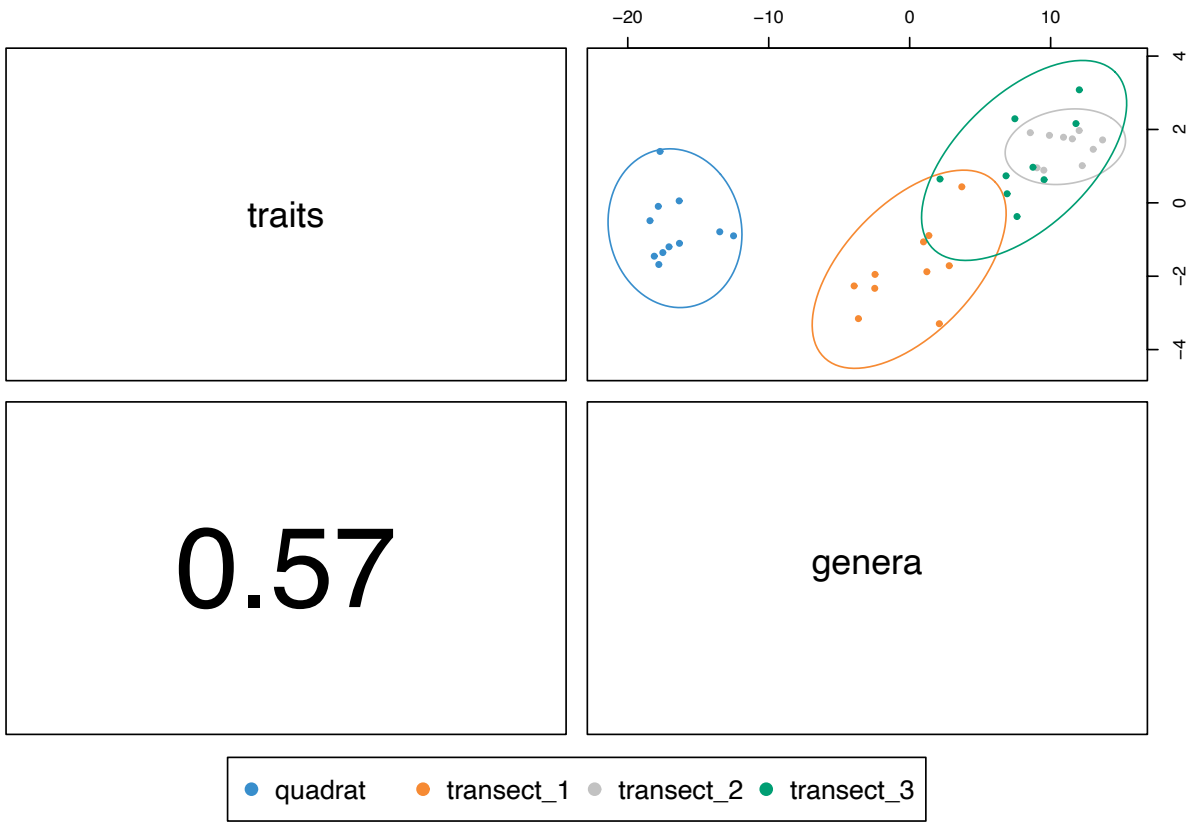

Figure C.2: Chapter 3 DIABLO plot based on site 4 samples. Sample type (quadrat, transect_1, transect_2, transect_3) as factor.

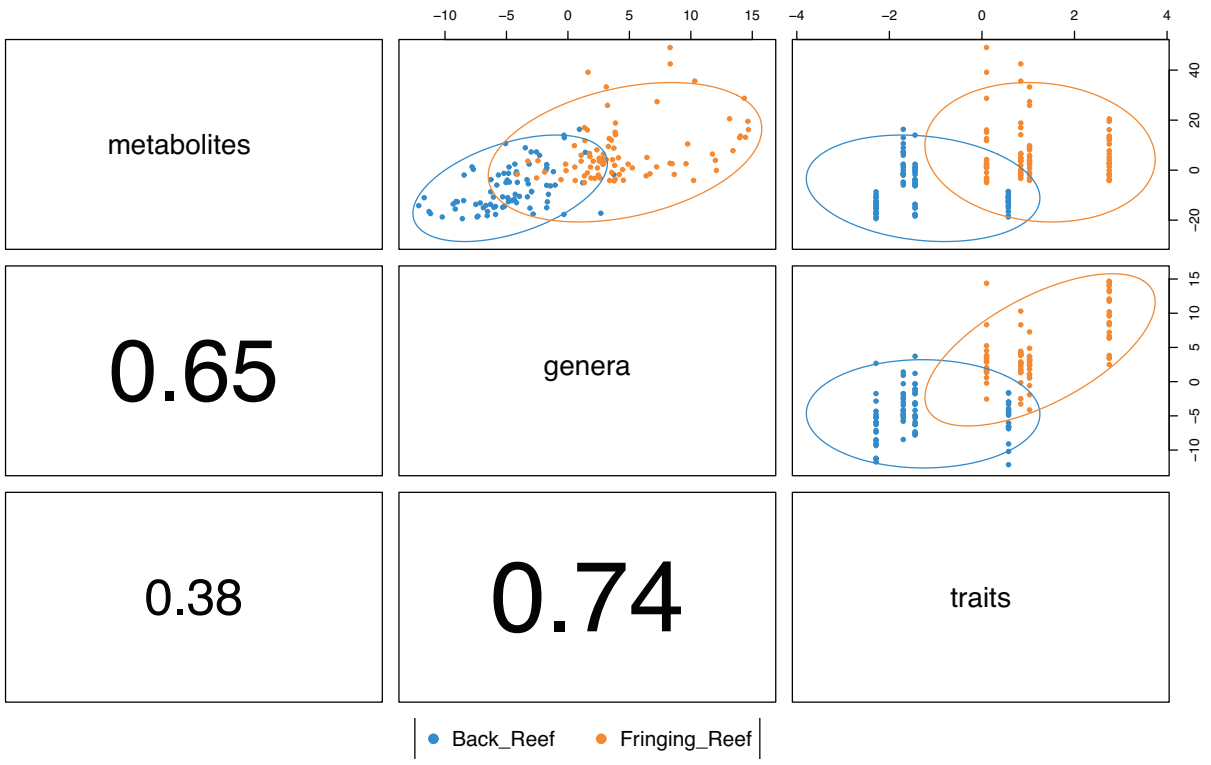

Figure C.3: Chapter 4 DIABLO plot with reef type as factor.

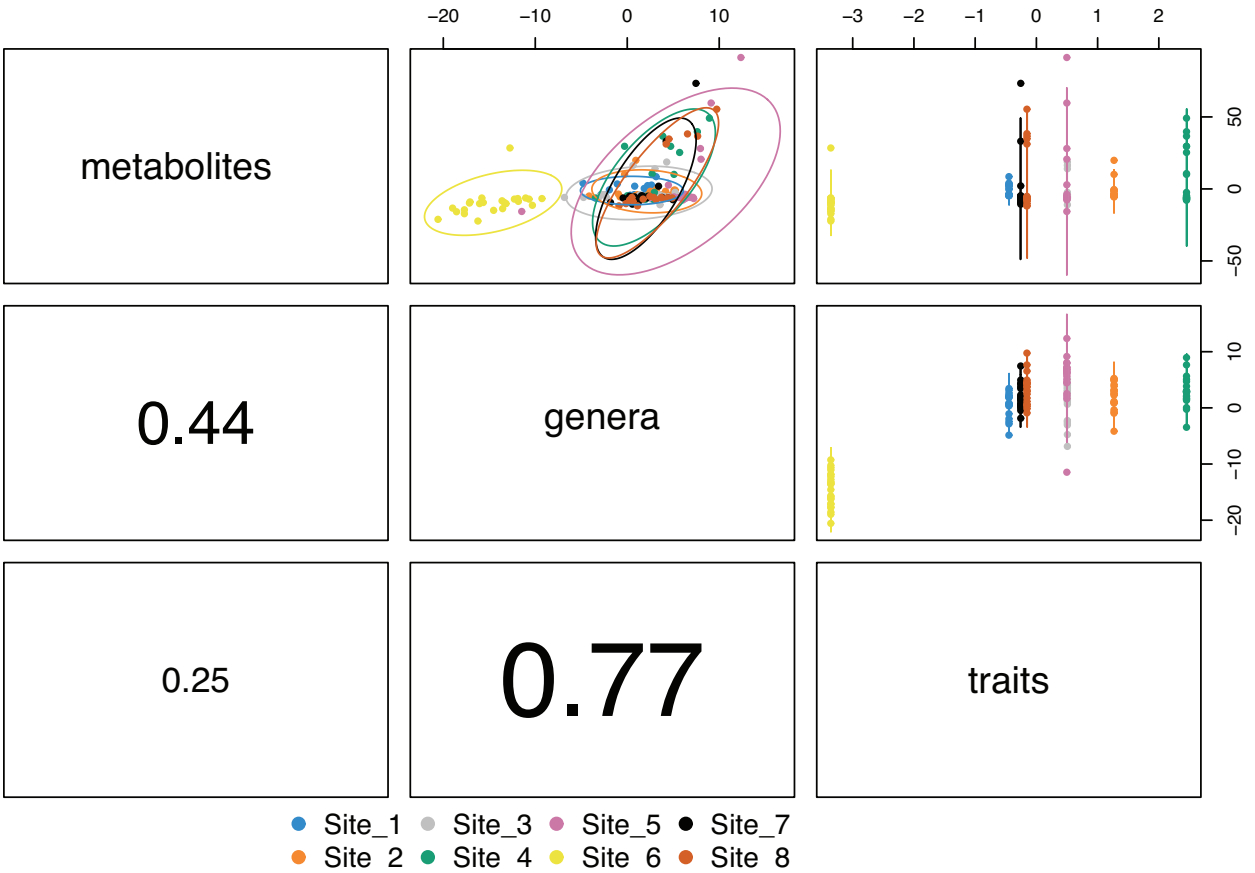

Figure C.4: Chapter 4 DIABLO plot with site as factor.