Characterizing the ecology of coral reef microorganisms across different scales within the Caribbean

by

Laura Weber
B. S., University of Maryland, 2012

to the Biology Department
Submitted in partial fulfillment of the requirements for the degree of
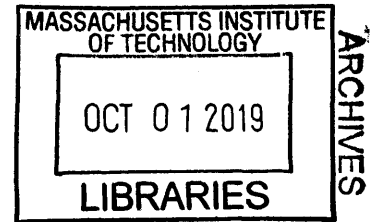Doctor of Philosophy
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
and the
WOODS HOLE OCEANOGRAPHIC INSTITUTION
February 2020

Signature of Author **Signature redacted**

Joint Program in Oceanography/Applied Ocean Science and Engineering
Massachusetts Institute of Technology
and Woods Hole Oceanographic Institution
September 25th, 2019

Certified by **Signature redacted**

Dr. Amy Apprill
Thesis Supervisor
Associate Scientist
Woods Hole Oceanographic Institution

Accepted by **Signature redacted**

Dr. Mick Follows
Chair, Joint Committee for Biological Oceanography
Massachusetts Institute of Technology

Characterizing the ecology of coral reef microorganisms across different scales within the
Caribbean
by
Laura Weber

Submitted to the MIT-WHOI Joint Program in Oceanography and Applied Ocean Science and
Engineering on September 25, 2019, in partial fulfillment of the requirements for the degree of Doctor
of Philosophy in Biological Oceanography

ABSTRACT

Microorganisms sustain the high productivity of coral reefs and support one of the most diverse, valuable, and threatened ecosystems on Earth. Despite the importance of reef microorganisms, there is a lack of understanding about their ecology, especially on Caribbean reefs. Furthermore, the hastening degradation of reefs due to anthropogenic stressors has made it difficult to understand natural patterns in microbial communities in the context of larger-scale ecosystem changes. Using genomics and metabolomics approaches paired with biogeochemical and physicochemical measurements as well as quantification of cell abundances, this dissertation provides optimized methods for studying the coral microbiome, investigates potential interactions between corals and seawater microorganisms, measures changes in the composition and diversity of reef seawater microorganisms over different spatial and temporal scales, and provides baseline information about microbial ecology, biogeochemistry, and metabolite compositions of a protected and relatively-healthy Cuban coral reef-system to fill these critical knowledge gaps. I found that coral species and reef location influenced the composition of bacteria and archaea within the seawater surrounding coral colonies and this seawater was enriched with microbial colonization and interaction genes, providing evidence of a distinct microbial environment surrounding corals named the coral ecosphere. In a separate study, diel and daily variation superseded spatial variation in terms of influencing shifts in the microbial community. At a larger scale, seawater microbial communities collected from the protected reef-system of Jardines de la Reina, Cuba had higher alpha diversity and community similarity, lower nutrient concentrations, and higher abundances of picocyanobacteria compared to less protected reef-systems within Los Canarreos, Cuba and the Florida Keys, U.S.A and seawater microbial communities collected from each reef-system were influenced by hydrogeography and environmental gradients. Lastly, the extracellular metabolite composition of reef seawater collected across Jardines de la Reina was highly similar, suggesting homogenous environmental and hydrogeographic conditions across these forereefs. Overall, this dissertation characterizes reef seawater microbial communities across different scales and provides novel, baseline information about a protected and understudied Cuban reef-system, offering critical information about the ecology of reef microorganisms within the Caribbean.

Thesis Supervisor: Dr. Amy Apprill
Title: Associate Scientist
Woods Hole Oceanographic Institution

ACKNOWLEDGEMENTS

During my experience as a graduate student within the WHOI-MIT Joint PhD program, I have spent many hours underwater collecting samples and in the lab processing samples, directly observed the humbling beauty and complexity of coral reef ecosystems, and worked alongside truly amazing mentors and collaborators. I have many people to thank for supporting me in this PhD journey and I am eternally grateful to these humans for the time, patience, and encouragement they have given me along the way.

First, I would like to thank my graduate advisor, Dr. Amy Apprill, for her steadfast support, guidance, trust, and wisdom. A number of years ago, Amy decided to take a chance on me and offered me a position as a guest investigator in her lab after I sent her an email out of the blue. A few months later, I arrived in Woods Hole and was immediately immersed into the world of microbiology and molecular biology, fascinated with using methods like DNA extraction, PCR amplification, sequencing, and microscopy to understand and in some cases visualize the microbial world. The rest is history – within a year I had applied and been admitted to the Joint PhD program. On the personal side, due to the nature of our work, Amy and I have spent a lot of time together in remote locations like Jardines de la Reina, Cuba and St. John, U.S. Virgin Islands. While each field experience is unique, I was always amazed by Amy's determination, energy, and enthusiasm while working in the field. As an advisor, Amy has always listened and had high yet reasonable expectations, allowing me time to maintain a balanced existence and giving me freedom to pursue other hobbies outside of my dissertation. I cannot thank her enough for taking me on as her first graduate student and I look forward to our future work together!

I would also like to thank the other members of my dissertation committee – Elizabeth Kujawinski, Stefan Sievert, and Greg Fournier – for invaluable feedback and guidance throughout my dissertation. I am very grateful to have such a supportive, communicative, and collaborative dissertation committee. Liz Kujawinski, thank you for your time, mentorship and for graciously inviting me into your lab and giving me the opportunity to learn metabolomics methodologies. Additionally, thank you to Colleen Hansel, another inspirational scientist who I have had the opportunity to work with in the field in Hawaii and Cuba, for chairing my dissertation defense.

I would like to thank my various co-authors – Emelia DeForce, Maickel Armenteros, Patricia Gonzalez-Diaz, Fernando Bretos, Erich Bartels, Victor Ferrer, and Alyson Santoro – for providing me useful comments and suggestions throughout the publication process.

I'd like to thank Karen Selph for training me how to run and analyze flow cytometry data, Krista Longnecker for advice and guidance in how to analyze the metabolite data, Melissa

Soule for running the metabolite samples, and Gretchen Swarr for teaching me the method of solid phase extraction and sample processing advice.

I would like to thank the other members of the Apprill lab – former and current – for their support and company. I would especially like to thank Dr. Carolyn Miller who has been my office buddy and confidant over the past six years. Our conversations and laughs have gotten me through stressful situations and your friendship is truly cherished. I would like to thank Cynthia Becker, my best field work buddy, for her companionship, determination, enthusiasm for singing along to music while we process samples, and expertise throughout our adventures together in the Virgin Islands and the Florida Keys. I look forward to many more song-filled field experiences together. I would like to thank Matt Neave, a former post-doc in the Apprill lab, who gave me a crash course in coding, taught me how to interact with computers through the command line, and introduced me to the wonders of bioinformatics.

A big thank you to the crews of the Felipe Poey, La Reina, and the M/V Alucia vessels for giving me the opportunity to access the remote reefs of Cuba. Thank you to Justin Suca, Ian Jones, Ashlee Lillis, Jessie Perelman, Alexis Earl, Jessica Dehn, Rod Catanach, Cathy Catanach, Joel Llopiz, Aram Mooney and the other members of the Coral Chorus team for allowing me to join these trips and participate in this important research! I will never forget the bond that was formed between us as we spent many sun-soaked and water-logged days collecting samples and deploying sound- and light-traps. I would also like to thank the staff of the Virgin Islands Environmental Resources Station (VIERS) for supporting the work of our team in St. John prior to the devastating Hurricanes of 2017.

To my friends within the MIT-WHOI JP and Woods Hole community, thank you for being my support system throughout this process. Jenny, Megan, Hanny, Mayra, Rene, Casey, Kalina, Cynthia, Tyler, and Deepa, I cherish the laughs and conversations we have had together and I wish you all the very best in your future endeavors. I will hold onto the memories we have made together - taking spin classes, baking delectable cakes and treats, playing board games, attending a wedding, traveling together, and taking long beach walks – but will hold out for more fun experiences with you all in the future.

To my loving family and my parents, Jane and Michael Weber, who have inspired me to pursue my dreams and who have always supported me. I would not be in this position today if my parents had not taken me to the shore every other summer. These early experiences instilled a deep love of and curiosity about the ocean and have led me to where I am today. Lastly, to my partner Henry, who has made the last couple years the best years of my life. Your smile, humor, and unwavering encouragement give me the strength and motivation to be the best scientist and human I can be. I look forward to our future land-bound and sea-faring adventures together and cannot wait to see where they lead.

This thesis is dedicated to my grandmother, Phyllis J. A. Weber, whose unwavering support and love during the formative years of my life helped me get where I am today.

TABLE OF CONTENTS

List of Tables

List of Figures

Chapter 1: Introduction

Chapter 2: Optimization of DNA extraction for advancing coral microbiota investigations

Chapter 3: The coral ecosphere: a unique coral reef habitat that fosters coral-microbial interactions

Chapter 4: Diel, daily, and spatial variation of coral reef seawater microbial communities

Chapter 5: Microbial signatures of protected and impacted Northern Caribbean reefs: changes from Cuba to the Florida Keys

Chapter 6: Extracellular reef metabolites across the protected Jardines de la Reina, Cuba reef-system

Chapter 1

Introduction

1.1 Background and motivation

Coral reefs are one of the most diverse and productive ecosystems on Earth (Odum and Odum 1955; Moberg and Folke 1999; Chavanich et al. 2015). They are essential for global ocean health and intricately tied to humans, providing a myriad of ecological services including coastline protection and sustaining up to 25% of marine fish species (reviewed within Moberg et al. 1999; Barbier et al. 2011; Brander and Van Beukering 2013; Lavides et al. 2016; Storlazzi et al. 2019). The immense productivity of coral reefs has always mystified scientists because measured concentrations of essential limiting nutrients like nitrogen and phosphorous in coral reef seawater are generally low, yet coral reefs support a tremendous diversity of organisms and high rates of primary productivity (Odum et al. 1955). The solution to this conundrum lies with the microorganisms living within reef animals, sediments, and seawater. Reef microorganisms drive nutrient cycling and productivity on coral reefs by retaining limiting nutrients like nitrogen and phosphorous and transferring benthic-derived organic matter into the water column and the reef sediment (Sorokin 1973; Atkinson 2010). Despite our general knowledge of the roles that these microorganisms occupy on reefs, there are many gaps in our fundamental understanding of the diversity and ecology of these microorganisms.

Over the past 35 years, the interactive threats of global climate change (e.g. elevated sea-surface temperatures, increased storm activity, ocean acidification), pollution, and overfishing have degraded these fragile and important ecosystems, leading to coral reefs dominated by macroalgae or other non-coral organisms (Done 1992; Norström et al. 2009), robbed of fish biomass (Valdivia et al. 2017), and plagued by coral mortality through bleaching and disease (Bourne et al. 2009, Porter et al. 2016). Amidst all of these changes, there is evidence that coral reef microbial communities have shifted in composition and function globally (Bruce et al. 2012;

Kelly et al. 2012; Kelly et al. 2014; Haas et al. 2016), potentially leading to further reef degradation.

As a result of these anthropogenic-induced changes, there is a pressing need to address fundamental questions about reef microbial community ecology from relatively healthy reefs to provide baseline information about the composition, diversity, function, and variations of these communities. The goals of this dissertation were to provide critical information about microbial, biogeochemical, and metabolite signatures of protected coral reefs, measure changes in the community ecology of reef seawater microorganisms over different spatial and temporal scales, and investigate potential interactions between corals and seawater microorganisms. An additional goal of this dissertation was to provide a standardized method for extracting microbial DNA from coral tissue, filling a community-based need to generate optimized protocols that overcome methodological issues associated with extracting DNA from coral biomass. The ultimate goal of this work was to provide essential information about reef microbial community ecology that can be integrated into conservation efforts to increase our chances of preserving and rehabilitating these fragile and important ecosystems.

In this introductory chapter, I will review important insights about the microbial ecology of coral reefs by discussing microbial contributions and dynamics within corals, reef sediment, and reef seawater. Following this review, I will also provide a discussion of the 'omics methods used to study microorganisms in this dissertation. The last section provides an overview of each dissertation chapter.

1.2 Coral-associated microbial communities

Reef-building scleractinian corals house dinoflagellates within their tissue that photosynthesize and translocate fixed carbon back to the coral host (Muscatine and Porter 1977). The coral can use this photosynthate for energy, respiring carbon dioxide and other waste products in the process (Muscatine et al. 1981; Falkowski et al. 1984). In addition to these symbiotic dinoflagellates, corals host diverse communities of bacteria, archaea, viruses, other protists, and fungi (reviewed within Thompson et al. 2014). These communities are considered instrumental to the health and growth of corals and likely play important roles in nutrient cycling within corals (Lesser et al. 2004; Raina et al. 2009; Ceh et al. 2013), pathogen defense through production of antibiotic compounds (Ritchie 2006; Krediet et al. 2013), and coral recruitment to the reef (Webster et al. 2004; Sneed et al. 2014). Sponges and macroalgae also populate the reef substrate and contain their own associated microbial communities (Barott et al. 2011; Thomas et al. 2016).

Coral-associated microbial communities are quite taxonomically diverse (Knowlton and Rohwer 2003; Sunagawa et al. 2010), but there is a large effort to understand which taxa are transient and influenced by local environmental conditions compared to taxa that are members of the core microbial community (Ainsworth et al. 2015; Hernandez-Agreda et al. 2018). This is because microbial taxa with core membership are more likely to be microbial symbionts of corals, playing important roles in coral health and growth. A number of specific bacterial taxa have been identified as potential coral symbionts based on their frequent detection in coral tissue. For example, *Endozoicomonas* spp. within the order Oceanospirillales (Class of Gammaproteobacteria) are frequently found in association with scleractinian corals as well as other marine invertebrates (Neave et al. 2016; Huggett and Apprill 2019). Genomic evidence suggests that *Endozoicomonas* bacteria exhibit a flexible lifestyle, perhaps giving them the ability to live both externally and internally within the coral host (Neave et al. 2016; Neave et al.

2017). Additionally, the presence of *Endozoicomonas* within coral tissue has been confirmed several times, providing concrete evidence of the residence of this putative coral symbiont within the host (Bayer et al. 2013; Neave et al. 2017). Lastly, *Endozoicomonas* is usually present and abundant in apparently healthy corals, further demonstrating that these cells may important roles in coral health and metabolism (Bayer et al. 2013; Meyer et al. 2014; Neave et al. 2017). Despite its presence and prevalence within coral tissue, the exact role and interactions of *Endozoicomonas* with the coral host are unknown to date. Bacteria identifying within the candidate genera "*Candidatus* Amoebophilus" have been frequently detected in coral tissue (Sunagawa et al. 2010; Apprill et al. 2016; Pollock et al. 2018; Huggett et al. 2019). Based on the lifestyle and genome of the first described species in this candidate genus, "*Candidatus* Amoebophilus asiaticus", "Ca Amoebophilus" taxa within coral tissue could be associated with other single-celled eukaryotic organisms within corals (Horn et al. 2001; Schmitz-Esser et al. 2010; Apprill et al. 2016). Other putative bacterial symbionts have been suggested for corals (Ainsworth et al. 2015; Hernandez-Agreda et al. 2018; Pollock et al. 2018), but very few studies have definitively investigated the interactions of these bacteria with coral hosts. However, the work conducted by Pollock et al. (2018) and Huggett and Apprill (2019) provide a solid platform for future coral-microbial symbioses research.

1.3 Reef sediment microbial communities

Reef sediment microbial communities recycle nutrients derived from particulate organic matter (e.g. coral mucus aggregates) that settles out of the water column, effectively serving as a filter for the reef by degrading this organic matter (Wild et al. 2004). Reef sediment microbial communities can also fix nitrogen or photosynthesize, depending on their location and oxygen

availability within the sediment (Werner et al. 2006; Werner et al. 2008). Reef sediment is typically carbonate- or silicate-based and this mineralogy can influence the taxonomic composition of microorganisms living with the sediment (Schottner et al. 2011) as well as their metabolic capacities (Wild et al. 2005). Additionally, the taxonomic composition of reef sediment microbial communities varies seasonally, geographically, and by depth within the sediment (Schottner et al. 2011). Generally, microbial community dynamics and metabolisms within reef sediment is not well understood, but these communities are likely critical for nutrient cycling on coral reefs.

1.4 Near coral seawater microbial communities and their potential interactions with corals

Sessile reef organisms, like corals, are continuously bathed in planktonic and particle-associated microorganisms living within the surrounding seawater. Multiple lines of evidence support the hypothesis that corals influence these seawater microbial communities. Corals release dissolved organic matter (DOM) that can be degraded by, and impact, the composition of planktonic microbial communities (Haas et al. 2013; Nelson et al. 2013). Specific metabolites commonly detected in coral mucus, including dimethylsulfoniopropionate (DMSP) and various amino acids (valine, lysine, threonine, aspartic acid, tryptophan, and casamino acids), have also been shown to initiate chemotactic responses in *Vibrio coralliilyticus* (a coral pathogen) (Garren et al. 2014) and natural assemblages of reef seawater bacteria (Tout et al. 2015). This evidence demonstrates that motile marine bacteria can detect and move towards gradients in coral metabolite concentrations. Coral colonies also interrupt water flow within the reef boundary layer, creating momentum boundary layers that are several centimeters thick surrounding individual coral colonies (Shashar et al. 1996). The hydrodynamics within these momentum

18

boundary layers may cause stagnation of flow and could have important influences on the advection of planktonic microorganisms close to the coral surface, but this has not been explored.

A number of field-based studies have investigated if seawater microbial communities surrounding corals are distinct in microbial composition. In a study conducted on the Great Barrier Reef, seawater microbial communities surrounding coral colonies were enriched with copiotrophic bacteria and bacterial motility and chemotaxis, membrane transport, and iron acquisition genes (in addition to other genes not listed here) when compared to surface reef seawater microbial communities (Tout et al. 2014). A study conducted in the Red Sea detected more coral-associated bacteria closer to coral colonies compared to further away in tandem with evidence of coral metabolite gradients surrounding each coral colony (Ochsenkuhn et al. 2018). Despite the above findings, a few field-based investigations did not observe differences in the seawater microbial composition or function surrounding corals compared to seawater sampled further away from the reef substrate (e. g. Silveira et al. 2017; Walsh et al. 2017), demonstrating that there is still much to be learned about potential coral – microbial interactions and the significance of these interactions for pathogen/symbiont recruitment and benthic-pelagic coupling on coral reefs. For example, how do environmental differences between reef locations impact coral-microbial dynamics surrounding individual coral colonies? Additionally, what is the variation of these near coral seawater microbial communities over temporal and spatial scales? These questions need to be addressed to understand the interactions between reef seawater microorganisms and corals, the engineers of coral reef ecosystems.

1.5 Reef seawater microbial communities

Smaller-scale interactions between corals and their surrounding seawater microorganisms occur within the context of reef seawater microbial community dynamics that transpire at the scale of the reef. As such, it is important to understand the influences and feedbacks in microbial community dynamics between these two spatial scales. Reef seawater microbial communities are typically comprised of autotrophic picocyanobacteria (primarily *Prochlorococcus* and *Synechoccocus*), autotrophic eukaryotic phytoplankton, heterotrophic bacteria, viruses, and heterotrophic protistan grazers. Autotrophic cells use the sun's energy to fix carbon and build biomass. A high proportion of this carbon is released as DOC from cells or as detrital matter through the process of sloppy feeding by heterotrophic protist and zooplankton grazing (reviewed by Thornton 2014). Heterotrophic bacteria respire carbon and recycle limiting nutrients like nitrogen and phosphorous, making these nutrients available for the rest of the microbial community (Azam and Malfatti 2007; Haas et al. 2013; Nakajima et al. 2017). Viruses lyse heterotrophic bacterial and eukaryotic cells, transferring carbon, nitrogen, phosphorous, and other limited micronutrients back into the dissolved phase via the viral shunt (Suttle 2007; Lara et al. 2017).

Additionally, benthic primary production carried out by symbiotic dinoflagellates living within coral tissue, algae, and cyanobacterial mats contributes to the overall primary production of coral reefs (reviewed within Atkinson 2010). Corals release a portion of dinoflagellate photosynthate as exudates and coral mucus into the surrounding seawater (reviewed within Crossland 1987; Edmunds and Davies 1989; Brown and Bythell 2005), and algae and cyanobacterial mats are direct sources of particulate and dissolved organic carbon to the reef seawater (Brocke et al. 2015; Mueller et al. 2016). Compared to benthic primary production, estimates of planktonic primary production pale in comparison, although these studies were only

conducted on two reefs within the Pacific Ocean and the Red Sea (Alldredge et al. 2013; Cardini et al. 2016). While planktonic primary production is small, the contribution to net community respiration by planktonic heterotrophic microorganisms is likely quite large, although there are no field studies that have actually demonstrated this in coral reef energy budgets. As such, respiration of DOC and POC as well as recycling of limited macronutrients by the heterotrophic microbial community retain nutrients within the coral reef food web and maintain high rates of primary productivity.

Accordingly, microbial community diversity, composition, and function generally reflect overall reef health. In fact, reef seawater microorganisms are sensitive to changes in environmental conditions and can be used as bioindicators of general reef health if their populations are monitored in concert with other biogeochemical and physicochemical measurements (Glasl et al. 2017). Seawater bacterial and archaeal communities sampled from relatively healthy coral reefs are mostly comprised of picocyanobacteria (*Prochlorococcus* and *Synechococcus*), Thaumarchaeota and Marine group II archaea within the Phylum Euryarchaeota, Alphaproteobacterial SAR11 sub-clades, and less abundant, but typically present Beta-, Gamma-, and Deltaproteobacteria, Bacteroidetes, Actinobacteria, and Firmicutes (Dinsdale et al. 2008; Nelson et al. 2011; Jeffries et al. 2015). This community shifts from net autotrophic to net heterotrophic on coral reefs that are subjected to nutrient disturbances and/or other anthropogenic impacts (Dinsdale et al. 2008; Moreira et al. 2015; Ziegler et al. 2016). Microorganisms that typically increase in abundance with nutrient shifts include more copiotrophic, heterotrophic Gammaproteobacteria such as *Alteromonas*, *Vibrio*, and *Pseudomonas*, and Alphaproteobacteria (e.g. *Ruegeria*, *Roseobacter*) (Moreira et al. 2015). Shifts in community metabolism affiliated with nutrient disturbances have also been linked to

enhanced detection of virulence-associated genes (Dinsdale et al. 2008; Kelly et al. 2012; Kelly et al. 2014; Moreira et al. 2015) and iron-acquisition genes (Kelly et al. 2012), suggesting that elevated nutrients select for dominance of copiotrophic taxa, increased competition, and enhanced expression of virulence. Additionally, it has been proposed that the process of microbialization, a shift in community metabolism that favors copiotrophic, heterotrophic bacteria, is occurring on reefs with high algae cover (Haas et al. 2016). In this positive feedback loop, named the dissolved organic carbon (DOC), disease, algae, microorganism (DDAM) model, algae release large quantities of dissolved organic matter that fuel heterotrophic microbial activity, resulting in a net draw down of DOC and the selection of copiotrophic and potentially pathogenic taxa (Haas et al. 2016).

That being said, shifts in microbial community composition, diversity, and function along various environmental and anthropogenic gradients as well as evidence for the microbialization hypothesis have not been extensively investigated on reefs within the Caribbean. Over the past 35 years, Caribbean reefs have suffered drastic losses in coral cover (Gardner et al. 2003) and reef rugosity (Alvarez-Filip et al. 2009) due to sea urchin die offs and overfishing that released algae from top-down grazing pressure, nutrient disturbances, as well as coral diseases (reviewed within Weil 2004). Many Caribbean reefs are also impacted substantially by human activities because they are located close to populated islands, accessible to fisherman and tourists, and influenced by coastal development. Altogether, these anthropogenic-induced stressors have led to the degradation of most Caribbean reefs, leaving only a few reef-systems that can provide baseline understanding of the microbial ecology of healthy Caribbean reefs. Obtaining baseline information is fundamental for understanding how microbial community diversity, composition, and function change along anthropogenic and environmental gradients as well as other metrics of

general reef health (e.g. reef composition, fish abundance). It is also essential for evaluating and testing the tenents of the microbialization hypothesis.

In the context of larger trophic and metabolic changes, seawater microbial communities also shift temporally in response to different environmental conditions or biological events (reviewed within Giovannoni and Vergin 2012; Teeling et al. 2012; Martin-Platero et al. 2018) . However, few studies have documented these changes in coral reef seawater microbial communities over time, making it difficult to discern normal diel, daily, or seasonal temporal variation from longer-term changes in reef microbial composition and metabolism. For example, an investigation of reef seawater microbial communities sampled from a reef flat and next to corals observed changes in community diversity over a diel cycle, but provided no insights into the taxonomic composition or direction of this change (Sweet et al. 2010). Diel and synchronous changes in reef seawater microbial community function and composition were observed on several Pacific reefs (Kelly et al. 2019), but this study only monitored populations over one diel cycle and noted daytime population increases of specific bacteria, *Psychrobacter* sp., that are not commonly detected in coral reef seawater. In contrast, no significant diel fluctuations occurred in seawater microbial communities collected from the momentum boundary layers surrounding the coral *Mussimilia braziliensis* or the above water column (Silveira et al. 2017), a surprising result considering the findings of the other two studies as well as the considerable metabolic changes that occur between day and night on reefs due to photosynthetic activity. In order to use microbial communities as reef bioindicators, there has to be an appreciation of normal variation and stochasticity in these communities over time as well as a way to integrate this variation into our assessment of larger shifts in reef community composition due to anthropogenic disturbances.

1.6 Techniques and methodologies for analyzing coral-associated and reef seawater microbial communities

The growing realization that microorganisms play essential roles in coral reef ecosystems (and any ecosystem) was definitively recognized in the 1970's and 80's with the use of microscopy, carbon-14 rate measurements, and thymidine incorporation rate estimates that made preliminary study of these populations possible (e. g. Sorokin 1973; Moriarty 1979; Moriarty et al. 1985). Additionally, there was a growing interest in studying the coral – dinoflagellate symbiosis, first chronicled in detail by Yonge et al. in 1932 (Yonge et al. 1932; Muscatine et al. 1977). Around the same time period, Carl Woese and George Fox used ribosomal RNA to discern the three domains of life (Woese and Fox 1977). This discovery led to the method of amplifying and sequencing conserved and hypervariable regions of the 16S rRNA gene of bacteria and archaea in order to identify these taxa within samples and compare the composition of microbial communities between different environments (Lane et al. 1985).

Ever since then, the development and advancement of methods for analyzing microbial populations using genomics as well as other 'omics methods (e. g. proteomics, metabolomics) has been at the forefront of the fields of microbial ecology and microbiology. In fact, it is common for one study to employ several of these 'omics methods to comprehensively document microbial community dynamics and generate hypotheses about the underlying mechanisms behind these dynamics. In the following paragraphs, I will briefly review the 'omics approaches that have been implemented in this dissertation.

Amplicon sequencing of targeted genes is a relatively fast and economical way to identify specific microorganisms as well as compare microbial community composition between

different environments or treatments to discover patterns within microbial communities. For this method, DNA is extracted and purified from biomass collected within a particular environment and polymerase chain reaction (PCR) amplification is used to amplify specific copies of marker genes (e.g. 16S rRNA genes of bacteria and archaea) that are subsequently sequenced. Following sequencing, the sequences/reads are filtered for quality, grouped together based on a particular level of similarity (e.g. operational taxonomic units, amplicon sequence variants, or minmum entropy decomposition (MED) nodes) with algorithms, and then matched to reference reads contained within a sequence database at a specific threshold to assign taxonomy to each unique sequence. The relative abundance or number of counts of each taxonomic group is then compared during microbial community analyses. While amplicon-based sequencing is fairly easy to implement, there are a number of obstacles, biases, and caveats that need to be considered. I will name a few below, but this list is not comprehensive and is offered to serve as an example of the major considerations for using this method.

Extracting microbial DNA can be difficult, especially for host-associated coral tissue microbial communities. DNA is extracted and purified from a coral tissue slurry (combination of mucus, tissue, and skeleton) or isolated coral tissue, and amplified during PCR to make many copies of a specific gene (e.g. 16S rRNA gene) in order to select for a particular fraction of the community for sequencing (e.g. bacteria and archaea in this example). It is particularly difficult to extract DNA from coral tissue for downstream analyses due to the presence of humic substances that inhibit PCR amplification (Bourne and Munn 2005; Sunagawa et al. 2009; Schrader et al. 2012) and the high proportion of mitochondrial coral and chloroplast DNA compared to bacterial and archaeal DNA that can lead to DNA swamping (Galkiewicz and Kellogg 2008; Feehery et al. 2013) and non-specific amplification during PCR (Galkiewicz et al.

2008; Klindworth et al. 2013). These obstacles lengthen the time needed to process and analyze microbial communities associated with coral tissue samples and often require researchers to develop their own DNA extraction and PCR amplification protocols based on their particular samples in order to generate useable data, thereby decreasing the ability to standardize methodology across different investigations. Compared to coral tissue, it is much easier to extract DNA from seawater microbial communities living within the euphotic zone.

The process of PCR may lead to biases in microbial community analyses downstream. During PCR, the primers used to select and copy the gene(s) of interest can be non-specific or fail to amplify all of the members of the community (e. g. Apprill et al. 2015). Standard PCR is not quantitative due to the inherent biases of the method as well as the fact that certain microbial cells may have more than one copy of a specific marker gene (reviewed within Vetrovsky and Baldrian 2013). The marker gene selected may not be representative of true ecological variability and the role of a microbe in a specific environment cannot be definitively inferred from matching its sequence to another sequence in a database.

There are other genomics and 'omics approaches that provide information about microbial community composition, but with more concrete insights about microbial functions and responses. Shotgun metagenomic sequencing, a method that sequences all of the genes within a sample, bypasses the biases introduced during PCR and provides a more complete picture of the potential function of the microbial community. Another genomics approach, metatranscriptomics, is used to understand which genes were actively transcribed by the organism at the time of cell death. As such, this approach gives a more definitive answer as to how microorganisms respond to environmental changes. Outside of genomics, the methods of

proteomics and metabolomics measure and identify the proteins or metabolites produced by microorganisms, respectively.

Metabolites are organic molecules produced by all living organisms. They constitute essential building blocks of proteins (amino acids) and DNA and RNA biomolecules (nucleosides and amino sugars), are involved in essential metabolic processes (metabolic intermediates and vitamins), and can even be used by microorganisms to communicate via quorum sensing (Rajamani et al. 2008; Johnson et al. 2016) as well as establish microbial symbioses (Dakora et al. 2015). Metabolomics is the study of metabolites using nuclear magnetic resonance spectroscopy (NMR) or mass spectrometry (MS) coupled with chromatography (Riekeberg and Powers 2017). Both approaches have trade-offs in terms of instrument sensitivity and the ionization potential of molecules (Riekeberg et al. 2017). Within MS-based metabolomics, metabolites within a sample can also be analyzed using targeted and untargeted approaches (Kido Soule et al. 2015). Untargeted metabolomics is a method that allows for semi-quantitative, multivariate comparison of metabolite composition across samples whereas targeted metabolomics involves detecting and quantifying concentrations of known compounds. Using both targeted and untargeted approaches, an investigation can query the overall metabolite composition and potentially identify new compounds of interest as well as obtain quantitative measurements of specific compounds within each sample.

The field of marine metabolomics is quite young compared to other fields with metabolomics applications (e.g. human medicine, plants) (reviewed by Kido Soule et al. 2015), but has tremendous potential to reveal the complexity of microbial interactions and exchange of organic matter within the ocean. For that matter, there is growing interest in the field of coral microbial ecology to apply metabolomics to understand the complexities and dynamics of

27

benthic-pelagic coupling, cycling of nutrients, and coral-algal interactions at a higher resolution

(Kelly et al. 2018). A handful of studies have applied intracellular metabolomics to understand

how corals respond to bleaching and ocean acidification (Sogin et al. 2016) or competitive

interactions with algae (Quinn et al. 2016). The coral tissue metabolome also corresponded with

the coral microbiome and *Symbiodinium* sp. composition, demonstrating intricate metabolic

connections between the coral host and its microbial community (Sogin et al. 2017). There is less

known about the extracellular metabolite composition of reef seawater, but it has been

demonstrated that reef sponges alter the composition of inhalant reef seawater and release

metabolites into the water through their exhalant current (Fiore et al. 2017). Another study found

evidence of a coral metabolite gradient in the seawater surrounding corals, suggesting that corals

may moderate and influence surrounding microbial communities via their metabolite exudates

(Ochsenkuhn et al. 2018). That being said, we have very little baseline knowledge about

identities, concentrations, distributions, and compositions of metabolites that can be found

extracellularly, constituting DOM, in coral reef seawater.

1.7 Dissertation overview

This dissertation seeks to address several voids in our understanding of the community ecology

of reef microorganisms. A theme of this dissertation is characterizing microbial (namely

bacterial and archaeal) community diversity, composition, and function across various spatial

and temporal scales as well as along environmental and anthropogenic gradients. Additionally,

this dissertation integrates various 'omics approaches (e. g. amplicon-based and shotgun

metagenomics; metabolomics) with bulk biogeochemical and physicochemical measurements as

well as reef composition surveys to comprehensively query how these parameters influence reef

microbial communities. An overview of the chapters included in this dissertation is presented in Figure 1.

Chapter two of this dissertation draws attention to the methodological limitations commonly encountered by investigations seeking to survey and characterize the microbial community composition of the coral microbiome. Through a series of methodological experiments, this chapter identifies a DNA extraction method that is best suited for extracting microbial DNA from seven different coral species and investigates if different extraction methods influence microbial community analyses downstream. Overall, this chapter satisfies a community need for a standardized and more optimal DNA extraction method that can help alleviate the costly obstacles encountered downstream of extractions as well as provides recommendations for future optimization studies for coral-associated microbial communities.

The remaining field-based chapters (3 – 6) mark a departure from the work completed in chapter two. These chapters catalogue reef seawater microbial community dynamics across reefs within the Caribbean. Chapters three and four investigate smaller-scale changes in microbial community composition, diversity, and function in the seawater adjacent to coral colonies whereas chapters five and six chronicle changes in microbial communities across reefs and reef-systems.

The work completed in chapter three was motivated by previous evidence that microorganisms can detect and grow off of coral exudates and that coral species can impact their surrounding seawater microbial communities (in some cases). However, these studies were limited to sampling one or two coral species at one reef and sampled seawater at large volumes, potentially mixing up the water column and entraining seawater from farther away into the samples (Tout et al. 2014; Silveira et al. 2017; Walsh et al. 2017). In this chapter, small-volume

(60 mL) near coral seawater samples were collected surrounding 5 Caribbean coral species across 10 reefs within the Jardines de la Reina and Canarreos reef-systems in Cuba. This work was completed to a) corroborate the findings that near coral seawater microbial communities are distinct in composition and function from reef seawater collected higher above the reef, (b) to tease apart how corals and environmental conditions at each reef site influence community composition, and (c) to determine if near coral microbial communities are species-specific. Broadly, this chapter seeks to understand how reef seawater microbial community dynamics are influenced by corals and what this could mean for pathogen or symbiont recruitment to corals as well as how reef microbial community dynamics within the seawater can ultimately impact coral health.

Building off of chapter three, chapter four addresses how reef microbial communities change over diel and daily time scales as well as in proximity to *Porites astreoides* coral colonies on a reef in St. John, U.S. Virgin Islands. There are few investigations that address consistent temporal fluctuations in reef seawater microbial communities or how sampling location (e.g. surface, reef-depth, near coral) can influence microbial community composition. As such, this study was designed to provide critical information about the natural variation of reef seawater microbial communities in a highly replicated diel timeseries that was conducted over the course of three days and nights.

Chapter five provides a comparative description of reef microbial community ecology at the scale of reef-system and along gradients of anthropogenic impact and environmental differences. Specifically, microbial community diversity, composition, and potential function as well as inorganic and organic macronutrient concentrations, microbial cell abundances, and net community respiration rates were assessed at reefs within the protected and remote Jardines de la

Reina reef-system and compared to these parameters surveyed and measured from reefs within

the Canarreos, Cuba reef-system as well as offshore and nearshore reefs within the Florida Keys,

USA. This chapter aims to identify signatures of Jardines de la Reina and compare these

signatures to reef-systems subjected to a higher degree of human impact, providing a necessary

baseline understanding of the microbial ecology of a relatively healthy and remote reef-system

within the Caribbean. Within this context, reefs were sampled over physicochemical and

hydrogeographical gradients within Jardines de la Reina and the Florida Keys (e.g. reefs located

offshore vs. nearshore) to obtain an understanding of the biogeochemical and microbial variation

within each reef-system. Lastly, this chapter investigates various aspects of the widely supported

microbialization hypothesis to provide further support for this hypothesis as well as its potential

nuances.

The work completed in chapter five served as motivation to return to the Jardines de la

Reina Archipelago in order to comprehensively survey the fore reefs within this reef-system. In

chapter six, untargeted and targeted environmental metabolomics approaches were applied, along

with biogeochemical and physicochemical measurements as well as reef composition surveys, to

examine the extracellular metabolite composition of surface and reef-depth seawater as well as

surface seawater from off-reef "bluewater" locations. Using these approaches, this chapter

provides the first comparative overview of the untargeted metabolite feature composition across

nine fore reefs in Jardines de la Reina, identifies and quantifies the environmental concentrations

of specific metabolites, and investigates trends between metabolite composition and microbial

biomass as well as inorganic and organic macronutrient concentrations. This work not only

provides novel information about the metabolite composition of reef seawater, thereby

illuminating the complexity and potential diversity of DOM cycling within reef ecosystems, but

also provides baseline information about a relatively healthy and understudied remote reef-system that can be used as a reference for future studies.

While each chapter of this dissertation independently addresses critical questions about the microbial ecology of coral reef ecosystems, the information gleaned from each chapter can be integrated to inform our holistic understanding of the complexity, intricacy, variation, and importance of microorganisms and their metabolisms in coral reef environments across various scales.

1.8 References

Ainsworth, T.D., Krause, L., Bridge, T., Torda, G., Raina, J.B., Zakrzewski, M. *et al.* (2015) The coral core microbiome identifies rare bacterial taxa as ubiquitous endosymbionts. *ISME J* 9: 2261-2274.

Alldredge, A., Carlson, C., and Carpenter, R. (2013) Sources of Organic Carbon to Coral Reef Flats. *Oceanography* 26: 108-113.

Alvarez-Filip, L., Dulvy, N.K., Gill, J.A., Cote, I.M., and Watkinson, A.R. (2009) Flattening of Caribbean coral reefs: region-wide declines in architectural complexity. *Proc Biol Sci* 276: 3019-3025.

Apprill, A., Weber, L.G., and Santoro, A.E. (2016) Distinguishing between microbial habitats unravels ecological complexity in coral microbiomes. *mSystems* 1: e00143-00116.

Apprill, A., Weber, L.G., and Santoro, A.E. (2016) Distinguishing between microbial habitats unravels ecological complexity in coral microbiomes. *mSystems* 1.

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.

Atkinson, M.J. (2010) Biogeochemistry of nutrients. In *Coral reefs an ecosystem in transition*. Dubinsky, Z., and Stambler, N. (eds): Springer Science & Business Media, pp. 199-206.

Azam, F., and Malfatti, F. (2007) Microbial structuring of marine ecosystems. *Nat Rev Microbiol* 5: 782-791.

Barbier, E.B., Hacker, S.D., Kennedy, C., Koch, E.W., Stier, A.C., and Silliman, B.R. (2011) The value of estuarine and coastal ecosystem services. *Ecological Monographs* 81: 169-193.

Barott, K.L., Rodriguez-Brito, B., Janouskovec, J., Marhaver, K.L., Smith, J.E., Keeling, P., and Rohwer, F.L. (2011) Microbial diversity associated with four functional groups of benthic reef algae and the reef-building coral Montastraea annularis. *Environ Microbiol* 13: 1192-1204.

Bayer, T., Neave, M.J., Alsheikh-Hussain, A., Aranda, M., Yum, L.K., Mincer, T. *et al.* (2013) The microbiome of the Red Sea coral Stylophora pistillata is dominated by tissue-associated Endozoicomonas bacteria. *Appl Environ Microbiol* 79: 4759-4762.

Bourne, D.G., and Munn, C.B. (2005) Diversity of bacteria associated with the coral Pocillopora damicornis from the Great Barrier Reef. *Environ Microbiol* 7: 1162-1174.

Brander, L.M., and Van Beukering, P. (2013) The total economic value of U.S. coral reefs: a review of the literature. In. Silver Spring, MD: NOAA Coral Reef Conservation Program, p. 32 p.

Brocke, H.J., Wenzhoefer, F., de Beer, D., Mueller, B., van Duyl, F.C., and Nugues, M.M. (2015) High dissolved organic carbon release by benthic cyanobacterial mats in a Caribbean reef ecosystem. *Sci Rep* **5**: 8852.

Brown, B.E., and Bythell, J.C. (2005) Perspectives on mucus secretion in reef corals. *Mar Ecol Prog Ser* **296**: 291-309.

Bruce, T., Meirelles, P.M., Garcia, G., Paranhos, R., Rezende, C.E., de Moura, R.L. *et al.* (2012) Abrolhos bank reef health evaluated by means of water quality, microbial diversity, benthic cover, and fish biomass data. *PLoS One* **7**: e36687.

Cardini, U., Bednarz, V.N., van Hoytema, N., Rovere, A., Naumann, M.S., Al-Rshaidat, M.M.D., and Wild, C. (2016) Budget of Primary Production and Dinitrogen Fixation in a Highly Seasonal Red Sea Coral Reef. *Ecosystems* **19**: 771-785.

Ceh, J., Kilburn, M.R., Cliff, J.B., Raina, J.-B., van Keulen, M., and Bourne, D.G. (2013) Nutrient cycling in early coral life stages:Pocillopora damicornislarvae provide their algal symbiont (Symbiodinium) with nitrogen acquired from bacterial associates. *Ecol Evol* **3**: 2393-2400.

Chavanich, S., Soong, K., Zvuloni, A., Rinkevich, B., and Alino, P. (2015) Conservation, management, and restoration of coral reefs. *Zoology (Jena)* **118**: 132-134.

Crossland, C.J. (1987) In situ release of mucus and DOC-lipid from the corals *Acropora variabilis* and *Stylophora pistillata* in different light regimes. *Coral Reefs* **6**: 35-42.

Dakora, F.D., Matiru, V.N., and Kanu, A.S. (2015) Rhizosphere ecology of lumichrome and riboflavin, two bacterial signal molecules eliciting developmental changes in plants. *Front Plant Sci* **6**: 700.

Dinsdale, E.A., Pantos, O., Smriga, S., Edwards, R.A., Angly, F., Wegley, L. *et al.* (2008) Microbial ecology of four coral atolls in the Northern Line Islands. *PLoS One* **3**: e1584.

Done, T.J. (1992) Phase shifts in coral reef communities and their ecological significance. *Hydrobiologia* **247**: 121-132.

Edmunds, P.J., and Davies, P.S. (1989) An energy budget for *Porites porites* (Scleractinia) growing in a stressed environment. *Coral Reefs* **8**: 37-43.

Falkowski, P.G., Dubinsky, Z., Muscatine, L., and Porter, J.W. (1984) Light and the bioenergetics of a symbiotic coral. *BioScience* **34**: 705 - 709.

Feehery, G.R., Yigit, E., Oyola, S.O., Langhorst, B.W., Schmidt, V.T., Stewart, F.J. *et al.* (2013) A method for selectively enriching microbial DNA from contaminating vertebrate host DNA. *PLoS One* **8**: e76096.

Fiore, C.L., Freeman, C.J., and Kujawinski, E.B. (2017) Sponge exhalent seawater contains a unique chemical profile of dissolved organic matter. *PeerJ* **5**: e2870.

Galkiewicz, J.P., and Kellogg, C.A. (2008) Cross-kingdom amplification using bacteria-specific primers: complications for studies of coral microbial ecology. *Appl Environ Microbiol* **74**: 7828-7831.

Gardner, T.A., Cote, I.M., Gill, J.A., Grant, A., and Watkinson, A.R. (2003) Long - term region wide declines in Caribbean corals. *Science* **301**: 958-960.

Garren, M., Son, K., Raina, J.B., Rusconi, R., Menolascina, F., Shapiro, O.H. *et al.* (2014) A bacterial pathogen uses dimethylsulfoniopropionate as a cue to target heat-stressed corals. *ISME J* **8**: 999-1007.

Giovannoni, S.J., and Vergin, K.L. (2012) Seasonality in ocean microbial communities. *Science* **335**: 671-676.

Glasl, B., Webster, N.S., and Bourne, D.G. (2017) Microbial indicators as a diagnostic tool for assessing water quality and climate stress in coral reef ecosystems. *Marine Biology* **164**.

Haas, A.F., Nelson, C.E., Rohwer, F., Wegley-Kelly, L., Quistad, S.D., Carlson, C.A. *et al.* (2013) Influence of coral and algal exudates on microbially mediated reef metabolism. *PeerJ* **1**: e108.

Haas, A.F., Fairoz, M.F., Kelly, L.W., Nelson, C.E., Dinsdale, E.A., Edwards, R.A. *et al.* (2016) Global microbialization of coral reefs. *Nat Microbiol* **1**: 16042.

Hernandez-Agreda, A., Leggat, W., Bongaerts, P., Herrera, C., and Ainsworth, T.D. (2018) Rethinking the coral microbiome: simplicity exists within a diverse microbial biosphere. *mBio* **9**: e00812-00818.

Horn, M., Harzenetter, M.D., Linner, T., Schmid, E.N., Müller, K.-D., Michel, R., and Wagner, M. (2001) Members of the Cytophaga-Flavobacterium-Bacteroides phylum as intracellular bacteria of acanthamoebae: proposal of '*Candidatus* Amoebophilus asiaticus'. *Environ Microbiol* **3**: 440-449.

Huggett, M.J., and Apprill, A. (2019) Coral microbiome database: Integration of sequences reveals high diversity and relatedness of coral-associated microbes. *Environ Microbiol Rep* **11**: 372-385.

Jeffries, T.C., Ostrowski, M., Williams, R.B., Xie, C., Jensen, R.M., Grzymski, J.J. *et al.* (2015) Spatially extensive microbial biogeography of the Indian Ocean provides insights into the unique community structure of a pristine coral atoll. *Sci Rep* **5**: 15383.

Johnson, W.M., Kido Soule, M.C., and Kujawinski, E.B. (2016) Evidence for quorum sensing and differential metabolite production by a marine bacterium in response to DMSP. *ISME J* **10**: 2304-2316.

Kelly, L.W., Haas, A., and Nelson, C.E. (2018) Ecosystem microbiology of coral reefs: linking genomic, metabolomic, and biogeochemical dynamics from animal symbioses to reefscape processes. *mSystems* **3**: e00162-00117.

Kelly, L.W., Barott, K.L., Dinsdale, E., Friedlander, A.M., Nosrat, B., Obura, D. *et al.* (2012) Black reefs: iron-induced phase shifts on coral reefs. *ISME J* 6: 638-649.

Kelly, L.W., Williams, G.J., Barott, K.L., Carlson, C.A., Dinsdale, E.A., Edwards, R.A. *et al.* (2014) Local genomic adaptation of coral reef-associated microbiomes to gradients of natural variability and anthropogenic stressors. *Proc Natl Acad Sci U S A* 111: 10227-10232.

Kelly, L.W., Nelson, C.E., Haas, A.F., Naliboff, D.S., Calhoun, S., Carlson, C.A. *et al.* (2019) Diel population and functional synchrony of microbial communities on coral reefs. *Nat Commun* 10: 1691.

Kido Soule, M.C., Longnecker, K., Johnson, W.M., and Kujawinski, E.B. (2015) Environmental metabolomics: Analytical strategies. *Marine Chemistry* 177: 374-387.

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., and Glockner, F.O. (2013) Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res* 41: e1.

Knowlton, N., and Rohwer, F. (2003) Multispecies microbial mutualisms on coral reefs: The host as a habitat. *Am Nat* 162: S51-S62.

Krediet, C.J., Ritchie, K.B., Alagely, A., and Teplitski, M. (2013) Members of native coral microbiota inhibit glycosidases and thwart colonization of coral mucus by an opportunistic pathogen. *ISME J* 7: 980-990.

Lane, D.J., Pace, B., Olsen, G.J., Stahl, D.A., Sogin, E.M., and Pace, N.R. (1985) Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *P Natl Acad Sci USA* 82: 6955-6959.

Lara, E., Vaque, D., Sa, E.L., Boras, J.A., Gomes, A., Borrull, E. *et al.* (2017) Unveiling the role and life strategies of viruses from the surface to the dark ocean. *Science Advances* 3: e1602565.

Lavides, M.N., Molina, E.P., de la Rosa, G.E., Jr., Mill, A.C., Rushton, S.P., Stead, S.M., and Polunin, N.V. (2016) Patterns of Coral-Reef Finfish Species Disappearances Inferred from Fishers' Knowledge in Global Epicentre of Marine Shorefish Diversity. *PLoS One* 11: e0155752.

Lesser, M.P., Mazel, C.H., Gorbunov, M.Y., and Falkowski, P.G. (2004) Discovery of symbiotic nitrogen-fixing cyanobacteria in corals. *Science* 305: 997-1000.

Martin-Platero, A.M., Cleary, B., Kauffman, K., Preheim, S.P., McGillicuddy, D.J., Alm, E.J., and Polz, M.F. (2018) High resolution time series reveals cohesive but short-lived communities in coastal plankton. *Nat Commun* 9: 266.

Meyer, J.L., Paul, V.J., and Teplitski, M. (2014) Community shifts in the surface microbiomes of the coral *Porites astreoides* with unusual lesions. *PLoS One* 9: e100316.

Moberg, F., and Folke, C. (1999) Ecological goods and services of coral reef ecosystems. *Ecological Economics* 29: 215-233.

Moreira, A.P., Meirelles, P.M., Santos Ede, O., Amado-Filho, G.M., Francini-Filho, R.B., Thompson, C.C., and Thompson, F.L. (2015) Turbulence-driven shifts in holobionts and planktonic microbial assemblages in St. Peter and St. Paul Archipelago, Mid-Atlantic Ridge, Brazil. *Front Microbiol* **6**: 1038.

Moriarty, D.J.W. (1979) Biomass of suspended bacteria over coral reefs. *Marine Biology* **53**: 193-200.

Moriarty, D.J.W., Pollard, P.C., and Hunt, W.G. (1985) Temporal and spatial variation in bacterial production in the water column over a coral reef. *Marine Biology* **85**: 285-292.

Mueller, B., den Haan, J., Visser, P.M., Vermeij, M.J., and van Duyl, F.C. (2016) Effect of light and nutrient availability on the release of dissolved organic carbon (DOC) by Caribbean turf algae. *Sci Rep* **6**: 23248.

Muscatine, L., and Porter, J.W. (1977) Reef corals: a mutualistic symbioses adapted to nutrient-poor environments. *BioScience* **27**: 454-460.

Muscatine, L., McCloskey, L.R., and Marian, R.E. (1981) Estimating the daily contribution of carbon from zooxanthellae to coral animal respiration. *Limnol Oceanogr* **26**: 601-611.

Nakajima, R., Tanaka, Y., Guillemette, R., and Kurihara, H. (2017) Effects of coral-derived organic matter on the growth of bacterioplankton and heterotrophic nanoflagellates. *Coral Reefs* **36**: 1171-1179.

Neave, M.J., Apprill, A., Ferrier-Pages, C., and Voolstra, C.R. (2016) Diversity and function of prevalent symbiotic marine bacteria in the genus *Endozoicomonas*. *Appl Microbiol Biotechnol* **100**: 8315-8324.

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-UK* **7**.

Neave, M.J., Rachmawati, R., Xun, L., Michell, C.T., Bourne, D.G., Apprill, A., and Voolstra, C.R. (2017) Differential specificity between closely related corals and abundant *Endozoicomonas* endosymbionts across global scales. *ISME J* **11**: 186-200.

Nelson, C.E., Alldredge, A.L., McCliment, E.A., Amaral-Zettler, L.A., and Carlson, C.A. (2011) Depleted dissolved organic carbon and distinct bacterial communities in the water column of a rapid-flushing coral reef ecosystem. *ISME J* **5**: 1374-1387.

Nelson, C.E., Goldberg, S.J., Wegley Kelly, L., Haas, A.F., Smith, J.E., Rohwer, F., and Carlson, C.A. (2013) Coral and macroalgal exudates vary in neutral sugar composition and differentially enrich reef bacterioplankton lineages. *ISME J* **7**: 962-979.

Norström, A.V., Nyström, M., Lokrantz, J., and Folke, C. (2009) Alternative states on coral reefs: beyond coral–macroalgal phase shifts. *Mar Ecol Prog Ser* **376**: 295-306.

Ochsenkuhn, M.A., Schmitt-Kopplin, P., Harir, M., and Amin, S.A. (2018) Coral metabolite gradients affect microbial community structures and act as a disease cue. *Commun Biol* **1**: 184.

Odum, H.T., and Odum, E.P. (1955) Trophic structure and productivity of a windward coral reef community on Eniwetok Atoll. *Ecological Monographs* **25**: 291-320.

Pollock, F.J., McMinds, R., Smith, S., Bourne, D.G., Willis, B.L., Medina, M. *et al.* (2018) Coral-associated bacteria demonstrate phylosymbiosis and cophylogeny. *Nat Commun* **9**: 4921.

Quinn, R.A., Vermeij, M.J., Hartmann, A.C., Galtier d'Auriac, I., Benler, S., Haas, A. *et al.* (2016) Metabolomics of reef benthic interactions reveals a bioactive lipid involved in coral defence. *Proc Biol Sci* **283**.

Raina, J.B., Tapiolas, D., Willis, B.L., and Bourne, D.G. (2009) Coral-associated bacteria and their tole in the biogeochemical cycling of sulfur. *Appl Environ Microb* **75**: 3492-3501.

Rajamani, S., Bauer, W.D., Robinson, J.B., Farrow, J.M., 3rd, Pesci, E.C., Teplitski, M. *et al.* (2008) The vitamin riboflavin and its derivative lumichrome activate the LasR bacterial quorum-sensing receptor. *Mol Plant Microbe Interact* **21**: 1184-1192.

Riekeberg, E., and Powers, R. (2017) New frontiers in metabolomics: from measurement to insight. *F1000Res* **6**: 1148.

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

Schmitz-Esser, S., Tischler, P., Arnold, R., Montanaro, J., Wagner, M., Rattei, T., and Horn, M. (2010) The genome of the amoeba symbiont "*Candidatus* Amoebophilus asiaticus" reveals common mechanisms for host cell interaction among amoeba-associated bacteria. *J Bacteriol* **192**: 1045-1057.

Schottner, S., Pfitzner, B., Grunke, S., Rasheed, M., Wild, C., and Ramette, A. (2011) Drivers of bacterial diversity dynamics in permeable carbonate and silicate coral reef sands from the Red Sea. *Environ Microbiol* **13**: 1815-1826.

Schrader, C., Schielke, A., Ellerbroek, L., and Johne, R. (2012) PCR inhibitors - occurrence, properties and removal. *J Appl Microbiol* **113**: 1014-1026.

Shashar, N., Kinane, S., Jokiel, P.L., and Patterson, M.R. (1996) Hydromechanical boundary layers over a coral reef. *J Exp Mar Biol Ecol* **199**: 17-28.

Silveira, C.B., Gregoracci, G.B., Coutinho, F.H., Silva, G.G.Z., Haggerty, J.M., de Oliveira, L.S. *et al.* (2017) Bacterial community associated with the reef coral *Mussismilia braziliensis's* momentum boundary layer over a diel cycle. *Front Microbiol* **8**: 784.

Sneed, J.M., Sharp, K.H., Ritchie, K.B., and Paul, V.J. (2014) The chemical cue tetrabromopyrrole from a biofilm bacterium induces settlement of multiple Caribbean corals. *Proc Biol Sci* **281**.

Sogin, E.M., Putnam, H.M., Anderson, P.E., and Gates, R.D. (2016) Metabolomic signatures of increases in temperature and ocean acidification from the reef-building coral, Pocillopora damicornis. *Metabolomics* **12**.

Sogin, E.M., Putnam, H.M., Nelson, C.E., Anderson, P., and Gates, R.D. (2017) Correspondence of coral holobiont metabolome with symbiotic bacteria, archaea and Symbiodinium communities. *Environ Microbiol Rep* **9**: 310-315.

Sorokin, Y.I. (1973) Trophical role of bacteria in the ecosystem of the coral reef. *Nature* **242**: 415-418.

Storlazzi, C.D., Reguero, B.G., Cole, A.D., Lowe, E., Shope, J.B., Gibbs, A.E. *et al.* (2019) Rigorously valuing the role of U.S. coral reefs in coastal hazard risk reduction. In. Reston, Virginia: U. S. Geological Survey open-file report, p. 42 p.

Sunagawa, S., Woodley, C.M., and Medina, M. (2010) Threatened corals provide underexplored microbial habitats. *PloS One* **5**.

Sunagawa, S., DeSantis, T.Z., Piceno, Y.M., Brodie, E.L., DeSalvo, M.K., Voolstra, C.R. *et al.* (2009) Bacterial diversity and white plague disease-associated community changes in the Caribbean coral *Montastraea faveolata*. *ISME J* **3**: 512-521.

Suttle, C.A. (2007) Marine viruses--major players in the global ecosystem. *Nat Rev Microbiol* **5**: 801-812.

Sweet, M.J., Croquer, A., and Bythell, J.C. (2010) Temporal and spatial patterns in waterborne bacterial communities of an island reef system. *Aquat Microb Ecol* **61**: 1-11.

Teeling, H., Fuchs, B.M., Becher, D., Klockow, C., Gardebrecht, A., Bennke, C.M. *et al.* (2012) Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. *Science* **336**: 608-611.

Thomas, T., Moitinho-Silva, L., Lurgi, M., Bjork, J.R., Easson, C., Astudillo-Garcia, C. *et al.* (2016) Diversity, structure and convergent evolution of the global sponge microbiome. *Nat Commun* **7**: 11870.

Thompson, J.R., Rivera, H.E., Closek, C.J., and Medina, M. (2014) Microbes in the coral holobiont: partners through evolution, development, and ecological interactions. *Front Cell Infect Microbiol* **4**: 176.

Thornton, D.C.O. (2014) Dissolved organic matter (DOM) release by phytoplankton in the contemporary and future ocean. *European Journal of Phycology* **49**: 20-46.

Tout, J., Jeffries, T.C., Webster, N.S., Stocker, R., Ralph, P.J., and Seymour, J.R. (2014) Variability in microbial community composition and function between different niches within a coral reef. *Microb Ecol* **67**: 540-552.

Tout, J., Jeffries, T.C., Petrou, K., Tyson, G.W., Webster, N.S., Garren, M. *et al.* (2015) Chemotaxis by natural populations of coral reef bacteria. *ISME J* **9**: 1764-1777.

Valdivia, A., Cox, C.E., and Bruno, J.F. (2017) Predatory fish depletion and recovery potential on Caribbean reefs. *Science Advances* **3**: e1601303.

Vetrovsky, 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**: e57923.

Walsh, K., Haggerty, J.M., Doane, M.P., Hansen, J.J., Morris, M.M., Moreira, A.P.B. *et al.* (2017) Aura-biomes are present in the water layer above coral reef benthic macro-organisms. *PeerJ* **5**: e3666.

Webster, N.S., Smith, L.D., Heyward, A.J., Watts, J.E., Webb, R.I., Blackall, L.L., and Negri, A.P. (2004) Metamorphosis of a scleractinian coral in response to microbial biofilms. *Appl Environ Microbiol* **70**: 1213-1221.

Weil, E. (2004) Coral reef diseases in the wider Caribbean. In *Coral Health and Disease*. Rosenberg, E., and Loya, Y. (eds). Berlin: Springer, pp. 35-68.

Werner, U., Blazejak, A., Bird, P., Eickert, G., Schoon, R., Abed, R.M.M. *et al.* (2008) Microbial photosynthesis in coral reef sediments (Heron Reef, Australia). *Estuarine, Coastal and Shelf Science* **76**: 876-888.

Werner, U., Bird, P., Wild, C., Ferdelman, T., Polerecky, L., Eickert, G. *et al.* (2006) Spatial patterns of aerobic and anaerobic mineralization rates and oxygen penetration dynamics in coral reef sediments. *Mar Ecol Prog Ser* **309**: 93-105.

Wild, C., Huettel, M., Klueter, A., Kremb, S.G., Rasheed, M.Y.M., and Jorgensen, B.B. (2004) Coral mucus functions as an energy carrier and particle trap in the reef ecosystem. *Nature* **428**: 66-70.

Wild, C., Rasheed, M., Jantzen, C., Cook, P., Struck, U., Huettel, M., and Boetius, A. (2005) Benthic metabolism and degradation of natural particulate organic matter in carbonate and silicate reef sands of the northern Red Sea. *Mar Ecol Prog Ser* **298**: 69-78.

Woese, C.R., and Fox, G.E. (1977) Phylogenetic structure of the porkaryotic domain: the primary kingdoms. *P Natl Acad Sci USA* **74**: 5088-5090.

Yonge, C.M., Yonge, M.J., and Nicholls, A.G. (1932) Studies on the physiology of corals. In *Scientific Reports / Great Barrier Reef Expedition 1928-29*, pp. 135 - 176.

Ziegler, M., Roik, A., Porter, A., Zubier, K., Mudarris, M.S., Ormond, R., and Voolstra, C.R. (2016) Coral microbial community dynamics in response to anthropogenic impacts near a major city in the central Red Sea. *Mar Pollut Bull* **105**: 629-640.

1.9 Figure
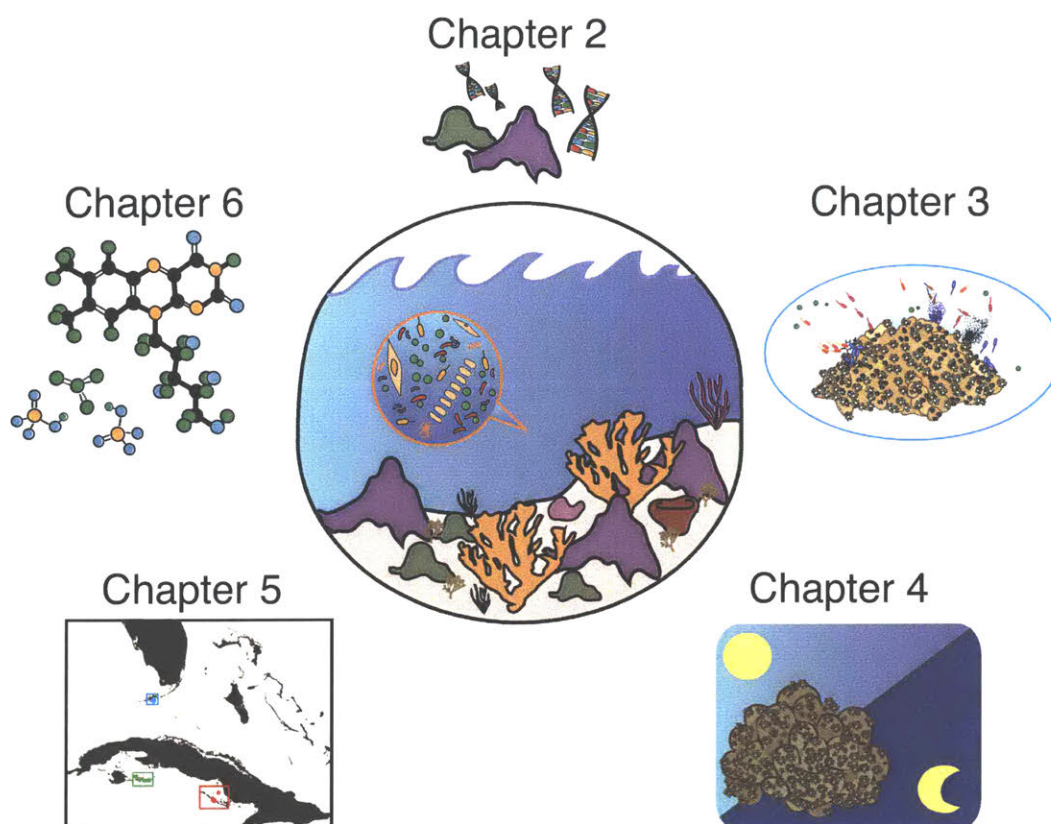
Figure 1. Overview of dissertation chapters.

Chapter 2

Optimization of DNA Extraction for Advancing Coral Microbiota Investigations

2.1 Abstract

DNA-based sequencing approaches are commonly used to identify microorganisms and their genes and document trends in microbial community diversity in environmental samples. However, extraction of microbial DNA from complex environmental samples like corals can be technically challenging and extraction methods may impart biases on microbial community structure. We designed a two-phase study in order to propose a comprehensive and efficient method for DNA extraction from microbial cells present in corals and investigate if extraction method influences microbial community composition. During phase I, total DNA was extracted from seven coral species in a replicated experimental design using four different MO BIO Laboratories, Inc. DNA Isolation kits: PowerSoil®, PowerPlant® Pro, PowerBiofilm®, and UltraClean® Tissue & Cells (with three homogenization permutations). Technical performance of the treatments was evaluated using DNA yield and amplification efficiency of small subunit ribosomal RNA (SSU rRNA) genes. During phase II, potential extraction biases were examined via microbial community analysis of SSU rRNA gene sequences amplified from the most successful DNA extraction treatments. In phase I of the study, the PowerSoil® and PowerPlant® Pro extracts contained low DNA concentrations, amplified poorly, and were not investigated further. Extracts from PowerBiofilm® and UltraClean® Tissue and Cells permutations were further investigated in phase II, and analysis of sequences demonstrated that overall microbial community composition was dictated by coral species and not extraction treatment. Finer pairwise comparisons of sequences obtained from *Orbicella faveolata, O. annularis,* and *Acropora humilis* corals revealed subtle differences in community composition between the treatments; PowerBiofilm®-associated sequences generally had higher microbial richness and the highest coverage of dominant microbial groups in comparison to the UltraClean® Tissue and Cells treatments, a result likely arising from using a combination of different beads during homogenization. Both the PowerBiofilm® and UltraClean® Tissue and Cells treatments are appropriate for large-scale analyses of coral microbiota. However, studies interested in detecting cryptic microbial members may benefit from using the PowerBiofilm® DNA treatment because of the likely enhanced lysis efficiency of microbial cells attributed to using a variety of beads during homogenization. Consideration of the methodology involved with microbial DNA extraction is particularly important for studies investigating complex host-associated microbiota.

2.2 Background

The coral holobiont (Rohwer et al. 2002; Knowlton and Rohwer 2003) consists of a network of interacting bacterial, archaeal, viral, fungal, protistan (i.e., *Symbiodinium* dinoflagellates), and coral cells (Thompson et al. 2014). While *Symbiodinium* are critical for providing carbon to the coral (Muscatine et al. 1981), bacteria and archaea may also play important roles by enhancing nutrient cycling (Lesser et al. 2004; Raina et al. 2009), inducing coral settlement (Sneed et al. 2014), and preventing coral diseases via production of antibiotic

44

compounds (Ritchie 2006; Krediet et al. 2013). The roles that bacteria and archaea may play in coral health and functioning have encouraged comprehensive investigations into the taxonomic identities and functional genes of microorganisms associated with globally distributed coral species. These studies have described widespread as well as health-related and ecologically important coral-microbial associations (Bourne et al. 2009; Vega Thurber et al. 2009; Apprill et al. 2013).

Cultivation-independent methods coupled with next generation sequencing technologies have been increasingly used to examine coral-microbial associations (Sunagawa et al. 2009; Vega Thurber et al. 2009; Mouchka et al. 2010; Apprill et al. 2013; Ainsworth et al. 2015)]. These methods rely on the extraction of nucleic acids (DNA and RNA) from environmental samples and are advantageous because they allow for the study of host-microbe interactions that are difficult to examine using cultivation-dependent methods (Thompson et al. 2014). The overall utility of these cultivation-independent approaches relies on the comprehensiveness of the extraction of nucleic acids from coral biomass. DNA extraction begins with a series of steps designed to rupture cells using chemical, enzymatic, physical, or mechanical means (Rudi et al. 2009). Investigators seeking to understand coral-associated microorganisms need to strive for representative lysis of morphologically diverse prokaryotic cells embedded within coral tissue (Apprill et al. 2009; Mouchka et al. 2010; Thompson et al. 2014; Ainsworth et al. 2015), and elution of high quality nucleic acids..

DNA extraction from coral biomass for investigation of associated bacteria and archaea is particularly subject to technical challenges and potential biases. Coral tissue is rife with polymerase chain reaction (PCR) inhibitors (Bourne and Munn 2005; Rudi et al. 2009; Sunagawa et al. 2009; Schrader et al. 2012), including humic acids and Ca^{2+} ions from the

residual coral skeleton. Co-elution of these inhibitors during extraction may decrease PCR efficiency and sensitivity, produce false-negative results (Schrader et al. 2012), delay investigations, and limit comparisons by decreasing sample size. In addition, lysis of microbial cells embedded within the matrix of larger eukaryotic coral cells (Apprill et al. 2009; Mouchka et al. 2010; Thompson et al. 2014; Ainsworth et al. 2015) may be particularly difficult to achieve because of the presence of the mesoglea, a supportive tissue layer comprised of strong collagen fibers that is sandwiched between the epidermal and gastrodermal coral tissue layers (Hausman and Burnett 1969). Moreover, lysis affinity for cells of a certain size or structure during sample homogenization could bias interpretation of microbial community composition from sequence-based data (Yuan et al. 2012; Wagner Mackenzie et al. 2015; Yang et al. 2015).

Differential lysis of cells during the extraction process may also compound biases associated with PCR amplification. For example, lysis methods with affinities for disrupting coral cells over microbial cells may increase the amount of eukaryotic DNA within the extract, therefore diluting the concentration of microbial relative to eukaryotic DNA. This swamping effect may reduce amplification of microbial DNA during PCR and decrease the overall efficiency of the reaction (Galkiewicz and Kellogg 2008; Feehery et al. 2013). In addition, non-specific amplification of more abundant chloroplast- and mitochondria-derived DNA from the eukaryotic cells by certain primers (Galkiewicz et al. 2008; Klindworth et al. 2013) may distort prokaryotic community structure and lead to exclusion of microbial groups found naturally associated with the coral (Sipos et al. 2007).

Commercial DNA extraction kits offer high-throughput and standardized protocols for streamlined sample processing. As such, using these kits minimizes technical variation and enables researchers to make meaningful comparisons between studies. In particular, kits

designed by MO BIO Laboratories, Inc. have been commonly used to extract DNA from coral

biomass for downstream analysis (Sunagawa et al. 2009; Sunagawa et al. 2010; Ceh et al. 2011;

Meron et al. 2011; Salerno et al. 2011; Morrow et al. 2012; Santos et al. 2012; Garcia et al. 2013;

Webster et al. 2013; Baker and Kellogg 2014; Kellogg et al. 2014; Ainsworth et al. 2015).

However, as described above, not all DNA extracts from coral biomass are amenable to PCR

amplification and this may be intensified for particular coral species. Several studies have

reported these methodological issues (Bourne et al. 2005; Vega Thurber et al. 2012; Kellogg et

al. 2014) and a few attempts have been made to optimize coral DNA extraction (Sekar et al.

2006; Lampert et al. 2008; Sunagawa et al. 2010; Santos et al. 2012; Baker et al. 2014). To date,

no large-scale studies have evaluated both the utility of and potential biases associated with

different DNA extraction treatments for extraction of DNA from disparate coral species.

In response, we conducted a two-phase experiment in order to 1) propose a

comprehensive and efficient method for extraction of microbial DNA from coral tissue and 2)

assess if DNA extraction treatment influences microbial community composition. Four

commercial DNA extraction kits and protocols supplied by MO BIO Laboratories, Inc.

(PowerSoil®, PowerPlant® Pro, PowerBiofilm®, and UltraClean® Tissue & Cells DNA

Isolation Kits) were used to extract DNA from seven different coral species during phase I of this

study. These kits were selected because they are commonly used to extract DNA from corals

(Baker et al. 2014) and each employ different combinations of chemical, enzymatic, and

mechanical disruption to lyse cells. DNA yield and microbial SSU rRNA gene amplification

efficiency were selected as initial screening parameters for phase I extractions because these

metrics are inexpensive and quantifiable indicators of DNA extraction success and amplification

amenability. The homogenization characteristics of the DNA extraction treatment that yielded

the highest average DNA concentrations and amplification efficiencies were further optimized

for DNA extraction. These extracts and the second highest performing extracts were then

subjected to SSU rRNA gene amplification and sequencing in phase II of this study to

investigate potential microbial community bias attributed to the different DNA extraction

methods.

2.3 Methods

Coral collection and processing

Coral fragments were collected by a scuba diver using a hammer and chisel during field

sampling trips to Kapangamarangi Atoll, Micronesia (November 2012), Florida Keys, U.S.A.

(May 2013), and Magnetic Bay, Australia (November 2013) (Additional file 1: Table 1).

Fragments were stored in a cooler containing ice until they were flash frozen in liquid nitrogen.

Fragments were obtained from 3 representative colonies of the following species: *Porites lobata*

(collected in Micronesia), *Pocillopora verrucosa* (Micronesia), *Acropora humilis* (Australia),

Orbicella faveolata, Montastraea cavernosa, O. annularis, and *Diploria strigosa* (Florida Keys).

Fragments were shipped back to Woods Hole Oceanographic Institution and stored at -80 °C

until they were processed.

Using an airbrush, an aerosolized jet of autoclaved 1X phosphate-buffered-saline (PBS)

was directed at freshly thawed coral fragments. This method physically separated the coral

mucus and tissue from the skeleton and suspended the cellular material in a slurry. The slurry

was homogenized and centrifuged at 4 °C for 20 minutes (5000 rpm) to form pellets comprised

of coral tissue and mucus. The PBS supernatant was removed and the tissue was evenly divided

into smaller sections using an ethanol-sterilized razor. To ensure that differing DNA yields were

solely attributed to the lysis efficiency of the extraction treatments, the amount of biomass

entering each extraction was standardized for all samples (38.7 ± 9.2 mg). Subsampled biomass

fractions were stored in separate tubes and frozen at -80 °C until they were extracted.

Phase I: DNA extractions

DNA was extracted from subsampled coral biomass using the PowerSoil® (cat # 12888),

PowerPlant® Pro (cat # 13400), PowerBiofilm® (cat # 24000), and UltraClean® Tissue & Cells

(cat # 12334) DNA Isolation kits following the manufacturer's protocols (MO BIO Laboratories,

Inc.) (Figure 1, Table 1). In this study, the treatments are referred to by their abbreviations:

PowerSoil® (PS), PowerPlant® Pro (PP), PowerBiofilm® (PB), and UltraClean® Tissue &

Cells (UC). In addition, manipulations to the mechanical lysis conditions for the UC extraction

were made, resulting in three permutations: Vortex Garnet (VG), Powerlyzer Glass (PG), and

Vortex Glass (VGl) (Table 1). The optional proteinase-K digestion step (15 µl; 20 mg/mL at 60°

C for 30 minutes) was implemented for all UC permutations. Samples were homogenized for 15

minutes using a vortex adaptor unless otherwise specified. Genomic DNA concentrations were

assessed using the dsDNA High Sensitivity Qubit 2.0 flourometric assay (Life Technologies).

After this study was conducted, MO BIO Laboratories merged with Qiagen and announced plans

to rebrand/discontinue some of their products as of January 1, 2017. To ease in this transition, we

have provided the original and new names for the kits used in this study: PS is the DNeasy

PowerSoil kit, PP is the DNeasy PowerPlant Pro kit, and PB is the DNeasy PowerBiofilm kit.

The UC kit has been discontinued.

DNA template was screened for PCR efficiency using the barcoded primer pair 515F and

806RB (Kozich et al. 2013; Apprill et al. 2015). PCR efficiency was determined for each species

X treatment pairing as the normalized percentage of successfully amplified amplicons of the

correct size (292 bp) out of all the extracts subjected to PCR. To assess PCR efficiency, unaltered DNA template (0.18 – 47 ng μl^{-1}) was amplified in 25 μl reactions containing 1.25 units of GoTaq Flexi DNA Polymerase (Promega), 5X Colorless GoTaq Flexi Buffer, 2.5 mM $MgCl_2$, 200 μM dNTP mix, and 200 nM of each barcoded primer in a thermocycler (Bio-Rad Laboratories). The following PCR reaction conditions were used: 95 °C for 2 minutes, followed by 40 cycles of 95 °C for 20 sec, 55 °C for 15 sec, and 72 °C for 5 minutes, concluding with an extension step of 72 °C for 10 minutes. PCR products were visually screened electrophoretically for quality using a 1% agarose and tris borate EDTA gel illuminated with ultraviolet light with the Hyperladder 50 bp DNA ladder (5 ng μl^{-1}) (Bioline). Positive amplification for each tested extract was denoted by the presence of a 292 bp sized band.

Mechanical lysis modifications

The extraction treatment that yielded extracts with the highest PCR efficiency for all coral colonies and species (UC) was selected to further examine if differences in bead type, homogenization method, and homogenization duration resulted in intra-treatment extraction biases (Figure 1, Table 1). The garnet beads provided with the UC kit were replaced with 0.1 mm glass beads (cat # 13118, MO BIO Laboratories, Inc.) (VGl). For the second modification, a PowerLyzer® 24 bench-top bead-based homogenizer (MO BIO Laboratories, Inc., cat # 13155) was used to homogenize the tissue instead of the vortex adaptor and garnet beads were replaced with 0.1 mm glass beads (PG). Samples were homogenized with the Powerlyzer for 45 seconds at 3500 rpm. DNA was not extracted from 3 of the 22 colonies (2 *P. verrucosa* and 1 *A. humilis)* using the VGl treatment because of limited biomass. DNA concentrations were quantified and PCR efficiency assessed using the methods outlined above.

Phase II: Library preparation and sequencing

Amplicons obtained from the PB, VG, PG, and VGI extractions were selected for sequencing based on overall DNA yield and PCR efficiency (Table 2). In addition, two positive DNA controls obtained from *E. coli* (Promega) and the Human Microbiome Project mock community DNA (BEI Resources, cat # HM-276D) and a negative control (U.V. sterilized DNA-free water) were amplified, barcoded, and included in the library pool. As an extra assessment of barcode reproducibility, each *O. annularis* extract was assigned two unique barcodes, amplified in separate reactions, and sequenced.

DNA template was amplified with the same V4 primer set using similar PCR reaction conditions to those described above, but with the number of cycles reduced to 35. Amplicons were purified using gel purification (MinElute PCR Purification Kit, Qiagen) so that only PCR products matching the approximate size of the V4 SSU rRNA gene amplicon were included in the final library pool. Samples were prepared for sequencing using the methods previously outlined by Apprill and colleagues (Apprill et al. 2015). The amplicon pool was shipped to the University of Illinois W. M. Keck Center for Comparative and Functional Genomics and sequenced using 2 x 250 bp paired-end MiSeq (Illumina) (Kozich et al. 2013; Apprill et al. 2015).

Sequence processing

Mothur software (Kozich et al. 2013) (v.1.33.3) was used to combine the de-multiplexed paired reads (8,344,281 contigs) and remove longer sequences (>275 bp) and sequences containing ambiguous base pairs. The expected length of the amplified region with the PCR-specific primers removed was 254 bp. A subset of longer sequences with read lengths exceeding 275bp were queried using the NCBI BLASTN 2.3.0 program (Zhang et al. 2000; Morgulis et al. 2008) to evaluate the taxonomic affiliation of these sequences. The remaining sequences were

classified using the SILVA SSU Ref database (Quast et al. 2013) (v. 119) and sequences

corresponding to Eukaryota, mitochondria, and 'unknown' lineages were discarded (2,802

sequences). Chloroplast sequences were retained to assess if more chloroplast sequences were

associated with a particular DNA extraction treatment. The UCHIME algorithm (Edgar et al.

2011) was used to identify and remove chimeric sequences (13,724 sequences total). Sequences

were not subsampled (de Carcer et al. 2011; McMurdie and Holmes 2014).

The sequences were grouped into nodes using the Minimum Entropy Decomposition

(MED) algorithm (Eren et al. 2013). These MED nodes are analogous to operational taxonomic

units (OTUs) and resolve biologically meaningful and distinct groups that can be separated by

<1% sequence disparities (Eren et al. 2013; Neave et al. 2017). Taxonomy was assigned to MED

nodes using the classify.seqs command in mothur (Kozich et al. 2013) and the SILVA database

(v. 119) (Quast et al. 2013). Sequences belonging to 'unclassified' MED nodes were re-aligned

using the SINA alignment service (Pruesse et al. 2012) (v. 1.2.11) and imported into ARB

(Ludwig et al. 2004) using the SILVA v. 123 database where phylogenetic comparisons were

made using neighbor joining algorithms to resolve 'unclassified' taxonomy. The mock

community and positive control DNA sample yielded the expected communities, replicate

barcoded samples produced repeatable results, and the negative control samples did not pass

quality control; these samples were then excluded from the analysis.

Statistical analysis

DNA concentrations were tested for normality using the Shapiro-Wilk test.

Concentrations were then subjected to one-way Analysis of Variance (ANOVA) or Friedman

Repeated Measures Analysis of Variance (FRMANOVA) on ranks tests, if data failed the

Shapiro-Wilk test, to assess if there were significant differences in mean DNA concentrations

between coral species or DNA extraction methods. When appropriate, Tukey's, Holm-Sidak, or Dunn's Method post-hoc tests were used to determine significantly different groups. P values $\leq$ 0.05 were accepted as being statistically significant. The above statistical tests were conducted using SigmaPlot software (v. 13).

Primer (v.7.0.9, Primer- E Ltd.) was used for a majority of the microbial community visualization and alpha diversity analysis. MED richness was calculated using the average number of unique MED nodes detected for each species x treatment grouping. MED species evenness was determined using the averaged Pielou's evenness index (J'). Bray-Curtis distances were calculated from normalized, square root transformed sequence data and used to conduct non-metric multidimensional scaling (nMDS) and nested two- and one-way analysis of similarity (ANOSIM) tests. Presence/absence heat maps of MED nodes detected from *O. faveolata*, *O. annularis*, and *A. humilis* associated amplicons were created using the phyloseq (McMurdie and Holmes 2013) R package and a custom script that was modified for this study. These heat maps were generated using distinct MED nodes that comprised 50% of all the reads obtained for each sample and thus represent the most dominant groups found within each colony. Frequency of MED node detection was determined for each treatment (and referred to as the top 50% MED coverage percentage) and the percentage of detection agreement between pairwise treatments within each colony was assessed. One-tailed t-tests were used to reveal significantly different MED node detection between treatments, and were conducted using SigmaPlot.

2.4 Results

Phase I: DNA yield and PCR efficiency

DNA concentrations varied among the extraction treatments (PS, PP, PB, VG, PG and

VGl) with PB yielding the highest average concentration of 12.53 ± 15.73 ng μl^{-1} (Figure 2,

Table 2) among all seven coral species. DNA concentrations obtained using the PB and VG

treatments had wider distributions than the other treatments, ranging from 0.22 - 46.40 and 0.18 -

32.80 ng μl^{-1}, respectively. Overall DNA yields from PG and VGl treatments were significantly

lower than yields from the PB, VG, and PS treatments (Figure 2, FRMANOVA, Df = 5, $p <$

0.001; Tukey Test, $p < 0.05$). Assessment of DNA yields by coral species revealed that PG and

VGl *P. lobata* extracts had significantly lower DNA yields than VG extracts (Table 2; one-way

FRMANOVA; Holm-Sidak Method, $p < 0.05$), but this trend was not observed for the other

species.

Gel screening was used to assess the efficiency of SSU rRNA gene amplification; VG

and VGl extracts had the highest species coverage and PCR efficiency, defined as the percentage

of extracts yielding visible and appropriately sized (~292 bp) bands in the gel (Table 2,

Additional file 1: Table 2). Similarly, efficiencies of PB and PG extracts were moderately high

(amplifying for 82% of all of the samples) and comparable with the VG and VGl extracts

(Additional file 1: Table 2). In contrast, efficiencies of PS and PP extracts were poor with the PS

extracts amplifying for 45% of samples and PP extracts only amplifying for 14% of all samples

(Additional file 1: Table 2). Non-specific priming, indicated by the presence of multiple larger or

smaller bands, was evident in a majority of the samples regardless of treatment. These

nonspecific bands (~200 and ~450 bp) were prominent in 58% and 52% of the PCR products

derived from PB and VG extracts, respectively.

During library preparation, 26% of the samples failed to amplify using the designated

temperature cycling conditions for the primers and 35 PCR cycles. Some of these samples may

have amplified at a higher number of PCR cycles or with dilution of the DNA template, but PCR optimization for every extract extended beyond the goals of this experiment. DNA extracts obtained from *D. strigosa* had the highest PCR failure rate of 75% in stark contrast to extracts from *O. faveolata* and *O. annularis* that had 100% PCR amplification success.

Phase II: Sequencing results

Amplicons obtained from the PB, VG, PG, and VGl treatments were prepared for sequencing of SSU rRNA genes in order to thoroughly assess the impact of DNA extraction treatment on microbial community composition. These amplicons were generated from 65 discrete coral colony and extraction treatment combinations representing all seven coral species. Regardless of treatment, there was a statistically significant disparity in the number of quality-filtered microbial sequences obtained from *P. lobata* and *P. verrucosa* corals in comparison to the other species (Figure 3). A majority of the reads obtained from *P. lobata* and *P. verrucosa* amplicons were too long and therefore were eliminated during preliminary quality-filtering. A subset of these longer reads corresponded to mitochondrial coral DNA sequences (NCBI accession numbers for top identities: JQ911534.1, e-value = 1e-102; EF597054.1, e-value = 1e-102; LN864762.1, e-value = 6e-101). *P. lobata* and *P. verrucosa* amplicons contributed a smaller proportion of sequences to the dataset in comparison to other species because of these disparities. In addition, all *M. cavernosa* PB extracts and *D. strigosa* VGl extracts had very low reads from the outset and were removed from analysis during quality-filtering.

Microbial community analysis of the SSU rRNA gene sequences demonstrated that, on a large scale, microbial community composition was significantly influenced by coral species and not DNA extraction treatment (Figure 4A, two-way nested ANOSIM, 7 coral species (B) nested within 4 extraction treatments (A), for A: R = -0.059, p = 0.798, for B: R = 0.684, p = 0.001).

Independent analysis of *O. faveolata*, *O. annularis*, and *A. humilis* amplicons (Figure 4B, C, D) revealed that microbial community composition was regulated by the coral colony used in the extraction (two-way nested ANOSIM, coral colonies (B) nested within 4 extraction treatments (A), for A: R = 0.013, p = 0.267, for B: R = 0.9, p = 0.001) rather than DNA extraction treatment. Additional testing within each coral species confirmed this observation that DNA extraction method did not significantly influence microbial community composition (one-way ANOSIM, R= 0.022, p = 0.364, *O. faveolata*; R = 0.065, p = 0.239, *O. annularis*; R = -0.044, p = 0.512, *A. humilis*). Non-metric multidimensional scaling analysis further supported this result, demonstrating that the same coral colonies clustered together regardless of the extraction treatment (Figure 4B, C, D). In-depth microbial community analysis was not possible for all species and treatments because of PCR inhibition and sequence disparities (Table 3).

Overall, no statistically significant differences in MED node richness were detected for a majority of the treatments, with the exception of amplicons from *P. verrucosa* (df = 4, one-tailed t-test, p = 0.03) and *A. humilis* (df = 2, one-way ANOVA, p = 0.023) (See Table 3). For *A. humilis*, Holm-Sidak method tests revealed higher MED node richness in PB compared to VG (p = 0.039) and VGI (p = 0.041) amplicons. Generally, the average number of MED nodes was lower for *P. lobata*, *P. verrucosa*, and *A. humilis* amplicons in comparison to the other species. Furthermore, amplicons generated from the PB treatment were more likely to have the highest MED richness out of the 4 treatments, with this being the case for five of the seven species (Table 3). MED species evenness (J') tended to be higher in all treatments except the PB treatment (Table 3). Overall, *D. strigosa* had the lowest evenness whereas *O. annularis* and *P. verrucosa* had the highest evenness.

To tease apart the pattern between MED richness and DNA extraction method on a finer

scale, pairwise investigations of differences between treatments were completed using

presence/absence analysis of MED nodes for *O. faveolata*, *O. annularis*, and *A. humilis* (Figure

5). Overall, there was fairly high MED node detection agreement between the sequences

generated from different treatments extracted from the same coral colony (average agreement

ranged from 72-85%). Closer inspection revealed that *O. faveolata* PB sequences had

significantly higher MED presence/absence coverage of dominant microbial groups compared to

VGl (one-tailed, paired t-test, $p = 0.013$) and VG ($p = 0.007$) sequences. Including technical

replicates, PB sequences from *O. annularis* had significantly higher MED coverage compared to

sequences from the PG ($p = 2.0 \times 10^{-4}$) and VGl ($p = 0.019$) treatments. A significant difference

in MED coverage was revealed between PG and VGl amplicons ($p = 0.004$) for *O. annularis*, but

this trend was not observed in *O. faveolata* or *A. humilis*. Sequences generated from *O. faveolata*

PB extracts contained more "*Candidatus* Branchiomonas" (MED node 4516) and

Thaumarchaeota (MED node 4459) reads in comparison to other treatments from this species.

Similarly, *Ca.* Branchiomonas (MED node 4517) was identified in more PB treatment sequences

from *O. annularis* compared to the other treatments. MED node presence/absence agreement in

O. annularis technical replicates was very high with only a few occurrences of disagreement

between 3 technical replicates (9 out of 182 possible disagreements). Statistically significant

differences in MED coverage of the dominant groups between treatments were not detected in

sequences obtained from *A. humilis* amplicons, but sequences from the PB treatment had the

highest coverage of dominant MED nodes (78%) out of all the treatments for this species.

2.5 Discussion

In this study, the PB and all variations of the UC (PG, VGl, VG) treatments were found to be technically suitable and reliable for extraction of microbial DNA from most colonies of *P. lobata*, *P. verrucosa*, *A. humilis*, *O. faveolata*, *O. annularis,* and *D. strigosa*. PCR inhibition during library preparation and significant removal of sequences during quality-filtering prevented highly resolved comparisons for *P. lobata*, *P. verrucosa*, *M. cavernosa*, and *D. strigosa*, demonstrating the importance of including many biological replicates for each species in sequencing-based investigations. Broadly, extraction treatment did not significantly bias microbial community composition, but finer scale investigations for *O. faveolata*, *O. annularis*, and *A. humilis* revealed minor differences in MED coverage and group sensitivity between the UC and PB derived communities.

Generally, all treatments yielded DNA concentrations that fell within the range of previously reported DNA concentrations for corals (Santos et al. 2012; Baker et al. 2014). While the VG and PB treatments yielded similar DNA concentrations, the PG and VGl treatments yielded extracts with lower DNA concentrations, an observation likely attributed to differences in the duration of mechanical lysis and bead size. In this study, significantly higher DNA yields were obtained from treatments that homogenized samples for 15 minutes (VG, PB) on a vortex adaptor in comparison to 45 seconds (PG) using a powerlyzer and this result aligns with the reported observation that DNA concentration increases with bead-beating duration (Henderson et al. 2013). Furthermore, larger beads are more likely to lyse eukaryotic coral cells and release more DNA (~420 Mbp coral genome[-1] (Shinzato et al. 2014)) whereas smaller beads are probably targeting the smaller microbial cells containing less DNA (~0.9-9.7 Mbp microbial genome[-1] (Koonin and Wolf 2008)). In this study, lower concentrations of DNA may have been obtained because the beads were either too large in diameter to effectively disrupt coral and

microbial cells (PP) or so small that they could not sufficiently rupture eukaryotic cells (PG, VGl). Lastly, the PB and UC treatments had less sample transfer steps (2) than the PS (4) and PP (3) kits. Minimizing steps during extraction likely helps maintain nucleic acid integrity and may also limit introduction of contaminants, reduce waste, and decrease extraction duration.

The UC (VG, PG, and VGl) and PB treatments yielded extracts that had similar PCR efficiencies. This may be because the UC and PB treatments physically lysed cells using high heat exposure (Middleberg 1995). Using an additional method to achieve cellular lysis may have increased the chance of disrupting cells from a wider variety of microorganisms and the overall concentration of prokaryotic DNA relative to eukaryotic DNA within the extraction. More importantly, the 100% amplification success of *O. annularis* UC and PB extracts in this study contrasts with the poor amplification (20-60%) reported for this species in a recent comparative DNA extraction optimization study using the PS and PP DNA treatments (Baker et al. 2014), and marks a promising advance in defining a suitable extraction method for this species.

PCR inhibition associated with particular coral species (*D. strigosa*) or colonies (*M. cavernosa*) may have arisen due to differences in PCR inhibitor carryover during initial sample processing. For example, we found that the calcium carbonate skeleton of *D. strigosa* colonies fractured more during sample processing in comparison with other species. This likely resulted in co-elution of calcium (Ca^{2+}) ions with DNA during the final step of the extraction. Because Ca^{2+} ions compete with magnesium (Mg^{2+}) ions as cofactors for DNA polymerase, higher concentrations of Ca^{2+} in *D. strigosa* extracts could have resulted in greater inhibition of DNA polymerase (Bickley et al. 1996; Al-Soud and Radstrom 2001). Baker & Kellogg (2014) also offered this hypothesis to explain differential PCR amplification between coral species and emphasized the importance of using multiple coral species for optimization experiments (Baker

et al. 2014). For future experiments, it may be appropriate to increase the Mg^{2+} concentration used during PCR to overcome this inhibition (Bickley et al. 1996; Al-Soud et al. 2001).

Unfortunately, neither DNA concentration nor PCR efficiency alone serve as definite indicators of sequence data quality, a concept not demonstrated in past coral DNA optimization studies (Santos et al. 2012; Baker et al. 2014), but supported by previous coral microbiota sequencing studies (Apprill et al. 2013; Ziegler et al. 2016) and the disparities between DNA concentration, PCR efficiency, and *P. lobata* and *P. verrucosa* sequence quality reported in this study. This observation can possibly be explained by the idea that extracts from *P. lobata* and *P. verrucosa* may be prone to more eukaryotic DNA swamping (Baker et al. 2014). Recent efforts for circumventing DNA swamping and non-specific amplification involve selectively enriching genomic extracts for prokaryotic DNA (Feehery et al. 2013) or designing new PCR primers (Mori et al. 2014). Alternatively, as the cost of sequencing declines, deep sequencing of shot-gun metagenomic DNA has increasingly been used to circumvent the issues associated with amplicon-based methods (Poretsky et al. 2014). While this approach may work well for some study systems (He et al. 2013; Nesme et al. 2014; Sunagawa et al. 2015), it proves difficult to use for studying complex communities within the coral holobiont; the abundance of coral and *Symbiodinium* genomic material requires deep sequencing and even size-fractionation may not effectively enrich the target communities of interest (Vega Thurber et al. 2009; Littman et al. 2011).

Microbial community analysis revealed that most of the variation in microbial community composition corresponded with coral species or colony rather than DNA extraction method. This agrees with the results of a human microbiome study that attributed most of the variation in microbial community composition to different human subjects rather than DNA

extraction method (Wagner Mackenzie et al. 2015). As a whole, this result suggests that the chosen DNA extraction method (either the PowerBiofilm® or the different permutations of the UltraClean® Tissue and Cells DNA Isolation kits) should not impart significant biases on microbial community composition if the aim of the study is to elucidate large differences in microbial community composition that correspond with changes in coral health, coral species, or other factors. Because many investigations are interested in making these larger comparisons, we recommend that both the PowerBiofilm® or the different permutations of the UltraClean® Tissue and Cells DNA Isolation kits are suitable for broad investigations of coral microbial dynamics.

This recommendation is verified by the finding that the dominant taxonomic classes of bacteria and archaea recovered in this dataset support the results of other coral microbiota taxonomic surveys. A recent review from Bourne and colleagues (Bourne et al. 2016) identified Gammaproteobacteria, Alphaproteobacteria, Actinobacteria, Bacteroidetes (esp. Flavobacteria), and Cyanobacteria as common coral-associated bacteria and all these groups were detected in this study. At a finer scale, we detected bacterial genera that have been previously identified as coral-associates. For example, in this study, MED nodes identified as *Endozoicomonas* (class Gammaproteobacteria) were present in all, but two of the samples (varying relative abundances of 0.002 – 80.02 %). *Endozoicomonas* bacteria are recognized as potentially important tissue and mucus-associates of corals (Apprill et al. 2016; Bourne et al. 2016; Neave et al. 2017), and their genomes suggest functional adaptations for residing with a host (Neave et al. 2016; Neave et al. 2017). *Ralstonia* spp. have also been detected in coral microbiota surveys of many different species (Sunagawa et al. 2010; Ainsworth et al. 2015; Apprill et al. 2016) and observed within coral-host cells in close proximity to symbiotic dinoflagellates (Ainsworth et al. 2015). The

functional role of *Ralstonia* spp. in corals has not been confirmed, but genetic evidence also suggests that they are well-suited for the symbiotic lifestyle (Ainsworth et al. 2015). We detected 4 distinct MED nodes associated with the *Ralstonia* genus in 74/77 of our samples, with the highest average relative abundances found in *Orbicella faveolata* (2.5 +/- 9.7%) and *Diploria strigosa* (17 +/- 30.1%) corals. This result demonstrates that the DNA extraction methods used in this study may have the capacity to lyse cells located within host-coral cells, thus confirming the use of these DNA extraction methods for studying complex, host-associated microbiomes.

However, if the goal of the investigation is to detect specific or cryptic/rare microorganisms (Ainsworth et al. 2015), care may need to be taken when choosing the DNA extraction method. This recommendation is supported by the minor, but important distinctions in microbial richness and coverage of top microbial groups between treatments that were detected during pairwise comparisons of the presence/absence of discrete MED nodes by coral colony. For example, nMDS and ANOSIM did not discern the higher MED coverage associated with PB extracts, but this trend was uncovered during presence/absence evaluation. Higher MED coverage, including the increased likelihood of detecting MED nodes not detected in sequences from other extractions (e.g., *Ca.* Branchiomonas, and *Thaumarchaeota*), may stem from using a mixture of bead sizes and types during PB DNA extraction. Crowder and colleagues (2010) (Crowder et al. 2010) used a mixture of 0.1 and 2.0 mm beads to extract DNA from ticks and reported that the 2.0 mm beads disrupted the tick exoskeleton, while the 0.1 mm beads disrupted soft tissue and microbial cells. The PB kit also uses a mixture of 0.1 and 0.5 mm glass beads and larger 2.4 mm ceramic beads to mechanically rupture cells. This bead combination may have facilitated lysis of more recalcitrant coral tissue with the ceramic beads (2.4 mm) and lysis of soft coral tissue (0.5 mm) and microbial cells (0.1 mm) with the glass beads. Using bead

mixtures during DNA extraction may be particularly important for studies investigating intracellular symbionts or rare microorganisms. Altogether, careful thought about the scope and expected outcomes of the planned research is needed because this may impact which DNA isolation treatment should be used.

2.6 Conclusions

This study demonstrates that the PowerBiofilm® and UltraClean® Tissue and Cells (and permutations) DNA Isolation kits are appropriate to use for extraction and amplification of microbial DNA from most colonies of *P. lobata*, *P. verrucosa*, *A. humilis*, *O. faveolata*, *O. annularis,* and *D. strigosa* corals. Subsequent microbial community analysis revealed that at a large scale, overall microbial community structure was significantly determined by coral species rather than DNA extraction treatment, a result that validates the use of either the PowerBiofilm® or UltraClean® Tissue and Cells (and permutations) DNA Isolation kits for broad coral microbiota comparisons of globally distributed coral species. On a finer scale, subtle, but potentially important differences in microbial community richness and coverage of top microbial groups were detected, trends that may stem from using different bead mixtures during mechanical lysis of the coral tissue. Based upon these results, we suggest that the PowerBiofilm® DNA extraction kit is the most reliable and comprehensive kit to use for small scale cultivation-independent characterization of coral microbiota.

As reliance on sequence data for scientific inquiry grows, acknowledgement of biases introduced to samples via methods is important not only for investigation of error and replication, but also for the detection of ecologically meaningful patterns. Methods vigilance, obtained by conducting dedicated method optimization studies, is a cornerstone for cultivation-

independent investigations of microbe-host associations. Understanding the influence of technical bias aids our detection of biologically relevant patterns from sequence data and deepens our understanding of the coral microbiome as well as other complex host environments.

2.7 Acknowledgments

2.8 References

Ainsworth, T.D., Krause, L., Bridge, T., Torda, G., Raina, J.B., Zakrzewski, M. *et al.* (2015) The coral core microbiome identifies rare bacterial taxa as ubiquitous endosymbionts. *ISME J* 9: 2261-2274.

Al-Soud, W.A., and Radstrom, P. (2001) Purification and characterization of PCR-inhibitory components in blood cells. *J Clin Microbiol* 39: 485-493.

Apprill, A., Hughen, K., and Mincer, T. (2013) Major similarities in the bacterial communities associated with lesioned and healthy *Fungiidae* corals. *Environ Microbiol* 15: 2063-2072.

Apprill, A., Weber, L.G., and Santoro, A.E. (2016) Distinguishing between microbial habitats unravels ecological complexity in coral microbiomes. *mSystems* 1.

Apprill, A., Marlow, H.Q., Martindale, M.Q., and Rappe, M.S. (2009) The onset of microbial associations in the coral *Pocillopora meandrina*. *ISME J* 3: 685-699.

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.

Baker, E.J., and Kellogg, C.A. (2014) Comparison of three DNA extraction kits to establish maximum yield and quality of coral-associated microbial DNA. *US Geological Survey Open-File Report 2014-1066*.

Bickley, J., Short, J.K., McDowell, D.G., and Parkes, H.C. (1996) Polymerase chain reaction (PCR) detection of *Listeria monocytogenes* in diluted milk and reversal of PCR inhibition caused by calcium ions. *Letters in Applied Microbiology* 22: 153-158.

Bourne, D.G., and Munn, C.B. (2005) Diversity of bacteria associated with the coral *Pocillopora damicornis* from the Great Barrier Reef. *Environ Microbiol* 7: 1162-1174.

Bourne, D.G., Morrow, K.M., and Webster, N.S. (2016) Insights into the coral microbiome: underpinning the health and resilience of reef ecosystems. *Annual Reviews Microbiology* 70: 317-340.

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.

Ceh, J., Van Keulen, M., and Bourne, D.G. (2011) Coral-associated bacterial communities on Ningaloo Reef, Western Australia. *FEMS Microbiol Ecol* 75: 134-144.

Crowder, C., Rounds, M.A., Phillipson, C.A., Picuri, J.M., Matthews, H.E., Halverson, J. *et al.* (2010) Extraction of total nucleic acids from ticks for the detection of bacterial and viral pathogens. *Journal of Medical Entomology* 47: 89-94.

de Carcer, D.A., Denman, S.E., McSweeney, C., and Morrison, M. (2011) Evaluation of subsampling-based normalization strategies for tagged high-throughput sequencing data sets from gut microbiomes. *Appl Environ Microbiol* **77**: 8795-8798.

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

Eren, A.M., Maignien, L., Sul, W.J., Murphy, L.G., Grim, S.L., Morrison, H.G., and Sogin, M.L. (2013) Oligotyping: Differentiating between closely related microbial taxa using 16S rRNA gene data. *Methods Ecol Evol* **4**.

Feehery, G.R., Yigit, E., Oyola, S.O., Langhorst, B.W., Schmidt, V.T., Stewart, F.J. *et al.* (2013) A method for selectively enriching microbial DNA from contaminating vertebrate host DNA. *PLoS One* **8**: e76096.

Galkiewicz, J.P., and Kellogg, C.A. (2008) Cross-kingdom amplification using bacteria-specific primers: complications for studies of coral microbial ecology. *Appl Environ Microbiol* **74**: 7828-7831.

Garcia, G.D., Gregoracci, G.B., Santos Ede, O., Meirelles, P.M., Silva, G.G., Edwards, R. *et al.* (2013) Metagenomic analysis of healthy and white plague-affected *Mussismilia braziliensis* corals. *Microb Ecol* **65**: 1076-1086.

Hausman, R.E., and Burnett, A.L. (1969) The mesoglea of hydra. I. Physical and histochemical properties. *Journal of Experimental Zoology* **171**: 7-13.

He, S., Ivanova, N., Kirton, E., Allgaier, M., Bergin, C., Scheffrahn, R.H. *et al.* (2013) Comparative metagenomic and metatranscriptomic analysis of hindgut paunch microbiota in wood- and dung-feeding higher termites. *PLoS One* **8**: e61126.

Henderson, G., Cox, F., Kittelmann, S., Miri, V.H., Zethof, M., Noel, S.J. *et al.* (2013) Effect of DNA extraction methods and sampling techniques on the apparent structure of cow and sheep rumen microbial communities. *PLoS One* **8**: e74787.

Kellogg, C.A., Piceno, Y.M., Tom, L.M., DeSantis, T.Z., Gray, M.A., and Andersen, G.L. (2014) Comparing bacterial community composition of healthy and dark spot-affected *Siderastrea siderea* in Florida and the Caribbean. *PLoS One* **9**: e108767.

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., and Glockner, F.O. (2013) Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res* **41**: e1.

Knowlton, N., and Rohwer, F. (2003) Multispecies microbial mutualisms on coral reefs: The host as a habitat. *Am Nat* **162**: S51-S62.

Koonin, E.V., and Wolf, Y.I. (2008) Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Res* **36**: 6688-6719.

Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K., and Schloss, P.D. (2013) Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol* **79**: 5112-5120.

Krediet, C.J., Ritchie, K.B., Alagely, A., and Teplitski, M. (2013) Members of native coral microbiota inhibit glycosidases and thwart colonization of coral mucus by an opportunistic pathogen. *ISME J* **7**: 980-990.

Lampert, Y., Kelman, D., Nitzan, Y., Dubinsky, Z., Behar, A., and Hill, R.T. (2008) Phylogenetic diversity of bacteria associated with the mucus of Red Sea corals. *FEMS Microbiol Ecol* **64**: 187-198.

Lesser, M.P., Mazel, C.H., Gorbunov, M.Y., and Falkowski, P.G. (2004) Discovery of symbiotic nitrogen-fixing cyanobacteria in corals. *Science* **305**: 997-1000.

Littman, R., Willis, B.L., and Bourne, D.G. (2011) Metagenomic analysis of the coral holobiont during a natural bleaching event on the Great Barrier Reef. *Environ Microbiol Rep* **3**: 651-660.

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar *et al.* (2004) ARB: a software environment for sequence data. *Nucleic Acids Res* **32**: 1363-1371.

McMurdie, P.J., and Holmes, S. (2013) phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* **8**: e61217.

McMurdie, P.J., and Holmes, S. (2014) Waste not, want not: why rarefying microbiome data is inadmissible. *Plos Computational Biology* **10**: e1003531.

Meron, D., Atias, E., Iasur Kruh, L., Elifantz, H., Minz, D., Fine, M., and Banin, E. (2011) The impact of reduced pH on the microbial community of the coral Acropora eurystoma. *ISME J* **5**: 51-60.

Middleberg, A.P.J. (1995) Process-scale disruption of microorganisms. *Biotechnol Adv* **13**: 491-551.

Morgulis, A., Coulouris, G., Raytselis, Y., Madden, T.L., Agarwala, R., and Schaffer, A.A. (2008) Database indexing for production MegaBLAST searches. *Bioinformatics* **24**: 1757-1764.

Mori, H., Maruyama, F., Kato, H., Toyoda, A., Dozono, A., Ohtsubo, Y. *et al.* (2014) Design and experimental application of a novel non-degenerate universal primer set that amplifies prokaryotic 16S rRNA genes with a low possibility to amplify eukaryotic rRNA genes. *DNA Res* **21**: 217-227.

Morrow, K.M., Moss, A.G., Chadwick, N.E., and Liles, M.R. (2012) Bacterial associates of two Caribbean coral species reveal species-specific distribution and geographic variability. *Appl Environ Microbiol* **78**: 6438-6449.

Mouchka, M.E., Hewson, I., and Harvell, C.D. (2010) Coral-associated bacterial assemblages: current knowledge and the potential for climate-driven impacts. *Integr Comp Biol* **50**: 662-674.

Muscatine, L., McCloskey, L.R., and Marian, R.E. (1981) Estimating the daily contribution of carbon from zooxanthellae to coral animal respiration. *Limnol Oceanogr* **26**: 601-611.

Neave, M.J., Apprill, A., Ferrier-Pages, C., and Voolstra, C.R. (2016) Diversity and function of prevalent symbiotic marine bacteria in the genus *Endozoicomonas*. *Appl Microbiol Biotechnol* **100**: 8315-8324.

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-UK* **7**.

Neave, M.J., Rachmawati, R., Xun, L.P., Michell, C.T., Bourne, D.G., Apprill, A., and Voolstra, C.R. (2017) Differential specificity between closely related corals and abundant Endozoicomonas endosymbionts across global scales. *ISME J* **11**: 186-200.

Nesme, J., Cecillon, S., Delmont, T.O., Monier, J.M., Vogel, T.M., and Simonet, P. (2014) Large-scale metagenomic-based study of antibiotic resistance in the environment. *Curr Biol* **24**: 1096-1100.

Poretsky, R., Rodriguez, R.L., Luo, C., Tsementzi, D., and Konstantinidis, K.T. (2014) Strengths and limitations of 16S rRNA gene amplicon sequencing in revealing temporal microbial community dynamics. *PLoS One* **9**: e93827.

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

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

Raina, J.B., Tapiolas, D., Willis, B.L., and Bourne, D.G. (2009) Coral-associated bacteria and their role in the biogeochemical cycling of sulfur. *Appl Environ Microb* **75**: 3492-3501.

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

Rohwer, F., Seguritan, V., Azam, F., and Knowlton, N. (2002) Diversity and distribution of coral-associated bacteria. *Mar Ecol Prog Ser* **243**: 1-10.

Rudi, K., Wilson, R.C., Jakobsen, K.S., and Liu, D. (2009) Preparation of bacteria samples for direct molecular applications. Boca Raton: CRC Press.

Salerno, J.L., Reineman, D.R., Gates, R.D., and Rappé, M.S. (2011) The Effect of a Sublethal Temperature Elevation on the Structure of Bacterial Communities Associated with the CoralPorites compressa. *Journal of Marine Biology* **2011**: 1-9.

Santos, H.F., Carmo, F.L., Leite, D.C.A., Jesus, H.E., De Carvalho Maalouf, P., Almeida, C. *et al.* (2012) Comparison of different protocols for the extraction of microbial DNA from reef corals. *Brazilian Journal of Microbiology*: 517-527.

Schrader, C., Schielke, A., Ellerbroek, L., and Johne, R. (2012) PCR inhibitors - occurrence, properties and removal. *J Appl Microbiol* **113**: 1014-1026.

Sekar, R., Mills, D.K., Remily, E.R., Voss, J.D., and Richardson, L.L. (2006) Microbial communities in the surface mucopolysaccharide layer and the black band microbial mat of black band-diseased *Siderastrea siderea. Appl Environ Microbiol* **72**: 5963-5973.

Shinzato, C., Mungpakdee, S., Satoh, N., and Shoguchi, E. (2014) A genomic approach to coral-dinoflagellate symbiosis: studies of *Acropora digitifera* and *Symbiodinium minutum. Front Microbiol* **5**: 336.

Sipos, R., Szekely, A.J., Palatinszky, M., Revesz, S., Marialigeti, K., and Nikolausz, M. (2007) Effect of primer mismatch, annealing temperature and PCR cycle number on 16S rRNA gene-targetting bacterial community analysis. *FEMS Microbiol Ecol* **60**: 341-350.

Sneed, J.M., Sharp, K.H., Ritchie, K.B., and Paul, V.J. (2014) The chemical cue tetrabromopyrrole from a biofilm bacterium induces settlement of multiple Caribbean corals. *Proc Biol Sci* **281**.

Sunagawa, S., Woodley, C.M., and Medina, M. (2010) Threatened corals provide underexplored microbial habitats. *PloS One* **5**.

Sunagawa, S., DeSantis, T.Z., Piceno, Y.M., Brodie, E.L., DeSalvo, M.K., Voolstra, C.R. *et al.* (2009) Bacterial diversity and white plague disease-associated community changes in the Caribbean coral *Montastraea faveolata. ISME J* **3**: 512-521.

Sunagawa, S., Coelho, L.P., Chaffron, S., Kultima, J.R., Labadie, K., Salazar, G. *et al.* (2015) Structure and function of the global ocean microbiome. *Science* **348**.

Thompson, J.R., Rivera, H.E., Closek, C.J., and Medina, M. (2014) Microbes in the coral holobiont: partners through evolution, development, and ecological interactions. *Front Cell Infect Microbiol* **4**: 176.

Vega Thurber, R., Burkepile, D.E., Correa, A.M., Thurber, A.R., Shantz, A.A., Welsh, R. *et al.* (2012) Macroalgae decrease growth and alter microbial community structure of the reef-building coral, Porites astreoides. *PLoS One* **7**: e44246.

Vega Thurber, R., Willner-Hall, D., Rodriguez-Mueller, B., Desnues, C., Edwards, R.A., Angly, F. *et al.* (2009) Metagenomic analysis of stressed coral holobionts. *Environ Microbiol* **11**: 2148-2163.

Wagner Mackenzie, B., Waite, D.W., and Taylor, M.W. (2015) Evaluating variation in human gut microbiota profiles due to DNA extraction method and inter-subject differences. *Front Microbiol* **6**: 130.

Webster, N.S., Negri, A.P., Flores, F., Humphrey, C., Soo, R., Botte, E.S. *et al.* (2013) Near-future ocean acidification causes differences in microbial associations within diverse coral reef taxa. *Environ Microbiol Rep* **5**: 243-251.

Yang, Q., Franco, C.M., and Zhang, W. (2015) Sponge-associated actinobacterial diversity: validation of the methods of actinobacterial DNA extraction and optimization of 16S rRNA gene amplification. *Appl Microbiol Biotechnol* **99**: 8731-8740.

Yuan, S., Cohen, D.B., Ravel, J., Abdo, Z., and Forney, L.J. (2012) Evaluation of methods for the extraction and purification of DNA from the human microbiome. *PLoS One* **7**: e33865.

Zhang, Z., Schwartz, S., Wagner, L., and Miller, W. (2000) A greedy algorithm for aligning DNA sequences. *J Comput Biol* **7**: 203-214.

Ziegler, M., Roik, A., Porter, A., Zubier, K., Mudarris, M.S., Ormond, R., and Voolstra, C.R. (2016) Coral microbial community dynamics in response to anthropogenic impacts near a major city in the central Red Sea. *Mar Pollut Bull* **105**: 629-640.

2.9 Tables

Table 1. Comparison of DNA treatment extraction characteristics.

MO BIO Extraction Characteristics	PS	PP	PB	VG	PG	VGl
Bead Diameter (mm)	0.7	2.38	0.1, 0.5, 2.4*	0.7	0.1	0.1
Bead Type	Garnet	Metal	Glass, Ceramic	Garnet	Glass	Glass
Homogenization Method	Vortex	Vortex	Vortex	Vortex	Powerlyzer	Vortex
Homogenization Duration	15 min	15 min	15 min	15 min	45 sec	15 min
Surfactant	<5%	<5%	X	X	X	X
Protein Precipitant	20-40%	10-20%	10-15%	X	X	X
Guanidine Thiocyanate	X	<3%	<3%, 60-80%	X	X	X
Inhibitor Removal	<10%	<5%	<10%	X	X	X
Proteinase K	X	X	X	1-5%	1-5%	1-5%
RNase	X	25%	X	X	X	X
Phenolic Separation Solution	X	5-15%	X	X	X	X

X indicates that the parameter was not included. PS = PowerSoil, PP = PowerPlant Pro, PB = PowerBiofilm, VG = UC Vortex Garnet, PG = UC Powerlyzer Glass, VGl = UC Vortex Glass.

*PB kit uses 0.1 and 0.5 mm glass beads and 2.4 mm ceramic beads.

Table 2. Summary of DNA extraction yield and PCR efficiency for extractions performed in phase I, with treatments selected for phase II bolded.

Treatment x Species (number of samples)	Average DNA Yield (S.D.) $(ng\ \mu l^{-1})\ ^\alpha$	PCR efficiency Ω
PS x *P. lobata* (3)	4.38 (2.60)[a,b]	1
PP x *P. lobata* (3)	3.73 (1.92)[a,b]	0.67
PB x *P. lobata* (3)	**3.27 (2.86)[a,b]**	**1**
VGl x *P. lobata* (3)	**0.74 (0.36)[b]**	**1**
PG x *P. lobata* (3)	**0.40 (0.06)[b]**	**0.33**
VG x *P. lobata* (3)	**10.31 (5.69)[a]**	**0.67**
PS x *P. verrucosa* (3)	7.35 (10.71)	0.67
PP x *P. verrucosa* (3)	7.96 (7.81)	0
PB x *P. verrucosa* (3)	**27.21 (23.58)**	**0.67**
VGl x *P. verrucosa** (1)	**4.14**	**1**
PG x *P. verrucosa* (3)	**6.06 (3.56)**	**1**
VG x *P. verrucosa* (3)	**12.17 (6.83)**	**1**
PS x *A. humilis* (3)	18.11 (25.20)	0
PP x *A. humilis* (3)	7.75 (11.33)	0.33
PB x *A. humilis* (3)	**38.20 (11.70)**	**1**
VGl x *A. humilis** (2)	**9.46 (5.43)**	**1**
PG x *A. humilis* (3)	**5.49 (4.56)**	**1**
VG x *A. humilis* (3)	**29.33 (3.31)**	**1**
PS x *O. faveolata* (4)	6.51 (2.95)	0.75
PP x *O. faveolata* (4)	2.24 (1.46)	0
PB x *O. faveolata* (4)	**8.05 (8.08)**	**1**
VGl x *O. faveolata* (4)	**8.24 (7.69)**	**1**
PG x *O. faveolata* (4)	**7.47 (9.79)**	**1**
VG x *O. faveolata* (4)	**12.19 (13.27)**	**1**
PS x *M. cavernosa* (3)	2.73 (2.10)	0.33
PP x *M. cavernosa* (3)	1.31 (0.42)	0
PB x *M. cavernosa* (3)	**1.62 (1.06)**	**0.67**
VGl x *M. cavernosa* (3)	**0.67 (0.36)**	**0.67**
PG x *M. cavernosa* (3)	**1.59 (1.58)**	**1**
VG x *M. cavernosa* (3)	**1.79 (1.73)**	**0.33**
PS x *O. annularis* (3)	7.23 (7.54)	0
PP x *O. annularis* (3)	1.97 (1.18)	0
PB x *O. annularis* (3)	**9.92 (7.31)**	**1**
VGl x *O. annularis* (3)	**1.63 (1.08)**	**1**
PG x *O. annularis* (3)	**1.47 (1.78)**	**1**
VG x *O. annularis* (3)	**11.47 (13.35)**	**1**
PS x *D. strigosa* (3)	2.28 (2.25)	0.33

PP x *D. strigosa* (3)	1.14 (0.29)	0
PB x *D. strigosa* (3)	**0.96 (0.62)**	**0.33**
VGl x *D. strigosa* (3)	**0.24 (0.16)**	**0.33**
PG x *D. strigosa* (3)	**0.33 (0.14)**	**0.33**
VG x *D. strigosa* (3)	**0.98 (0.41)**	**1**

[α, ab] Values within a column with different superscripts indicate significantly different DNA concentrations between treatments (one-way FRMANOVA; Holm-Sidak Method, $p < 0.05$).
[Ω]Values represent normalized PCR efficiency

Table 3. Summary of microbial community analysis conducted during phase II.

Treatment x Species	n	# reads[ε]	MED richness[ρ]	MED evenness (J')[δ]	Total Community Structure[φ]	Detailed Analysis	Top 50% MED Coverage
PB x *P. lobata*	2	142,167 (14,141)	129 (35)	0.40 (0.13)[b]	VGl, PB: R=0.75 VGl, VG: R=1	No; low reads	--------
VGl x *P. lobata*	3	401 (161)	86 (22)	0.70 (0.07)[a]	VG, PB: R=0.5 p =0.001 for all		
PG x *P. lobata*	1	641	100	0.61[b]	-----		
VG x *P. lobata*	3	2,594 (713)	74 (16)	0.29 (0.08)[b]	Tested above		
PB x *P. verrucosa*	1	13,153	161	0.53	-----	No; low reads; low biological replication	--------
VGl x *P. verrucosa**	1	603	117	0.78	-----		
PG x *P. verrucosa*	3	745 (691)	105 (19)[a]	0.77 (0.16)	No differences		
VG x *P. verrucosa*	3	152 (23)	65 (11)[b]	0.85 (0.04)			
PB x *A. humilis*	3	22,105 (20,156)	131 (30)[b]	0.31 (0.21)	No differences	Yes	0.78
VGl x *A. humilis**	2	21,133 (27,005)	60 (2)[a]	0.36 (0.41)			0.67
PG x *A. humilis*	2	42,241 (53,217)	85 (2)	0.40 (0.45)			0.67
VG x *A. humilis*	3	15,536 (11,314)	59 (5)[a]	0.27 (0.21)			0.7
PB x *O. faveolata*	4	72,587 (25,401)	378 (71)	0.60 (0.16)	No differences	Yes	0.84
VGl x *O. faveolata*	4	16,355 (18,421)	237 (90)	0.56 (0.17)			0.71
PG x *O. faveolata*	4	28,852 (17,895)	392 (87)	0.66 (0.12)			0.77
VG x *O. faveolata*	4	18,630 (9,437)	361 (43)	0.61 (0.10)			0.68
PB x *M. cavernosa*	0	n/a	n/a	n/a	n/a	No; not all treatments represented	--------
VGl x *M. cavernosa*	2	65,040 (3,908)	405 (22)	0.63 (0.12)	No differences		
PG x *M. cavernosa*	3	51,741 (26,956)	366 (40)	0.57 (0.04)			
VG x *M. cavernosa*	2	52,055 (7,028)	355 (115)	0.67 (0.14)			
PB x *O. annularis*	3	87,777 (11,390)	485 (60)	0.73 (0.09)	No differences	Yes	0.87
VGl x *O. annularis*	3	32,681 (4,358)	357 (45)	0.74 (0.08)			0.77
PG x *O. annularis*	3	24,582 (4,974)	313 (130)	0.75 (0.15)			0.67
VG x *O. annularis*	3	34,152 (8,088)	382 (123)	0.71 (0.16)			0.77

74

PB x *D. strigosa*	1	65,392	268	0.27	n/a	No; not all treatments represented; low biological replication	--------
VGl x *D. strigosa*	0	n/a	n/a	n/a			
PG x *D. strigosa*	1	60,393	211	0.23			
VG x *D. strigosa*	1	41,726	172	0.18			

n= number of samples included in microbial community analysis after quality-filtering sequences.

All values are presented as mean (standard deviation (S. D.)) when appropriate. Single values with no S.D. represent samples from treatments with no replicates and these values were not included in statistical significance testing.

ε Average number of reads obtained for that species x treatment grouping out of the total number of analyzed reads.

ρ [ab] Species values under ρ within a column with different superscripts indicate significantly different MED richness between treatments ($p < 0.05$, one-tailed t-test or one-way ANOVA with Holm-Sidak Method post-hoc test).

δ [ab] MED evenness values (J') for each treatment with different superscripts indicate significant differences in species evenness ($p < 0.05$, one-way ANOVA with Holm-Sidak Method post-hoc test).

φ Differences in community structure were first determined using one-way ANOSIM global tests within each coral species ($p < 0.05$ is significance threshold). If significant differences were found, pairwise tests were conducted between the different treatments. Species x treatment combinations with only 1 sample were excluded in this analysis.
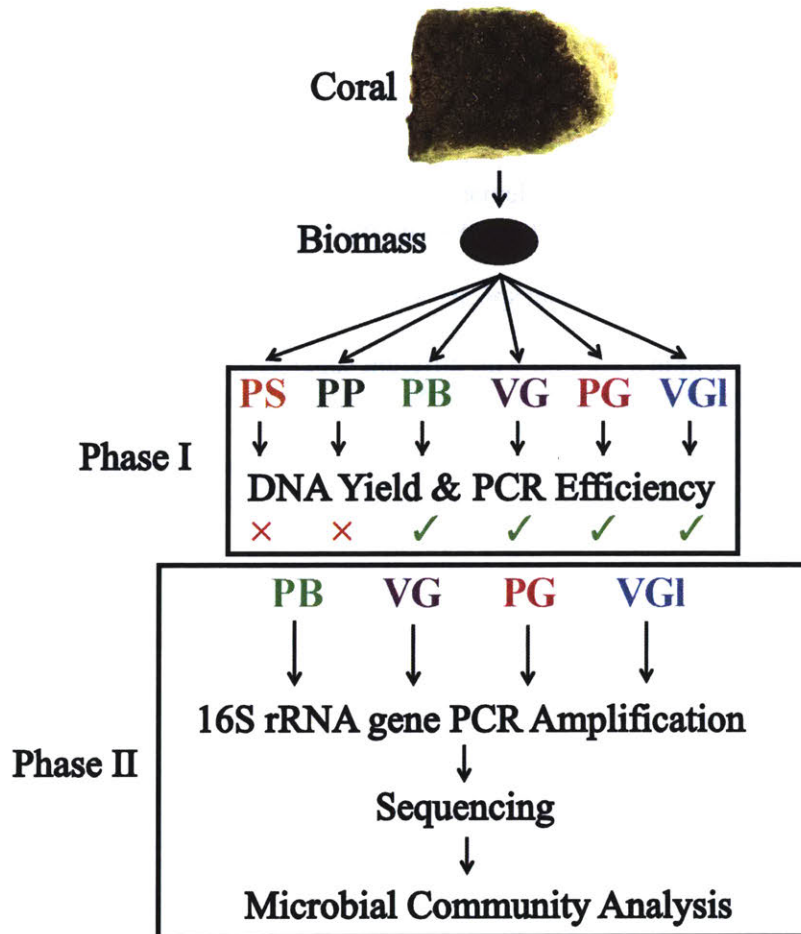
Figure 1. Overview of experimental design. During phase I, DNA extraction treatments were performed on subdivided tissue, with efficiency of SSU gene amplification assessed using gel screening of PCR products. The green check mark and red X indicate that amplicons from the treatment were and were not chosen for sequencing, respectively. During phase II, well-performing PB, VG, PG, and VGl extracts were amplified and sequenced for microbial community analysis. PS = PowerSoil, PP = PowerPlant Pro, PB = PowerBiofilm, VG = UC Vortex Garnet, PG = UC Powerlyzer Glass, VGl = UC Vortex Glass.
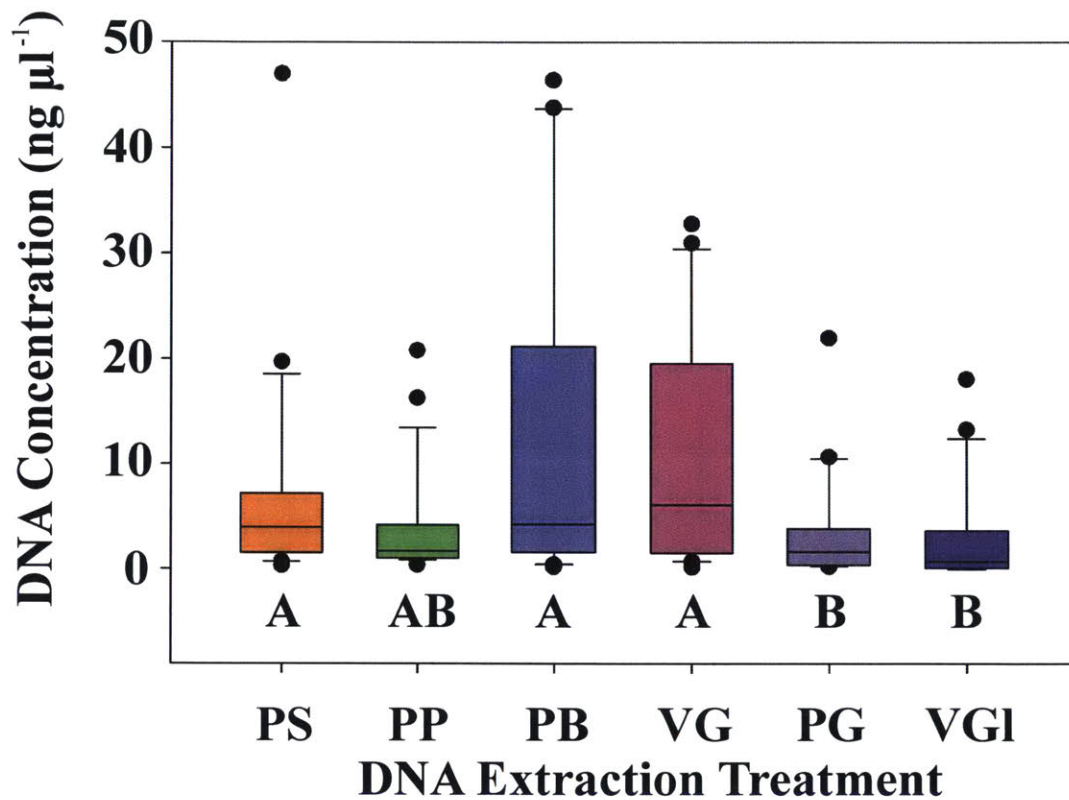
Figure 2. Boxplot of total DNA concentrations grouped by treatment (n=19-22 individual extractions per treatment). A and B letters differentiate significantly different groups (Tukey's test $p < 0.05$). Medians are indicated by the solid black lines and the 25% and 75% quartiles are represented by the lower and upper bounds of the box. Outliers are indicated as black circles and represent samples falling outside the 10% and 90% quartiles. PS = PowerSoil, PP = PowerPlant Pro, PB = PowerBiofilm, VG = UC Vortex Garnet, PG = UC Powerlyzer Glass, VGl = UC Vortex Glass.
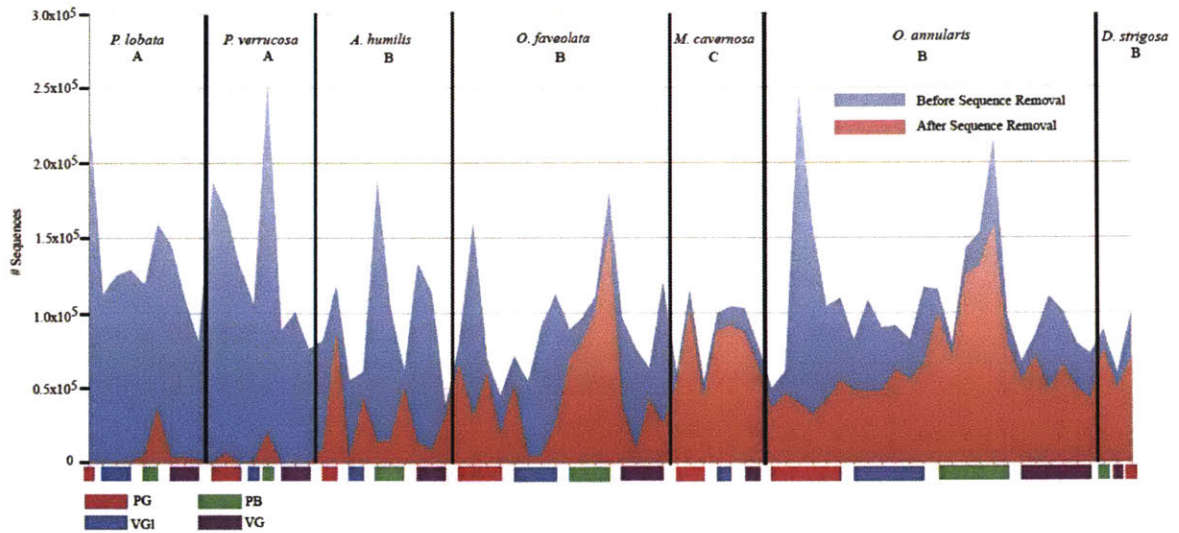
Figure 3. Number of sequences before and after quality-filtering and removal of low quality sequences. Samples are grouped by DNA extraction treatment nested within coral species. Different letters (A, B, and C) denote statistically significant differences between species (one-tailed t-test or Mann-Whitney ranked sums test, $p < 0.05$). PG = UC Powerlyzer Glass, VGl = UC Vortex Glass, PB = PowerBiofilm, VG = UC Vortex Garnet.
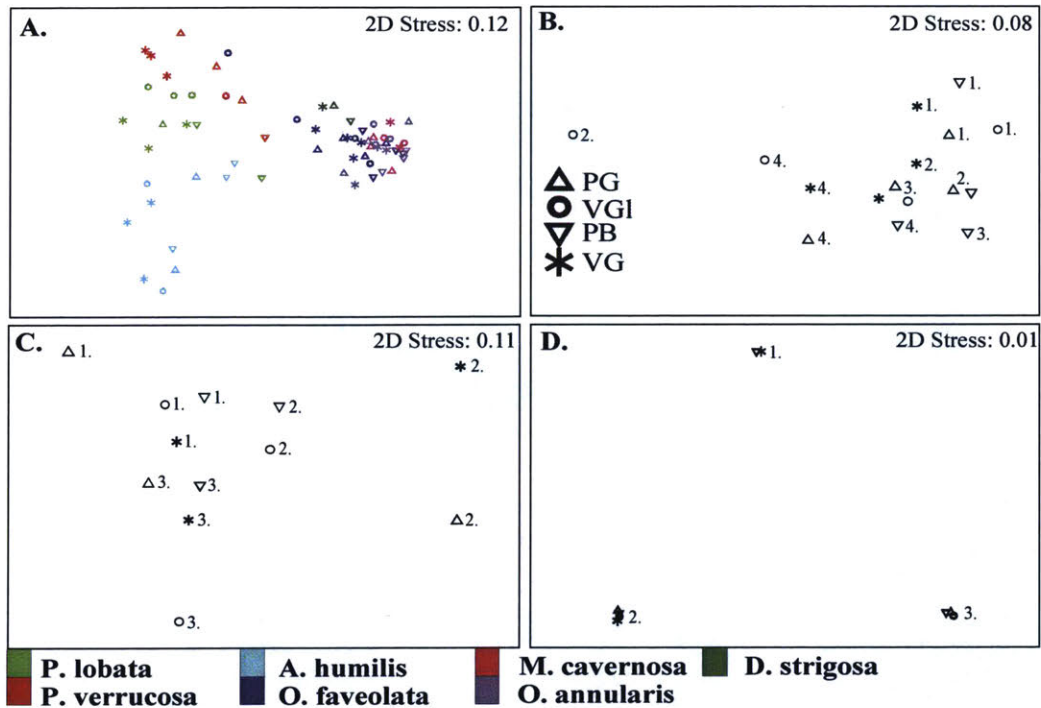
Figure 4. nMDS ordination of SSU rRNA gene sequences recovered from the different DNA extraction treatments and compared using Bray-Curtis distances for A) all species, B) *O. faveolata*, C) *O. annularis*, and D) *A. humilis*. In A, species groupings are designated by colors. In B, C, and D, samples from the same coral colony are designated by number.
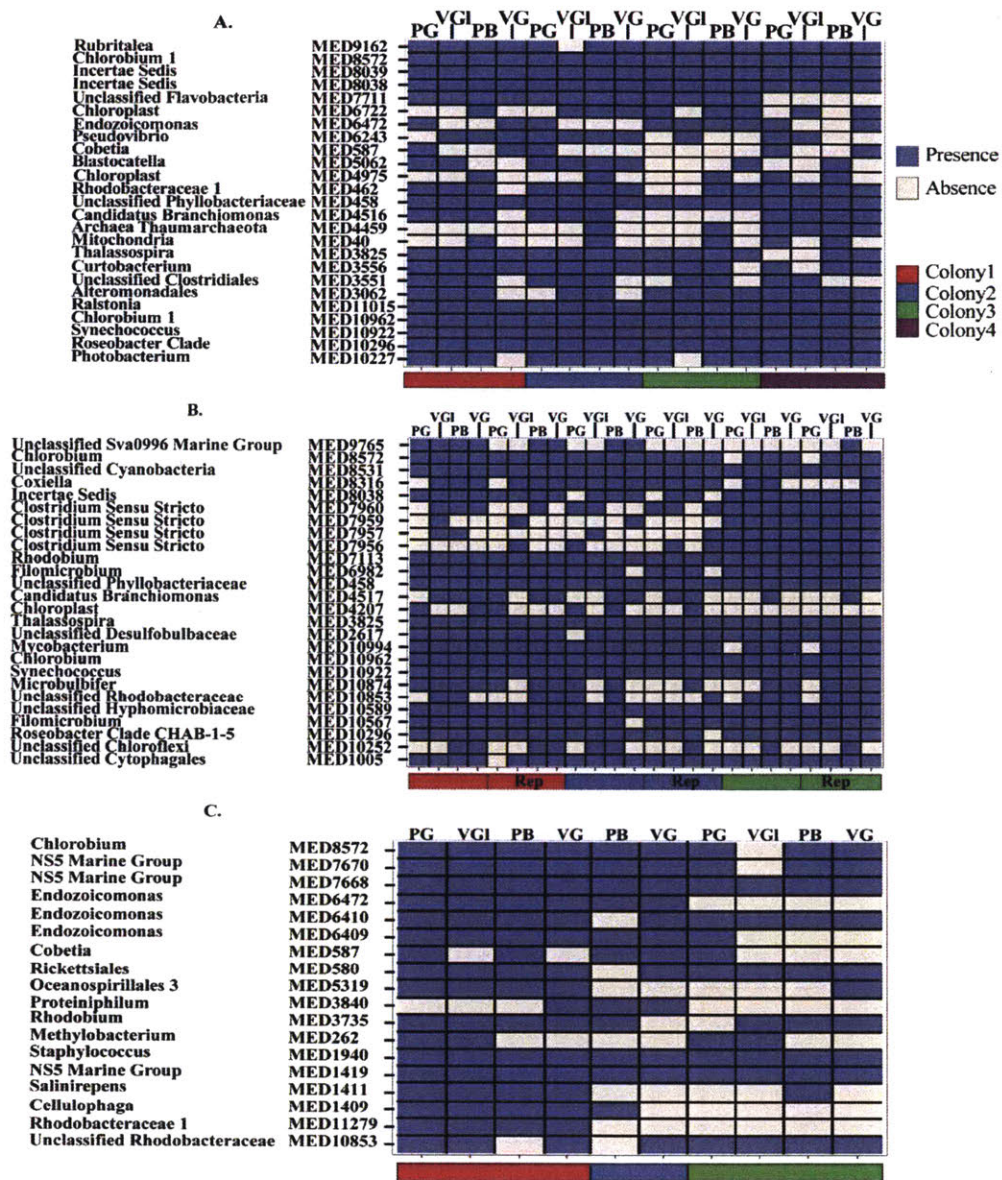
Figure 5. Heatmaps displaying the presence or absence of dominant MED nodes that ranked within the top 50% of the dataset for each species for A) *O. faveolata,* B) *O. annularis,* and C) *A. humilis.* 'Rep' designates technical replicates of identical *O. annularis* DNA extracts tagged with different barcodes during PCR. The colors designate different colonies of that species.

2.11 Supporting Information

Table S1. Coral species, collection site location, collection depth, and colony name for colonies.

Species	Collection Site	Latitude	Longitude	Depth (ft)	Colony name
Porites lobata	Shallow barrier reef/lagoon, Kapangamarangi Atoll, Micronesia	1.02695 N	154.774643 W	1	Plob1
Porites lobata	Shallow barrier reef/lagoon, Kapangamarangi Atoll, Micronesia	1.02695 N	154.774644 W	1	Plob2
Porites lobata	Shallow barrier reef/lagoon, Kapangamarangi Atoll, Micronesia	1.02695 N	154.774645 W	1	Plob3
Pocillopora verrucosa	Patch reef/lagoon, Kapangamarangi Atoll, Micronesia	1.0365 N	154.765717 W	33	Pverr1
Pocillopora verrucosa	Patch reef/lagoon, Kapangamarangi Atoll, Micronesia	1.0365 N	154.765719 W	33	Pverr2
Pocillopora verrucosa	Patch reef/lagoon, Kapangamarangi Atoll, Micronesia	1.02695 N	154.774645 W	33	Pverr3
Acropora humilis	Magnetic Bay, Australia	15.0883 S	146.8520 E	5	Ahum1
Acropora humilis	Magnetic Bay, Australia	15.0883 S	146.8520 E	5	Ahum2
Acropora humilis	Magnetic Bay, Australia	15.0883 S	146.8520 E	5	Ahum3
Diploria strigosa	Midpatch reef, Fl Keys, USA	24.59306 N	81.39111 W	18	Dstrig1
Diploria strigosa	Openwater patch reef, Fl Keys, USA	24.557886 N	81.436053 W	20	Dstrig2
Diploria strigosa	Midpatch reef, Fl Keys, USA	24.562519 N	81.500666 W	18	Dstrig3
Orbicella faveolata	Nearshore reef, Fl Keys, USA	24.59306 N	81.39111 W	9	Ofav1
Orbicella faveolata	Reef flat, Fl Keys, USA	24.606054 N	81.429342 W	19	Ofav2
Orbicella faveolata	Midpatch reef, Fl Keys, USA	24.59306 N	81.39111 W	18	Ofav3
Orbicella faveolata	Reef flat, Fl Keys, USA	24.562519 N	81.500666 W	17	Ofav4
Montastraea cavernosa	Reef flat, Fl Keys, USA	24.59306 N	81.39111 W	18	Mcav1
Montastraea cavernosa	Openwater patch reef, Fl Keys, USA	24.557886 N	81.436053 W	21	Mcav2
Montastraea cavernosa	Midpatch reef, Fl Keys, USA	24.562519 N	81.500666 W	17	Mcav3
Montastraea annularis	Openwater patch reef, Fl Keys, USA	24.557886 N	81.436053 W	22	Mann1
Montastraea annularis	Midpatch reef, Fl Keys, USA	24.562519 N	81.500666 W	18	Mann2
Montastraea annularis	Openwater patch reef, Fl Keys, USA	24.557886 N	81.436053 W	23	Mann3

Table S2: PCR efficiency displayed for each colony and extraction method.

Treatment	PS	PP	PB	VG	PG	VGl
Plob1	+	+	+	+	+	+
Plob2	+	-	+	-	-	+
Plob3	+	+	+	+	-	+
Pverr1	+	-	+	+	+	X
Pverr2	+	-	-	+	+	+
Pverr3	-	-	+	+	+	X
Ahum1	-	-	+	+	+	+
Ahum2	-	-	+	+	+	X
Ahum3	-	+	+	+	+	+
Ofav1	+	-	+	+	+	+
Ofav2	+	-	+	+	+	+
Ofav3	-	-	+	+	+	+
Ofav4	+	-	+	+	+	+
Mcav1	+	-	+	+	+	-
Mcav2	-	-	-	-	+	+
Mcav3	-	-	+	-	+	+
Mann1	-	-	+	+	+	+
Mann2	-	-	+	+	+	+
Mann3	-	-	+	+	+	+
Dstrig1	-	-	-	+	-	-
Dstrig2	-	-	-	+	-	-
Dstrig3	+	-	+	+	+	+
PCR Efficiency◆	45%	14%	82%	86%	82%	84%

◆PCR efficiency was calculated as the percentage of successfully amplified bands of the correct size (292 bp, including primers) out of the total number of samples that were subjected to PCR for each extraction treatment.

'+'= band present, '-'=band absent, 'X'= no sample.

Availability of Data and Material

Demultiplexed raw sequence reads supporting the conclusions of this article were deposited into

the Sequence Read Archive (SRA, http://www.ncbi.nlm.nih.gov/sra) under the BioProject

accession number SUB1357861 and BioSample accession numbers SAMN04531900 through

SAMN04531977. Information about the study can also be found at the Biological and Chemical

Oceanography Data Management Office website (http://lod.bco-dmo.org/id/dataset/662114).

Chapter 3

The coral ecosphere: A unique coral reef habitat that fosters coral–microbial interactions

3.1 Abstract

Scleractinian corals are bathed in a sea of planktonic and particle-associated microorganisms. The metabolic products of corals influence the growth and composition of microorganisms, but interactions between corals and seawater microorganisms are underexplored. We conducted a field-based survey to compare the biomass, diversity, composition, and functional capacity of microorganisms in small-volume seawater samples collected adjacent to five coral species with seawater collected >1m away from the reef substrate on the same reefs. Seawater collected close to corals generally harbored copiotrophic-type bacteria and its bacterial and archaeal composition was influenced by coral species as well as the local reef environment. Trends in picoplankton abundances were variable and either increased or decreased away from coral colonies based on coral species and picoplankton functional group. Genes characteristic of surface-attached and potentially virulent microbial lifestyles were enriched in near coral seawater compared to reef seawater. There was a prominent association between the coral *Porites astreoides* and the coral symbiont *Endozoicomonas*, suggesting recruitment and/or shedding of these cells into the surrounding seawater. This evidence extends our understanding of potential species-specific and reef site-influenced microbial interactions that occur between corals and microorganisms within this near coral seawater environment that we propose to call the 'coral ecosphere'. Microbial interactions that occur within the coral ecosphere could influence recruitment of coral-associated microorganisms and facilitate the transfer of coral metabolites into the microbial food web, thus fostering reef biogeochemical cycling and a linkage between corals and the water column.

3.2 Introduction

Marine organisms are bathed in seawater that is densely populated by protists, bacteria, archaea, and viruses. This continuous contact likely facilitates interactions between marine bacteria and archaea and single-celled or multicellular organisms. For example, heterotrophic bacteria residing within the microenvironment surrounding and directly attached to eukaryotic phytoplankton cells can interact on a cellular level with the host eukaryote. These heterotrophic cells respire the dissolved organic matter (DOM) released by the eukaryote, synthesize and transfer essential vitamins to the host, and/or engage with the host using infochemicals (Seymour et al. 2017). These interactions can be beneficial, neutral, and/or exploitative and may impact productivity, growth rates, and life cycles of specific phytoplankton, potentially influencing the primary productivity of the ecosystem (Seymour et al. 2017). We hypothesize that these

interactions may be present and even more pronounced for much larger sessile organisms such as kelp, corals, and sponges, as their fixed location on the seafloor provides the opportunity to foster specific host-microbial interactions.

The microbiomes of scleractinian corals are some of the most well-characterized host-associated communities in the marine environment (reviewed by Thompson et al. 2014; and Bourne et al. 2016), but much less is known about how corals interact with surrounding seawater microbial communities. Previous investigations of reef water microbial community dynamics have revealed relationships between the composition of reef macrofauna, the composition and metabolism of bacteria and archaea in reef seawater, the abundances of heterotrophic bacteria and virulence genes, and coral health (Dinsdale et al. 2008; Kelly et al. 2012; Kelly et al. 2014; Haas et al. 2016). In addition, recent studies have suggested the existence of a previously unrecognized coral-associated microbial environment: the seawater adjacent to corals (Tout et al. 2014; Silveira et al. 2017; Walsh et al. 2017; Ochsenkuhn et al. 2018).

Corals may indeed influence the composition, structure, and function of these surrounding planktonic microbial communities. For instance, corals secrete DOM that can be degraded by and even serve as chemical cues for motile marine bacteria (Nelson et al. 2013; Garren et al. 2014; Tout et al. 2015). Additionally, some corals graze on picoplankton and remove cells from the water column (Houlbrèque et al. 2006; McNally et al. 2017). Physically, coral colonies interrupt water flow at different scales and form centimeter-scale momentum boundary layers surrounding individual colonies as well as micro-vortices closer to the coral surface (Chamberlain and Graus 1975; Shashar et al. 1996; Kaandorp et al. 2003; Shapiro et al. 2014). Together, these factors suggest that distinct microbial communities may form surrounding corals within the coral momentum boundary layer. Furthermore, microbial interactions that occur

within this environment could influence microbial symbiont acquisition and pathogen recruitment to the coral surface.

In a preliminary investigation of two coral colonies, Tout and colleagues (2014) detected genomic differences between seawater collected above corals and surface reef seawater. Despite collecting large volumes (10 L) of seawater that may have integrated the microbial heterogeneity that exists at smaller scales, Tout et al. (2014) found enrichment of copiotrophic bacteria near the corals, as well as genes used for bacterial motility, chemotaxis, membrane transport, iron-acquisition, and metabolism of aromatic compounds in addition to other pathways (Tout et al. 2014). In contrast, Silveira et al. (2017) did not detect any significant differences in the functional or taxonomic microbial composition between large (80 L) volume samples of near coral seawater collected surrounding patches of the coral *Mussismilia braziliensis* and the water column. A study by Walsh and colleagues (2017) detected differences in the microbial communities of seawater within 5 cm of reef macro-organisms compared to seawater sampled 3 m off of the reef except for the coral *M. braziliensis*, similar to the study conducted by Silveira et al. (2017). Finally, a recent study used syringes (50 mL volume) to sample seawater 0, 5, and 50 cm away from individual colonies of *Acropora* and *Platygyra* spp. and reported that specific coral-associated bacteria were more abundant closer to *Acropora* (0 and 5 cm) compared to *Platygyra* colonies, attributing this to morphological differences between the corals that could impact momentum boundary layer dynamics and mixing processes (Ochsenkuhn et al. 2018).

In all cases, these investigations targeted their sampling within the diffusive or momentum boundary layer, the area in which microbial dynamics may be distinct from the overlying water column. However, most of the previous studies (Tout et al. 2014; Silveira et al. 2017; Walsh et al. 2017) did not sample smaller volumes of seawater (<10 L) that may be more

relevant for examining the momentum boundary layer surrounding corals, which has an approximate thickness of a few centimeters (Shashar et al. 1996). Additionally, these previous studies collected samples surrounding one or two different coral species at a single reef site and could not investigate if local environmental conditions influenced the microbial community composition of this seawater. While these efforts have brought attention to the importance of this near coral seawater environment, there is still a need to examine the microbial interactions at a higher resolution by collecting smaller volumes of seawater as well as by investigating the influences that different coral species or reef locations impart on near coral seawater microbial communities.

We designed this study to explore the hypotheses that 1) near coral seawater environments harbor taxonomically and functionally distinct microbial communities compared to the overlying water column and 2) that near coral seawater is also distinct by coral species. To test these hypotheses, we examined microbial communities collected using small (1 mL and 60 mL) volume seawater samples from distances generally thought to include the momentum boundary layer surrounding individual coral colonies (Shashar et al. 1996; Barott and Rohwer 2012). We compared these microbial communities to reef seawater microbial communities collected from within the benthic boundary layer across multiple reefs.

3.3 Experimental Procedures

Sampling design and sample collection

Seawater was collected near corals (≤ 30 cm away) as well as farther from corals (>1 m off the reef) at ten reefs during two separate field expeditions to the Cuban reef-systems of Jardines de la Reina (JR) and Los Canarreos (CAN) in February and April/May of 2015 (Supporting

Information Figure S1). Most reefs within JR lie within a marine protected area and they are some of the most protected and preserved reefs in the Caribbean. The surveyed reefs in JR included forereefs (JR1 and 2) that are located on the southern side of the reef tract as well as back reefs (JR 3, 4, 5, 6) that are located within the gulf of Ana Maria, lying between the island of Cuba and the reef tract (Supporting Information Figure S1). JR reefs include a variety of habitats and hydrodynamic regimes (i. e. tidal currents and wave exposure), contributing to microbiological differences between these reefs (Weber et al., unpublished). Reefs within Los Canarreos were hydrogeographically similar to each other and did not span distinct environmental gradients (Supporting Information Figure S1).

Seawater samples were collected near five species of coral (coral seawater, CSW) within distances thought to comprise the lower (<10 cm) and upper (30 cm) bounds of the momentum boundary layer surrounding individual coral colonies (Shashar et al. 1996; Barott et al. 2012). The corals *Orbicella faveolata* (Ellis and Solander, 1786), *Montastraea cavernosa* (Linnaeus, 1767), *Pseudodiploria strigosa* (Dana, 1846; formerly known as *Diploria strigosa*), and *Porites astreoides* (Lamark, 1816) were chosen because they are commonly observed on Cuban reefs and the first three species are major reef-builders in the Caribbean. *Acropora cervicornis* (Lamarck, 1816) was selected because this species was historically a major reef-builder on Caribbean reefs. However, disease outbreaks have decimated *Acropora* populations and this species is now listed as threatened under the U.S. Endangered Species Act.

To sample the near coral seawater (CSW) for genomic analyses, a scuba diver used sterile 60 mL syringes to collect seawater 30 cm away from at least three colonies of at least three species on each reef (Table 1). Colonies were generally isolated from other corals and colony replicates were separated by more than 2 m across each reef. We did not collect CSW

from colonies that were actively shedding their mucus in order to avoid potential mucus

contamination. The natural distribution of coral species varied between reefs therefore some

species (e. g. *P. astreoides*) were sampled more than others (e.g., *A. cervicornis*). In total, 49

CSW samples were collected across the two reef-systems, but six were removed from the final

analysis because they had low numbers of sequences.

Additionally, smaller volume (1 mL) seawater samples were collected via syringe to

examine general trends in microbial abundances along a gradient towards coral colonies and to

complement the genomic analyses. Two distances (0 and 5 cm) were sampled around each

colony in JR and three distances (0, 2, and 30 cm) were sampled around each colony in CAN.

We increased the number of sampling distances per colony for corals sampled in CAN so that we

could more comprehensively evaluate how microbial abundance changed over a small-scale

distance gradient from each colony. Additionally, control seawater samples were collected over

sand-patches (ranging from 0 to 30 cm away) at JR 6 and CAN 12, 14, 15, and 17. We compared

cell abundances in these control samples with cell abundances in CSW to investigate if the

presence of corals influenced the abundance of cells. Sampling distances were measured by

using the length of a custom syringe sampling device holder. Each sample was preserved with

1% paraformaldehyde (Electron Microscopy Sciences, Hatfield, Pennsylvania, USA) (final

concentration) and flow cytometry was used to quantify picoeukaryotes, *Prochlorococcus*,

Synechococcus, and unpigmented (heterotrophic) cells (Supporting Information).

Reef seawater samples (RSW) were collected in duplicate from >1 m above each reef at

approximately the same time as when the CSW samples were collected (Table 1). While

collecting paired RSW and CSW samples would have been ideal, we had limited bottom time to

collect paired samples using our syringe sampling approach and opted to integrate the RSW

samples per reef based on previous observations of relatively high similarity in RSW microbial

communities across individual reefs (Apprill et al. 2016). To collect RSW samples, seawater was

pumped to the surface from reef depth (>1 m off of the reef substrate) with a groundwater pump

(Mini-monsoon sampling pump, Proactive Environmental Products, Hamilton, New Jersey,

USA). We rinsed the acid-cleaned plastic tubing with reef-depth seawater for 30 seconds and

then collected 4.2 L of the seawater into acid-cleaned plastic bottles (for amplicon sequencing)

or duplicate 10 L acid-washed bottles (for metagenome sequencing). All samples were kept cold

in a cooler filled with ice until they were processed.

To filter RSW, the acid-cleaned tubing was rinsed with seawater and then duplicate 2 L

samples of seawater were filtered onto 0.22 μm, 25 mm Supor® filters (Pall Corporation, Ann

Arbor, Michigan, USA) using peristalsis. Hand filtration was used to filter the CSW samples

using the same filters. Additionally, 20 L of seawater from sites JR 2, 4, 5, and 6 were each

filtered onto 0.22 μm, 142 mm Supor® filters (Pall Corporation, Ann Arbor, Michigan, USA) in

order to concentrate microbial biomass for shotgun metagenomic sequencing (Table 2). This

seawater was not pre-filtered. All filters were flash frozen in liquid nitrogen, shipped back to the

U. S. in a charged dry shipper, and then stored at -80° C until DNA was extracted.

Amplicon sequencing of coral and reef seawater DNA

DNA was extracted from RSW and CSW filters using two extraction methods that were

performed sequentially, a modified sucrose-lysis extraction protocol (Santoro et al. 2010) as well

as a simplified phenol-chloroform extraction (Urakawa et al. 2010), to increase cellular lysis

efficiency and DNA yield (Supporting Information). The Genomic DNA Clean and Concentrator

kit (Zymo Research Corporation, Irvine, CA, USA) was used to combine the purified DNA

extracts yielded from both methods. Additionally, DNA extraction (n=2) and DNA pooling

(n=2) controls as well as a microbial mock community (HM-278D, BEI Resources, Manassas, VA, USA) were prepared to account for potential DNA extraction contamination as well as amplification and sequencing error.

The nucleic acids were submitted to the W. M. Keck Center for Comparative and Functional Genomics (University of Illinois, Urbana, IL) where V4 region SSU rRNA genes from bacteria and archaea were amplified using the Fluidigm® microfluidics quantitative PCR platform and prepared for 2x250 bp paired-end Illumina MiSeq sequencing (Supporting Information). These primers were chosen to specifically amplify 16S rRNA genes from bacteria and archaea based on their fairly comprehensive coverage of prokaryotes, their frequent use in marine studies, and their size for sequencing on an Illumina MiSeq (Kozich et al. 2013; Apprill et al. 2015; Parada et al. 2016). Furthermore, we used 515F-Y and 806R-B primers with degeneracies to minimize known amplification biases against Crenarachaeota/ Thaumarchaeota (Parada et al. 2016) and the SAR 11 clade (Apprill et al. 2015), taxa that are both found in marine microbial communities. The primer-sorted and demultiplexed reads were screened for quality using mothur v.1.36.1 (Schloss et al. 2009) (Supporting Information). The sequences were then subsampled to 8,500 reads per sample in order to minimize the impacts of uneven sequence coverage across samples, but retain as many samples within the dataset as possible. All of the raw sequences used for this analysis were deposited into the NCBI Sequence Read Archive (SRA) under BioProject PRJNA422534.

Reads were clustered into homogenous groups using Minimum Entropy Decomposition (MED), a program that selects specific information-rich nucleotide positions in the reads using Shannon entropy and decomposes these sequences into different groups referred to as MED nodes (Eren et al. 2015). The MED algorithm identifies closely related, but biologically distinct

organisms (MED nodes) using marker gene information and is valuable for examining patterns in microbial diversity that could be overlooked if sequences are grouped based on a lower level of similarity (Eren et al. 2015). We chose to use the MED algorithm because it has been used to examine microbial community diversity in coral tissue and seawater environments and can discern between closely related, but potentially ecologically distinct microorganisms (Neave et al. 2017; Ward et al. 2017; Weber et al. 2017). Sequences representing each MED node were classified in mothur using the Silva v119 database ('knn' method) (Pruesse et al. 2007). All MED node representative sequences were also re-aligned using the SINA alignment and taxonomic service (Quast et al. 2013) to verify taxonomic assignment of the reads (SILVA reference database v. 128). Sequences representing *Endozoicomonas* MED nodes were compared to each other and aligned using the NCBI BLASTN 2.8.0+ algorithm (Zhang et al. 2000) in order to investigate their similarity to each other as well as their similarity to other reported *Endozoicomonas* sequences.

Statistically significant enrichment comparisons of MED nodes between CSW and RSW were made using the differential expression package 'DESeq2' (Love et al. 2014) following previous methods (McMurdie and Holmes 2014; Neave et al. 2017) (Supporting Information). Within-site enrichment comparisons were conducted at each reef location in order to minimize geographic and depth-related variability. Enrichment tests were only completed for samples collected at JR reefs 1, 2, 5, and 6 because a majority of the CAN RSW samples were removed due to low sequence quality.

Microbial community visualization and statistical analyses were accomplished using several R packages (R Core Development Team 2017). To examine the similarity between RSW communities sampled across JR, we completed a cluster analysis (method = 'average linkage')

94

on the Bray-Curtis dissimilarity matrix using the 'Pvclust' R package (Suzuki and Shimodaira 2006). We also examined the similarity between CSW communities separately using the same method. Non-metric multidimensional scaling (NMDS) analysis was completed with the 'vegan' package (Oksanen et al. 2017) using the square-root transformed Bray-Curtis dissimilarity matrix to visually compare the degree of similarity between the CSW and RSW bacterial and archaeal communities. Nested Permutational Multivariate Analysis of Variance tests using distance matrices (PERMANOVA/Adonis) (Oksanen et al. 2017) were performed on the Bray-Curtis dissimilarity index using 999 permutations to determine the degree to which the different factors explained the microbial community composition of the samples ($p < 0.05$) (Supporting Information). Most of the nested PERMANOVA comparisons were completed using the CSW and RSW collected within JR. However, *P. astreoides* CSW was also sampled across three sites in CAN (12, 15, and 17) so we included these samples and one RSW sample (collected from CAN 15) in the NMDS and PERMANOVA tests for this species. We also collected CSW samples from *P. strigosa* and *O. faveolata* in CAN, but were unable to use these samples in the NMDS and PERMANOVA tests because too many sequences were removed during quality-filtering.

Shotgun metagenomic sequencing

We combined 3-5 CSW DNA extracts per species across samples collected within Jardines de la Reina and prepared the pooled mixtures for shotgun metagenomic sequencing (Table 2). Samples were pooled in order to increase the total concentration of DNA in each coral seawater sample. We recognize that this is not an ideal approach, but were concerned that the separate extracts were too diluted to be sequenced individually. DNA extractions were also performed on one half of each of four RSW metagenome filters (representing 10 liters of reef-depth seawater

sampled at JR sites 2, 4, 5, and 6) using a modified cetyl-trimethylammonium bromide (CTAB)-phenol: chloroform: isoamyl alcohol extraction and isopropanol precipitation (Table 2, Supporting Information).

Library preparation and sequencing of the pooled CSW and RSW DNA samples were completed at the W. M. Keck Center. Libraries were prepared using the Hyper Library construction kit (Kapa Biosystems, Inc, Wilmington, MA, USA) and sequenced using 2x150 bp paired-end Illumina HiSeq 4000 sequencing. The raw sequences used for this analysis were deposited into the NCBI Sequence Read Archive (SRA) under BioProject PRJNA422534. Fastq files were demultiplexed and library adaptors were trimmed from the 3' ends of the reads (Supporting Information). BBTools (Bushnell 2016) was used to quality-filter and prepare the raw metagenomic reads for functional analysis (Supporting Information).

The Functional Mapping and Analysis Pipeline for metagenomics and metatranscriptomics (FMAP) (Kim et al. 2016) was used to annotate genes with the mapping program DIAMOND (Buchfink et al. 2014) against the UniRef100 database (uniprot.org), calculate KEGG (Kyoto Encyclopedia of genes and genomes) gene abundances (Kanehisa et al. 2016), and identify significantly differentially abundant KEGG orthologs (KOs), pathways, and modules between CSW and RSW (Kruskal – Wallis test, Fisher's exact test, p-value < 0.05, FDR adjusted to control for false positives; Supporting Information).

Sampling volume comparisons

Because we collected seawater samples for genomic analysis over a range of volumes (60 mL to 2 L), we conducted a separate experiment to test if initial seawater sampling volume influenced alpha- and beta-diversity comparisons in seawater microbial communities. To do this, we collected replicate 60 mL, 1.5 L, or 2 L samples from surface reef seawater at two different sites

96

in St. John, US Virgin Islands (Table 1). We sequenced and analyzed these samples

independently to validate our analysis of the CSW and RSW samples collected in Cuba

(Supporting Information).

3.4 Results

Sample volume comparisons

Analyses of the SSU rRNA gene amplicon sequences from the seawater volume experiment

showed that samples of larger volume (1.5 or 2 L) had greater microbial species richness

compared to smaller volume (60 mL) samples (Supporting Information, Figures S7-S11).

However, sampling volume was not found to impact comparisons of beta diversity or enrichment

analysis (Supporting Information, Figures S12-S13). Based on these results, further comparisons

of alpha diversity were not made between Cuban CSW and RSW.

Coral seawater microbial communities are influenced by reef and coral species

Analysis of SSU rRNA gene sequences showed that RSW microbial communities from JR were

37 – 84% similar in terms of bacterial and archaeal community composition whereas CSW

microbial communities from JR were more similar to each other (51 – 84% similarity) (see

Supporting Information Figure S2 for class-level relative abundances). Individual non-metric

multidimensional scaling (NMDS) analyses of amplicon sequences by coral species

demonstrated that CSW communities generally separated from the RSW communities, although

there was a degree of overlap with RSW especially for *P. astreoides* CSW (Figure 1A-E). Nested

PERMANOVA (Adonis) tests on the amplicon sequence data confirmed that both sample type

(CSW vs. RSW) and reef site were significant determinants of community similarity (Figure 1A-

E). A NMDS including all CSW and RSW microbial communities revealed overlapping

community composition between RSW and CSW, with some distinction by species as indicated by the covariance ellipses (Figure 1F). Additionally, a nested PERMANOVA (Adonis) test completed on all CSW and RSW communities within JR demonstrated that both reef location and coral species significantly influenced microbial community structure (Figure 1F).

Differential enrichment analyses of the MED clustered amplicon sequences revealed that CSW microbial communities were distinct from RSW microbial communities with regard to specific bacterial taxa. Broadly, CSW was mostly enriched with copiotrophic lineages of Gammaproteobacteria when compared to RSW collected within JR (Table 3). MED enrichment in *P. astreoides* CSW compared to RSW was attributed to the Gammaproteobacteria genera *Alteromonas*, *Endozoicomonas*, and *Bermanella* (Supporting Information Table S1). *Endozoicomonas* MEDs were significantly enriched in *P. astreoides* CSW at reefs JR 2, 5, and 6. *Endozoicomonas* and *Alteromonas* were also enriched in *P. strigosa* coral seawater (Supporting Information Table S1). Similarly, *O. faveolata* CSW was enriched with *Alteromonas* as well as two MED nodes identifying as *Pyschrobacter* (Supporting Information Table S1). *Marinobacter* was enriched in CSW from corals collected from JR 1, but not enriched at the other reefs (Supporting Information Table S1). Additionally, non-Gammaproteobacteria taxa frequently identified in nutrient rich or sediment environments were significantly enriched in CSW, including *Propionigenium*, unclassified Bacillales, *Chitinophagaceae*, Deltaproteobacterial OM27 clade, *Owenweeksia*, and *Erythrobacter* (Supporting Information Table S1). RSW from JR was generally significantly enriched with MED nodes classifying as microbial taxa that are found within free-living seawater microbial communities, including Rhodobacteraceae, the ultra-small "*Candidatus* Actinomarina", SAR11, SAR86, and SAR116 clades, and AEGEAN-169 and NS5 marine groups (Supporting Information Table S1).

Diverse Endozoicomonas *bacteria associate with* P. astreoides *seawater*

Using the amplicon sequence data, we detected seven *Endozoicomonas* MED nodes in CSW,

demonstrating *Endozoicomonas* genotype diversity within the CSW (Figure 2). Two

Endozoicomonas MED nodes, MED3416 and 798, had the highest relative abundance in *P.*

astreoides CSW across JR and CAN (Figure 2). We compared the 16S rRNA gene sequence

similarity for these two MED nodes with other 16S rRNA genes in NCBI and found that the

MED3416 sequence was 98% similar to *Parendozoicomonas haliclonae*, a bacterial isolate from

a marine sponge (NCBI sequence ID: NR_157681.1) and 96% similar to *Endozoicomonas*

euniceicola (NCBI sequence ID: NR_109684.2), *E. numazuensis* (NCBI sequence ID:

NR_114318.1), and *E. montiporae* (NCBI sequence ID: NR_116609.1). The MED798

representative amplicon sequence was 96% similar to *Endozoicomonas* cultures isolated from

gorgonians and *E. montiporae*, as well as an isolate from the sea slug, *E. ornata* (NCBI sequence

ID: NR_041264.1). We also compared the representative *Endozoicomonas* sequences to each

other and found that some of the most abundant MED nodes detected in *P. astreoides* CSW,

MED3416, 798, 810, and 832, were 99% similar to each other whereas MED nodes detected at

lower relative abundances in *P. strigosa* and/or *O. faveolata* CSW were less similar (MED3145,

98%; MED2581, 96%; MED1451, 95% similar). In general, the relative abundances of

Endozoicomonas MED nodes were low in RSW (ranging from no detection to 1.7% relative

abundance).

Genomic evidence of surface-attached and dynamic microbial communities within coral

seawater

99

Comparisons between the pooled CSW and reef-depth RSW metagenomes revealed 1058

differentially abundant genes (Figure 3). CSW metagenomes were significantly enriched in

genes involved in 15 KEGG pathways (Table 3) and six KEGG modules (Table 5). The two-

component system was the most significantly enriched pathway in CSW (Table 3) and included

genes involved in cell-cycle and biofilm response regulation, signal transduction (histidine

kinsases), as well as chemotaxis (Supporting Information Table S2). The other enriched

pathways within CSW metagenomes included bacterial chemotaxis, flagellar assembly, biofilm-

formation (in *Pseudomonas aeruginosa* and *Vibrio cholera*), bacterial secretion systems, ABC

transporters, the *Caulobacter* cell cycle, lipopolysaccharide biosynthesis, nitrogen metabolism,

pentose and glucuronate interconversions, cationic antimicrobial peptide resistance, geraniol

degradation, and glycan degradation (Table 3).

The KEGG modules enriched within CSW included type II and IV secretion systems,

denitrification, the dipeptide transport system, and the CheA-CheYBV chemotaxis and PleC-

PleD cell fate control two-component regulatory systems (Table 4). The CSW metagenomes also

differed from each other by coral species (Figure 3), but the most abundant KEGG pathways

were the same.

Trends in microbial cell abundance over a distance gradient from the corals

Microbial cell abundances sampled over a distance gradient from each colony were highly

variable by coral species, microbial group, and reef (Figure 4 and Supporting Information

Figures S3-S6). For *P. astreoides* and *P. strigosa*, abundances of microorganisms generally

increased away from the colonies (Figure 4 and Supporting Information Figures S3-S6). This

trend was also observed for picoeukaryotes and unpigmented cells surrounding *M. cavernosa*

colonies (Figure 4 and Supporting Information Figures S5-S6). For *A. cervicornis*, the

abundance of picocyanobacteria (*Prochlorococcus* and *Synechococcus*) increased with distance

from the colonies, but picoeukaryotes and unpigmented cells displayed the opposite trend

(Figure 4 and Supporting Information Figures S3-S4). For *O. faveolata*, cell abundances from all

groups decreased with distance from the colonies, except for *Synechococcus* (Figure 4 and

Supporting Information Figures S3-S6). For the sand seawater controls in CAN, the increases

and decreases within microbial groups generally followed the trends observed for the corals

(Figure 4 and Supporting Information Figures S3-S6). The overall cell abundance of the different

microbial groups was related to reef location, with very strong site-specificity for

Prochlorococcus (Supporting Information Figure S3).

3.5 Discussion

In this study, we used genomics to determine that coral seawater microbial communities are

influenced by coral species and reef site. More specifically, we detected enrichment of

copiotrophic bacterial taxa and genes indicative of potential mobile, surface-attached, and

virulent microbial lifestyles within coral seawater compared to reef-depth seawater. Microbial

cell abundances collected along a gradient from coral surfaces were variable, but influenced by

coral species, reef site, and cell type. Overall, these results provide taxonomic and functional

genomic support for the existence of an environment that we term the 'coral ecosphere', a

distinct and dynamic environment for microorganisms that forms surrounding individual coral

colonies and that may serve as an interaction zone between the coral surface and the overlying

seawater. In Figure 5, we present a conceptual diagram of the microbial functions, potential

interactions, and bacterial taxa that are enriched within the coral ecosphere compared to the

surrounding seawater.

Enrichment of primary colonizers within the coral ecosphere may be influenced by coral-derived organic matter

We detected enrichment of several copiotrophic Gammaproteobacteria, including the genera *Endozoicomonas, Bermanella, Marinobacter,* and *Alteromonas,* within coral ecospheres. Several of these taxa have been commonly associated with corals and coral-derived organic matter (OM) (Nelson et al. 2013) and Gammaproteobacteria are typically early colonizers of marine surfaces (Dang and Lovell 2000; Sweet et al. 2011). *Endozoicomonas* are an established tissue and mucus symbiont of corals globally (Apprill et al. 2016; Glasl et al. 2016; Neave et al. 2017; Pollock et al. 2018) and our results extend the current knowledge of *Endozoicomonas* biogeography by indicating that *Endozoicomonas* may reside in the seawater surrounding corals (specifically *P. astreoides*). The other enriched bacteria, including members within the genus *Bermanella* and the order *Alteromonadales,* have previously been found in association with coral-derived particulate and dissolved OM including coral tissue homogenates (Randall et al. 2016), coral mucus (Sweet et al. 2011), the seawater close to corals (Tout et al. 2014), and within natural reef seawater cultures inoculated with coral mucus and exudates (Allers et al. 2008; Nelson et al. 2013).

The enrichment of copiotrophic groups in coral ecospheres compared to the reef seawater is also paired with enrichment of specific metabolic pathways involved in the cycling of OM. Genes used in the denitrification pathway were significantly enriched in coral ecospheres compared to reef seawater, possibly suggesting that the ecosphere environment is populated by anoxic or microaerobic regions where denitrification occurs, aligning with results of other studies that have investigated oxygen dynamics close to corals (Barott et al. 2012; Wangpraseurt et al. 2012; Haas et al. 2013). In addition, corals exude amino acids and other dissolved organic

nitrogen into the water column (Schlichter and Liebezeit 1991; Tanaka et al. 2009) and we

detected enrichment of dipeptide transport system genes (ABC transporter) within ecospheres,

suggesting that amino acid uptake could be an important source of nitrogen for microorganisms

surrounding corals. *Alteromonas*, a genus shown to dominate natural seawater assemblages after

the addition of dissolved OM produced by microbial communities fueled with nitrate and

ammonium (Goldberg et al. 2017), was also enriched in a majority of the ecospheres, further

suggesting that coral – derived OM may be influencing community composition within the

ecosphere. Our genomic evidence suggests that ecosphere microbial composition may be

influenced by the input of coral-derived OM and that microbial metabolisms within the

ecosphere may be important for recycling and transferring this OM into the water column

(Figure 5).

Furthermore, the variability and lack of consistent trends in cell abundance suggest that

interactions between multiple processes (including grazing or advection of cells) may mask

influences of coral-derived OM on overall growth of planktonic microorganisms surrounding

corals. It could also be that specific taxa, rather than the cell types we counted, respond to these

coral-derived exudates and that these subtle responses cannot be detected using more coarse

changes in microbial cell abundances. However, we did observe that microbial abundance was

influenced by both coral species and reef location, reflecting the factors that influenced microbial

community composition and suggesting that coral species influence these cell populations.

Microbes within the coral ecosphere are specialized for colonization and interaction with hosts

Coral ecospheres were enriched with microbial pathways characteristic of biofilm-forming,

surface-attached, and potentially virulent microbial communities (Figure 5). The two-component

system pathway was the most significantly enriched pathway within coral ecospheres and

included genes involved in response regulation, cell-cycles, signal transduction, and chemotaxis. Genes used in the two-component system are found in bacteria, archaea, and eukarya, but are most abundant in gram-negative bacteria and cyanobacteria (Capra and Laub 2012). The two-component signal transduction system permits bacteria to sense and respond to external stimuli (Capra et al. 2012), and is also involved in the development of virulence and antimicrobial resistance (Gooderham and Hancock 2009). Enrichment of the two-component system suggests that cells within the coral ecosphere may be able to rapidly respond to changes in this fluctuating marine environment. For example, bacteria may respond to an environmental cue by transcribing virulence genes that enable them to colonize a host and potentially cause disease (Ribet and Cossart 2015). Furthermore, the type II and IV bacterial secretion systems were also enriched within coral ecosphere metagenomes and these systems are typically used by bacteria to colonize surfaces, transport and secrete molecules, induce endocytosis within the host cell, acquire virulence genes, and disrupt host cell defenses (Kohler and Roy 2015; Green and Mecsas 2016). The enrichment of secretion systems near corals suggests that these infection strategies may be used by putative pathogens as well as symbionts residing within the ecosphere microbial community to colonize the coral host.

Additionally, the prevalent KEGG pathways and modules detected within coral ecospheres suggest that specific taxa within these microbial communities have the capacity to exhibit chemotaxis, transport solutes, as well as produce, secrete, and resist antibiotics (Figure 5). Many of the enriched coral ecosphere genes are also classified as interaction genes (Torto-Alalibo et al. 2009; Cardenas et al. 2018), genes that permit microorganisms to colonize and interact with hosts (Dale and Moran 2006). In support of our hypothesis, Tout and colleagues (2014) detected elevated abundances of these interaction genes, including bacterial chemotaxis

and motility, membrane transport, and cell signaling genes, within the seawater close to the corals *Acropora aspera* and *A. palifera* (Tout et al. 2014). Additionally, Walsh et al. (2017) detected enrichment of genes used for antibiotic resistance, resistance to toxic compounds (methicillin resistance), and motility and chemotaxis in the seawater adjacent to *Mussimilia braziliensis* (Walsh et al. 2017).

Furthermore, there are similarities between the potential microbial metabolic pathways detected within the coral ecosphere and coral tissue. Bacterial, archaeal, and fungal genes used to catalyze different conversions within the nitrogen cycle are commonly found in coral tissue metagenomes (Wegley et al. 2007; Vega Thurber et al. 2009; Kimes et al. 2010; Garcia et al. 2013) and we detected an enrichment of denitrification genes within the coral ecosphere. Metal tolerance and antimicrobial resistance genes as well as virulence genes have also been identified in coral tissue metagenomes (Vega Thurber et al. 2009; Kimes et al. 2010; Garcia et al. 2013) as well as in our study. In contrast, oxidative stress genes were more regularly detected in coral tissue metagenomes (Wegley et al. 2007; Vega Thurber et al. 2009) whereas motility and chemotaxis genes and genes used in the secretion of lipopolysaccharides and for biofilm formation were more commonly detected in coral ecosphere metagenomes.

The coral ecosphere may be a reservoir for potential coral symbionts and pathogens

Our results suggest that corals are bathed in microbial cells that are capable of colonizing and interacting with the coral surface. As such, the coral ecosphere may serve as a reservoir for coral symbionts or pathogens. In support of this hypothesis, we detected a prevalent association between the coral *P. astreoides* and *Endozoicomonas* bacteria within the coral ecosphere at sites JR 2, 5, and 6 using differential enrichment tests. *Endozoicomonas* MED nodes were also detected in *P. astreoides* ecosphere samples at JR 1, but were not significantly enriched relative

to RSW after p-value corrections for multiple testing were applied. Nevertheless, this association suggests that either *Endozoicomonas* cells reside in the seawater and are attracted (i.e. through some chemical cue) to the coral surface or that *Endozoicomonas* cells are shed from the coral mucus or tissue. There is evidence supporting both of these hypotheses. *Endozoicomonas* genomes are fairly large (>5 Mbp) and equipped with genes required for degrading amino and nucleic acids (Neave et al. 2014; Neave et al. 2017) as well as genes coding for enzymes that are used to degrade testosterone and glycosidic bonds (named Endo-AEmo) in glycoproteins (Ding et al. 2016). Ding and colleagues (2016) suggested that *Endozoicomonas* may be able to attach to the coral mucus layer, penetrate the mucus using the Endo-AEmo enzyme, and then enter the host tissue via endocytosis. Alternatively, *Endozoicomonas* genes may reside within the ecosphere because they have been shed from coral tissue and mucus. *Endozoicomonas* was identified as a dominant member of the newly formed communities in *P. astreoides* mucus (Glasl et al. 2016) and as cells die and mucus sloughs off into the water column, *Endozoicomonas* cells may be shed into the ecosphere environment. In addition to their putative roles as common coral symbionts, *Endozoicomonas* cells residing in the *P. astreoides* coral ecosphere may influence the ecosphere chemically through the production of extracellular superoxide. This ubiquitous molecule can be found in the coral ecosphere of *P. astreoides*, is produced by *Endozoicomonas* isolates, and likely plays important roles in bacterial interactions and coral health (Diaz et al. 2016; Zhang et al. 2016).

The exogenous recruitment of specific bacteria to corals has been investigated for larvae (Sharp et al. 2010; Apprill et al. 2012), but is still unresolved for adult colonies. Sweet and colleagues (2011) touched on this topic when they proposed that specific bacteria were recruited from the water column into the developing coral mucus biofilm through some selective process

or direct contact with another surface. Our study extends this hypothesis by demonstrating that some of the primary mucus colonizers detected by Sweet and colleagues (2011) were also enriched within the coral ecospheres. Detailed exploration of microbial interactions within the coral ecosphere will deepen our understanding of which microbes are available to the corals to serve as potential symbionts and how the coral host, as well as the external environmental conditions, influence these microbial interactions.

Considerations for studying the coral ecosphere

In our comparisons, the reef of collection was also identified as an important predictor of microbial community composition. This finding suggests that local environmental conditions, like current direction and speed, temperature, light, and nutrient availability, may also influence microbial growth and community composition within the ecosphere. For example, seaward reef locations within Jardines de la Reina are exposed to stronger currents (up to 40 cm s^{-1}) on average compared to locations within the Gulf of Ana Maria (13 cm s^{-1}) (Arriaza et al. 2008) and these conditions likely influence the flux of cells and nutrients within the coral ecosphere. Future studies could investigate the connection between water flow and microbial dynamics within coral ecospheres. The variability in microbial community similarity between RSW samples collected within Jardines de la Reina also corroborates the strong influence of reef-specific environmental conditions, but is within the range of variability observed in seawater microbial communities at smaller geographic scales (Apprill et al. 2016). Future studies of the dynamic environment of the coral ecosphere should undoubtedly account for reef-specific variation by collecting more ecosphere samples (biological replicates) at each site for every species surveyed. Additionally, while we were able to discern differences between CSW and RSW in this study, the magnitude

of this distinction may be greater if paired CSW and RSW samples were collected at a variety of different reefs.

Continued research into microbial interactions within the coral ecosphere requires recognition of potential methodological biases and improvements to these methods, contingent upon the available technology. We used 60 mL syringes to sample seawater from the coral ecosphere so that we could gently aspirate seawater close to each colony in a controlled manner. We then compared these ecosphere microbial communities to reef seawater communities that were sampled with larger volumes using a groundwater pump. We recognize that differences in sampling method could have led to potential biases in our analyses so we have made an effort to address the realized and potential impacts of these differences. First, we completed a field – based seawater volume experiment to understand how differences in sampling volume impacted microbial community analysis and found that sampling volume did not influence metrics of beta diversity or enrichment analysis. Secondly, while we cannot directly examine how sampling method (syringe vs. groundwater pump) would influence our results, we postulate that cells could have been exposed to different physical stressors or grazing pressures when they were sampled with different techniques. That being said, we have used groundwater pumps to collect samples for flow cytometry and for microbial respiration experiments in the past and have no reason to believe that this method is shearing cells or collecting water in a manner that makes these collections incomparable to syringe-based collection methods. Furthermore, we stored the samples on ice immediately after they were harvested to reduce the influence of grazing or altered growth dynamics within the collection containers. Even if growth and grazing did occur, these processes would be unlikely to influence our results because the average doubling time of microbial cells within reef seawater at ambient reef temperature is about a day (McNally et al.

2017) and zooplankton abundances are relatively low in reef seawater, present on the order of 0.06 zooplankter l⁻¹ (0.004 zooplankton within 60 mL) (Cox et al. 2006). Additionally, after sample collection, we also made efforts to standardize the concentrations of DNA used in PCR reactions in order to minimize the impact of sample volume and collection method. Lastly, we used conservative data analysis (sub-sampling and quality-filtering) and multivariate statistical approaches (e.g., Bray-Curtis dissimilarity, NMDS, PERMANOVA) to analyze the data.

3.6 Conclusions

We have shown that five reef-building coral species were surrounded by a distinct microbial environment, the coral ecosphere, that in turn was influenced by local environmental conditions at each reef. This coral ecosphere supports taxonomically and functionally distinct microbial communities and constitutes a dynamic seawater habitat harboring cells that seem capable of interacting with the coral surface. Recognition of the coral ecosphere provides new opportunities to study coral-microbial interactions within the water column and exogenous recruitment of microorganisms, including pathogens, to colonies. Future directions in coral ecosphere research include understanding the ecosphere microbial community variability in the context of changing environmental conditions, documenting how cells within the coral ecosphere use coral-derived organic matter, and exploring the significance and contribution of these interactions to biogeochemical cycling on coral reefs.

3.7 Acknowledgements

3.8 References

Allers, E., Niesner, C., Wild, C., and Pernthaler, J. (2008) Microbes enriched in seawater after addition of coral mucus. *Appl Environ Microb* **74**: 3274-3278.

Apprill, A., Weber, L.G., and Santoro, A.E. (2016) Distinguishing between microbial habitats unravels ecological complexity in coral microbiomes. *mSystems* **1**: e00143-00116.

Apprill, A., Marlow, H.Q., Martindale, M.Q., and Rappe, M.S. (2012) Specificity of associations between bacteria and the coral *Pocillopora meandrina* during early development. *Appl Environ Microbiol* **78**: 7467-7475.

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.

Arriaza, L., Simanca, J., Rodas, L., Lorenzo, S., Hernandez, E., Milian, D., and Romero, P. (2008) Corrientes marinas estimadas en la plataforma suroriental Cubana. *Serie Oceanológica* **4**.

Barott, K.L., and Rohwer, F.L. (2012) Unseen players shape benthic competition on coral reefs. *Trends Microbiol* **20**: 621-628.

Bourne, D.G., Morrow, K.M., and Webster, N.S. (2016) Insights into the coral microbiome: underpinning the health and resilience of reef ecosystems. *Annual Reviews Microbiology* **70**: 317-340.

Buchfink, B., Xie, C., and Huson, D.H. (2014) Fast and sensitive protein alignment using DIAMOND. *Nature Med* **12**: 59–60.

Bushnell, B. (2016) BBTools Software package. In. BBTools - DOE Joint Genome Institute.

Capra, E.J., and Laub, M.T. (2012) Evolution of two-component signal transduction systems. *Annu Rev Microbiol* **66**: 325-347.

Cardenas, A., Neave, M.J., Haroon, M.F., Pogoreutz, C., Radecker, N., Wild, C. *et al.* (2018) Excess labile carbon promotes the expression of virulence factors in coral reef bacterioplankton. *ISME J* **12**: 59-76.

Chamberlain, J.A., and Graus, R.R. (1975) Water-flow and hydromechanical adaptations of branched reef corals. *B Mar Sci* **25**: 112-125.

Cox, E.F., Ribes, M., and Kinzie III, R.A. (2006) Temporal and spatial scaling of planktonic responses to nutrient inputs into a subtropical embayment. *Mar Ecol Prog Ser* **324**: 19-35.

Dale, C., and Moran, N.A. (2006) Molecular interactions between bacterial symbionts and their hosts. *Cell* **126**: 453-465.

Dang, H., and Lovell, C.R. (2000) Bacterial primary colonization and early succession on surfaces in marine waters as determined by amplified rRNA gene restriction analysis and sequence analysis of 16S rRNA genes. *Appl Environ Microbiol* **66**: 467-475.

Diaz, J.M., Hansel, C.M., Apprill, A., Brighi, C., Zhang, T., Weber, L. *et al.* (2016) Species-specific control of external superoxide levels by the coral holobiont during a natural bleaching event. *Nat Commun* **7**: 13801.

Ding, J.Y., Shiu, J.H., Chen, W.M., Chiang, Y.R., and Tang, S.L. (2016) Genomic insight into the host-endosymbiont relationship of *Endozoicomonas montiporae* CL-33(T) with its coral host. *Front Microbiol* **7**: 251.

Dinsdale, E.A., Pantos, O., Smriga, S., Edwards, R.A., Angly, F., Wegley, L. *et al.* (2008) Microbial ecology of four coral atolls in the Northern Line Islands. *PLoS One* **3**: e1584.

Eren, A.M., Morrison, H.G., Lescault, P.J., Reveillaud, J., Vineis, J.H., and Sogin, M.L. (2015) Minimum entropy decomposition: unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences. *ISME J* **9**: 968-979.

Garcia, G.D., Gregoracci, G.B., Santos Ede, O., Meirelles, P.M., Silva, G.G., Edwards, R. *et al.* (2013) Metagenomic analysis of healthy and white plague-affected *Mussismilia braziliensis* corals. *Microb Ecol* **65**: 1076-1086.

Garren, M., Son, K., Raina, J.B., Rusconi, R., Menolascina, F., Shapiro, O.H. *et al.* (2014) A bacterial pathogen uses dimethylsulfoniopropionate as a cue to target heat-stressed corals. *ISME J* **8**: 999-1007.

Glasl, B., Herndl, G.J., and Frade, P.R. (2016) The microbiome of coral surface mucus has a key role in mediating holobiont health and survival upon disturbance. *ISME J* **10**: 2280-2292.

Goldberg, S.J., Nelson, C.E., Viviani, D.A., Shulse, C.N., and Church, M.J. (2017) Cascading influence of inorganic nitrogen sources on DOM production, composition, lability and microbial community structure in the open ocean. *Environ Microbiol* **19**: 3450-3464.

Gooderham, W.J., and Hancock, R.E. (2009) Regulation of virulence and antibiotic resistance by two-component regulatory systems in *Pseudomonas aeruginosa*. *FEMS Microbiol Rev* **33**: 279-294.

Green, E.R., and Mecsas, J. (2016) Bacterial secretion systems: An overview. *Microbiol Spectr* **4**.

Haas, A.F., Gregg, A.K., Smith, J.E., Abieri, M.L., Hatay, M., and Rohwer, F. (2013) Visualization of oxygen distribution patterns caused by coral and algae. *PeerJ* **1**: e106.

Haas, A.F., Fairoz, M.F., Kelly, L.W., Nelson, C.E., Dinsdale, E.A., Edwards, R.A. *et al.* (2016) Global microbialization of coral reefs. *Nat Microbiol* **1**: 16042.

Houlbrèque, F., Delesalle, B., Blanchot, J., Montel, Y., and Ferrier-Pagès, C. (2006) Picoplankton removal by the coral reef community of La Prévoyante, Mayotte Island. *Aquat Microb Ecol* **44**: 59-70.

Kaandorp, J.A., Koopman, E.A., Sloot, P.M., Bak, R.P., Vermeij, M.J., and Lampmann, L.E. (2003) Simulation and analysis of flow patterns around the scleractinian coral *Madracis mirabilis* (Duchassaing and Michelotti). *Philos Trans Soc Lond B Biol Sci* **358**: 1551-1557.

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016) KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* **44**: D457-462.

Kelly, L.W., Barott, K.L., Dinsdale, E., Friedlander, A.M., Nosrat, B., Obura, D. *et al.* (2012) Black reefs: iron-induced phase shifts on coral reefs. *ISME J* **6**: 638-649.

Kelly, L.W., Williams, G.J., Barott, K.L., Carlson, C.A., Dinsdale, E.A., Edwards, R.A. *et al.* (2014) Local genomic adaptation of coral reef-associated microbiomes to gradients of natural variability and anthropogenic stressors. *Proc Natl Acad Sci U S A* **111**: 10227-10232.

Kim, J., Kim, M.S., Koh, A.Y., Xie, Y., and Zhan, X. (2016) FMAP: Functional Mapping and Analysis Pipeline for metagenomics and metatranscriptomics studies. *BMC Bioinformatics* **17**: 420.

Kimes, N.E., Van Nostrand, J.D., Weil, E., Zhou, J., and Morris, P.J. (2010) Microbial functional structure of *Montastraea faveolata*, an important Caribbean reef-building coral, differs between healthy and yellow-band diseased colonies. *Environ Microbiol* **12**: 541-556.

Kohler, L.J., and Roy, C.R. (2015) Biogenesis of the lysosome-derived vacuole containing *Coxiella burnetii*. *Microbes Infect* **17**: 766-771.

Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K., and Schloss, P.D. (2013) Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol* **79**: 5112-5120.

Love, M.I., Huber, W., and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**: 550.

McMurdie, P.J., and Holmes, S. (2014) Waste not, want not: why rarefying microbiome data is inadmissible. *Plos Computational Biology* **10**: e1003531.

McNally, S.P., Parsons, R.J., Santoro, A.E., and Apprill, A. (2017) Multifaceted impacts of the stony coral *Porites astreoides* on picoplankton abundance and community composition. *Limnol Oceanogr* **62**: 217-234.

Neave, M.J., Michell, C.T., Apprill, A., and Voolstra, C.R. (2014) Whole-genome sequences of three symbiotic *Endozoicomonas* strains. *Genome Announc* **2**: e00802-00814.

Neave, M.J., Rachmawati, R., Xun, L., Michell, C.T., Bourne, D.G., Apprill, A., and Voolstra, C.R. (2017) Differential specificity between closely related corals and abundant *Endozoicomonas* endosymbionts across global scales. *ISME J* **11**: 186-200.

Nelson, C.E., Goldberg, S.J., Wegley Kelly, L., Haas, A.F., Smith, J.E., Rohwer, F., and Carlson, C.A. (2013) Coral and macroalgal exudates vary in neutral sugar composition and differentially enrich reef bacterioplankton lineages. *ISME J* **7**: 962-979.

Ochsenkuhn, M.A., Schmitt-Kopplin, P., Harir, M., and Amin, S.A. (2018) Coral metabolite gradients affect microbial community structures and act as a disease cue. *Commun Biol* **1**: 184.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D. *et al.* (2017) vegan: community ecology package. R package version 2.4-5.

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.

Pollock, F.J., McMinds, R., Smith, S., Bourne, D.G., Willis, B.L., Medina, M. *et al.* (2018) Coral-associated bacteria demonstrate phylosymbiosis and cophylogeny. *Nat Commun* **9**: 4921.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., and Glöckner, F.O. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* **35**: 7188-7196.

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

R Core Development Team (2017) R: A language and environment for statistical computing. In. Vienna, Austria: R Foundation for Statistical Computing.

Randall, C.J., Jordan-Garza, A.G., Muller, E.M., and van Woesik, R. (2016) Does dark-spot syndrome experimentally transmit among Caribbean corals? *PLoS One* **11**: e0147493.

Ribet, D., and Cossart, P. (2015) How bacterial pathogens colonize their hosts and invade deeper tissues. *Microbes Infect* **17**: 173-183.

Santoro, A.E., Casciotti, K.L., and Francis, C.A. (2010) Activity, abundance and diversity of nitrifying archaea and bacteria in the central California Current. *Environ Microbiol* **12**: 1989-2006.

Schlichter, D., and Liebezeit, G. (1991) The natural release of amino acids from the symbiotic coral *Heteroxenia fuscescens* (Ehrb.) as a function of photosynthesis. *J Exp Mar Bio Ecol* **150**: 83-90.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B. *et al.* (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75**: 7537-7541.

Seymour, J.R., Amin, S.A., Raina, J.B., and Stocker, R. (2017) Zooming in on the phycosphere: the ecological interface for phytoplankton-bacteria relationships. *Nat Microbiol* **2**.

Shapiro, O.H., Fernandez, V.I., Garren, M., Guasto, J.S., Debaillon-Vesque, F.P., Kramarsky-Winter, E. *et al.* (2014) Vortical ciliary flows actively enhance mass transport in reef corals. *P Natl Acad Sci USA* **111**: 13391-13396.

Sharp, K.H., Ritchie, K.B., Schupp, P.J., Ritson-Williams, R., and Paul, V.J. (2010) Bacterial acquisition in juveniles of several broadcast spawning coral species. *Plos One* **5**.

Shashar, N., Kinane, S., Jokiel, P.L., and Patterson, M.R. (1996) Hydromechanical boundary layers over a coral reef. *J Exp Mar Bio Ecol* **199**: 17-28.

Silveira, C.B., Gregoracci, G.B., Coutinho, F.H., Silva, G.G.Z., Haggerty, J.M., de Oliveira, L.S. *et al.* (2017) Bacterial community associated with the reef coral *Mussismilia braziliensis's* momentum boundary layer over a diel cycle. *Front Microbiol* **8**: 784.

Suzuki, R., and Shimodaira, H. (2006) Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **22**: 1540-1542.

Sweet, M.J., Croquer, A., and Bythell, J.C. (2011) Development of bacterial biofilms on artificial corals in comparison to surface-associated microbes of hard corals. *PLoS One* **6**: e21195.

Tanaka, Y., Miyajima, T., Umezawa, Y., Hayashibara, T., Ogawa, H., and Koike, I. (2009) Net release of dissolved organic matter by the scleractinian coral *Acropora pulchra*. *J Exp Mar Bio Ecol* **377**: 101-106.

Thompson, J.R., Rivera, H.E., Closek, C.J., and Medina, M. (2014) Microbes in the coral holobiont: partners through evolution, development, and ecological interactions. *Front Cell Infect Microbiol* **4**: 176.

Torto-Alalibo, T., Collmer, C.W., and Gwinn-Giglio, M. (2009) The Plant-Associated Microbe Gene Ontology (PAMGO) Consortium: community development of new Gene Ontology terms describing biological processes involved in microbe-host interactions. *BMC Microbiol* **9 Suppl 1**: S1.

Tout, J., Jeffries, T.C., Webster, N.S., Stocker, R., Ralph, P.J., and Seymour, J.R. (2014) Variability in microbial community composition and function between different niches within a coral reef. *Microb Ecol* **67**: 540-552.

Tout, J., Jeffries, T.C., Petrou, K., Tyson, G.W., Webster, N.S., Garren, M. *et al.* (2015) Chemotaxis by natural populations of coral reef bacteria. *ISME J* **9**: 1764-1777.

Urakawa, H., Martens-Habbena, W., and Stahl, D.A. (2010) High abundance of ammonia-oxidizing Archaea in coastal waters, determined using a modified DNA extraction method. *Appl Environ Microbiol* **76**: 2129-2135.

Vega Thurber, R., Willner-Hall, D., Rodriguez-Mueller, B., Desnues, C., Edwards, R.A., Angly, F. *et al.* (2009) Metagenomic analysis of stressed coral holobionts. *Environ Microbiol* **11**: 2148-2163.

Walsh, K., Haggerty, J.M., Doane, M.P., Hansen, J.J., Morris, M.M., Moreira, A.P.B. *et al.* (2017) Aura-biomes are present in the water layer above coral reef benthic macro-organisms. *PeerJ* **5**: e3666.

Wangpraseurt, D., Weber, M., Røy, H., Polerecky, L., de Beer, D., Suharsono, and Nugues, M.M. (2012) In situ oxygen dynamics in coral-algal interactions. *PLoS ONE* **7**.

Ward, C.S., Yung, C.M., Davis, K.M., Blinebry, S.K., Williams, T.C., Johnson, Z.I., and Hunt, D.E. (2017) Annual community patterns are driven by seasonal switching between closely related marine bacteria. *ISME J* **11**: 1412-1422.

Weber, L., DeForce, E., and Apprill, A. (2017) Optimization of DNA extraction for advancing coral microbiota investigations. *Microbiome* **5**.

Wegley, L., Edwards, R., Rodriguez-Brito, B., Liu, H., and Rohwer, F. (2007) Metagenomic analysis of the microbial community associated with the coral *Porites astreoides*. *Environ Microbiol* **9**: 2707-2719.

Zhang, T., Diaz, J.M., Brighi, C., Parsons, R.J., McNally, S., Apprill, A., and Hansel, C.M. (2016) Dark production of extracellular superoxide by the coral *Porites astreoides* and representative symbionts. *Front Mar Sci* **3**: 232.

Zhang, Z., Schwartz, S., Wagner, L., and Miller, W. (2000) A greedy algorithm for aligning DNA sequences. *J Comput Biol* **7**: 203-214.

3.9 Tables

Table 1. Description of reef locations and number and types of samples collected.

Reef system* and site	Date	Coral species for CSW[†]	RSW[‡]	Sand Seawater controls[§]	Seawater experiment samples[∥]	Latitude	Longitude
JR 1	2/8/2015	*M. cavernosa* (3/3) *P. astreoides* (3/3) *P. strigosa* (3/3)	surface (1/0) reef-depth (1/2)			20.77453	-78.91517
JR 2	2/9/2015	*M. cavernosa* (3/2) *P. astreoides* (3/3) *P. strigosa* (3/1)	surface (1/0) reef-depth (1/2)			20.82598	-78.97931
JR 3	2/8/2015		reef-depth (0/2)			20.81478	-78.88320
JR 4	2/10/2015		reef-depth (0/2)			20.87765	-78.97028
JR 5	2/11/2015	*O. faveolata* (3/2) *P. astreoides* (3/3) *P. strigosa* (3/2)	reef-depth (1/2)			21.09232	-78.73354
JR 6	2/12/2015	*O. faveolata* (3/3) *P. astreoides* (3/3) *A. cervicornis* (3/3)	reef-depth (1/2)	sand (1)		21.10845	-78.72080
CAN 12	4/28/2015 (CSW) 4/30/2015 (RSW)	*P. astreoides* (4/3)	reef-depth (1/1)	sand (1)		21.58387	-81.62795
CAN 14	4/30/2015	*O. faveolata* (1/1) *P. strigosa* (2/1)	surface (1/0) reef-depth (1/1)	sand (1)		21.56893	-81.63820
CAN 15	5/06/2015	*P. astreoides* (4/3)	surface (1/0) reef-depth (1)	sand (1)		21.55521	-81.76323
CAN 17	5/04/2015 (CSW) 4/05/2015 (RSW)	*P. astreoides* (4/4)	surface (1/0) reef-depth (1/0)	sand (1)		21.60200	-81.93400
USVI – Tektite	10/29/2016				7	18.3095	-64.7219
USVI – Dock	10/29/2016				5	18.3182	-64.7241

*JR, Jardines de la Reina, Cuba; CAN = Los Canarreos, Cuba; USVI = St. John, USVI

[†]Number of colonies sampled for coral seawater (CSW); CSW samples for microbial cell counts were taken at 0 and 5 cm (JR) or 0, 2, and 30 cm (CAN) away from the colony and this number is the first number in the parentheses; community DNA samples were taken 30 cm away from all JR and CAN coral colonies. The second number in parentheses reflects the number of samples that made it past sequence quality-filtering and that were used in amplicon analysis.

[‡]RSW = number of reef seawater (RSW) samples that were collected. Surface RSW samples were collected 1 m from the surface of the seawater. RSW (reef-depth) samples were collected > 1 m off of the reef. The first number in parentheses indicates the number of samples collected for flow cytometry. The second number in parentheses indicates the number of samples that made it past sequence quality-filtering and/or that were used in amplicon analysis.

[§]Number of seawater samples taken over sand for microbial cell counts. In JR samples were taken at 5 cm and in CAN samples were taken at 0, 2, and 30 cm away from the sand.

[∥]Seawater experiment samples = number of samples for seawater volume experiment.

Table 2. Summary of DNA samples collected from JR by reef location and by the coral species (i.e. CSW) used to create the pooled samples for metagenomic sequencing.

Metagenome sample	Site						Pooled samples
	JR1	JR2	JR3	JR4	JR5	JR6	
Montastraea cavernosa CSW	3	2					5
Porites astreoides CSW		2			1	2	5
Pseudodiploria strigosa CSW	3				1		4
Orbicella faveolata CSW					2	3	5
Acropora cervicornis CSW						3	3
RSW*		1		1	1	1	-

* RSW samples were not pooled.

Table 3. Significantly enriched KEGG pathways in microbial metagenomes from coral seawater (CSW) compared to reef seawater (RSW) in Jardines de la Reina.

Number*	Pathway definition	Orthology count†	Coverage‡	p-value§
02020	Two-component system	136	0.28	1.57E-12
02030	Bacterial chemotaxis	19	0.73	1.41E-10
02040	Flagellar assembly	23	0.58	2.29E-09
02025	Biofilm formation – *P. aeruginosa*	34	0.38	3.11E-07
03070	Bacterial secretion system	30	0.41	3.38E-07
02010	ABC transporters	107	0.22	1.00E-04
04112	Cell cycle - *Caulobacter*	14	0.45	1.19E-04
00540	Lipopolysaccharide biosynthesis	16	0.40	2.26E-04
02026	Biofilm formation - *Escherichia coli*	21	0.34	3.08E-04
00910	Nitrogen metabolism	19	0.32	1.84E-03
00040	Pentose and glucuronate interconversions	20	0.29	5.37E-03
01503	Cationic antimicrobial peptide (CAMP) resistance	16	0.30	8.34E-03
00281	Geraniol degradation	6	0.40	2.24E-02
00511	Other glycan degradation	7	0.37	2.24E-02
05111	Biofilm formation – *V. cholerae*	25	0.23	2.88E-02

*KEGG pathway map number.
†Number of individual KOs from this study that are included within this pathway.
‡Normalized coverage of orthologs within each pathway.
§p-values were calculated using Fisher's exact test.

Table 4. Significantly enriched KEGG modules in microbial metagenomes from CSW compared to RSW in JR.

Module Number*	Module Definition	Orthology Count[†]	Coverage[‡]	p-value[§]
00333	Type IV secretion system	11	0.92	2.20E-03
00331	Type II general secretion pathway	13	0.76	1.73E-02
00529	Denitrification, nitrate => nitrogen	9	0.82	2.51E-02
00324	Dipeptide transport system	5	1	2.62E-02
00506	CheA-CheYBV (chemotaxis) two-component regulatory system	5	1	2.62E-02
00511	PleC-PleD (cell fate control) two-component regulatory system	5	1	2.62E-02

*KEGG module map number.
[†]Number of individual KOs from this study that are included within this module.
[‡]Normalized coverage of orthologs within each module.
[§]p values were calculated using Fisher's exact test.

3.10 Figures
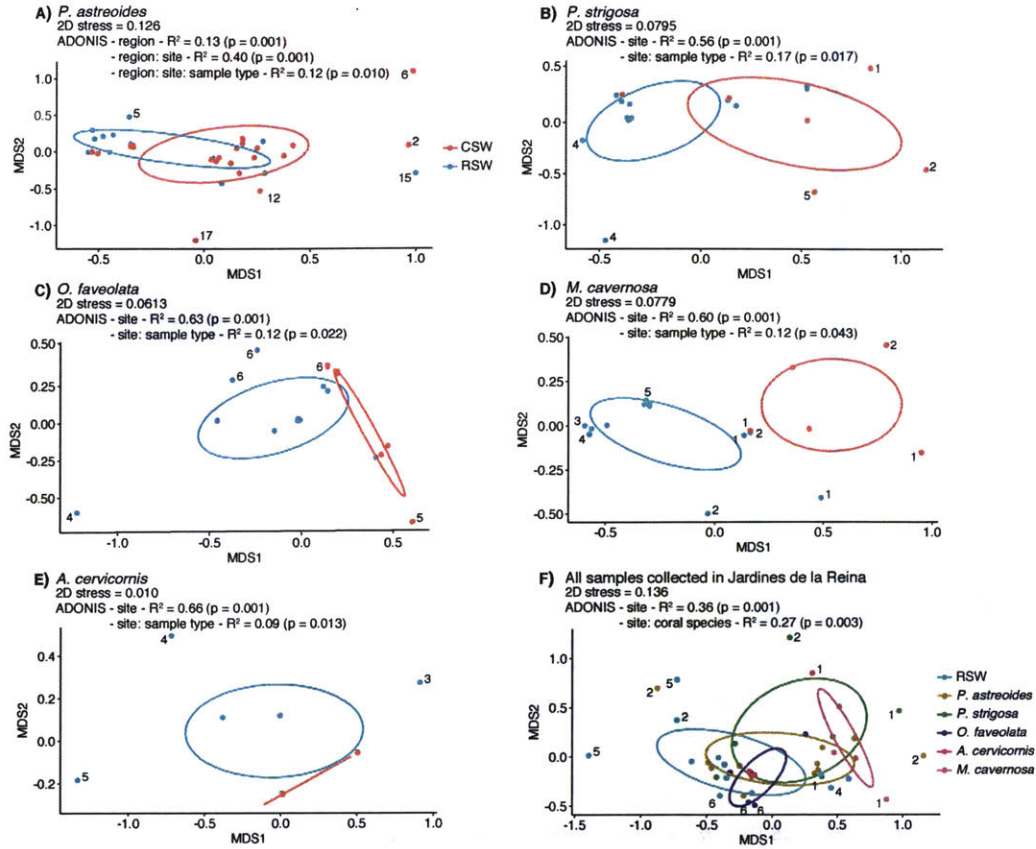
Figure 1. Non-metric multidimensional scaling (NMDS) analyses performed on Bray-Curtis dissimilarity indices that were obtained from square-root transformed relative abundances of 16S rRNA amplicon sequences grouped into MED nodes and results of the nested PERMANOVA (Adonis) tests. Ellipses are drawn using the group mean and covariance for each species. A) *P. astreoides* CSW compared to RSW in JR and CAN, B) *P. strigosa* CSW compared to RSW in JR, C) *O. faveolata* CSW compared to RSW in JR, D) *M. cavernosa* CSW compared to RSW in JR, E) *A. cervicornis* CSW compared to RSW in JR, and F) All JR samples. The number next to each symbol indicates the reef location of collection and only the symbols outside of the covariance ellipses are labeled. The NMDS ordination stress and results of the nested PERMANOVA (Adonis) tests (factors, R^2 value, p - value) are included for each comparison (A-E). The colon between factors indicates the nested structure of each PERMANOVA (Adonis) test; for example, 'site: sample type' indicates that the factor 'sample type' is nested within the factor 'site'. CSW = coral seawater, RSW= reef seawater.
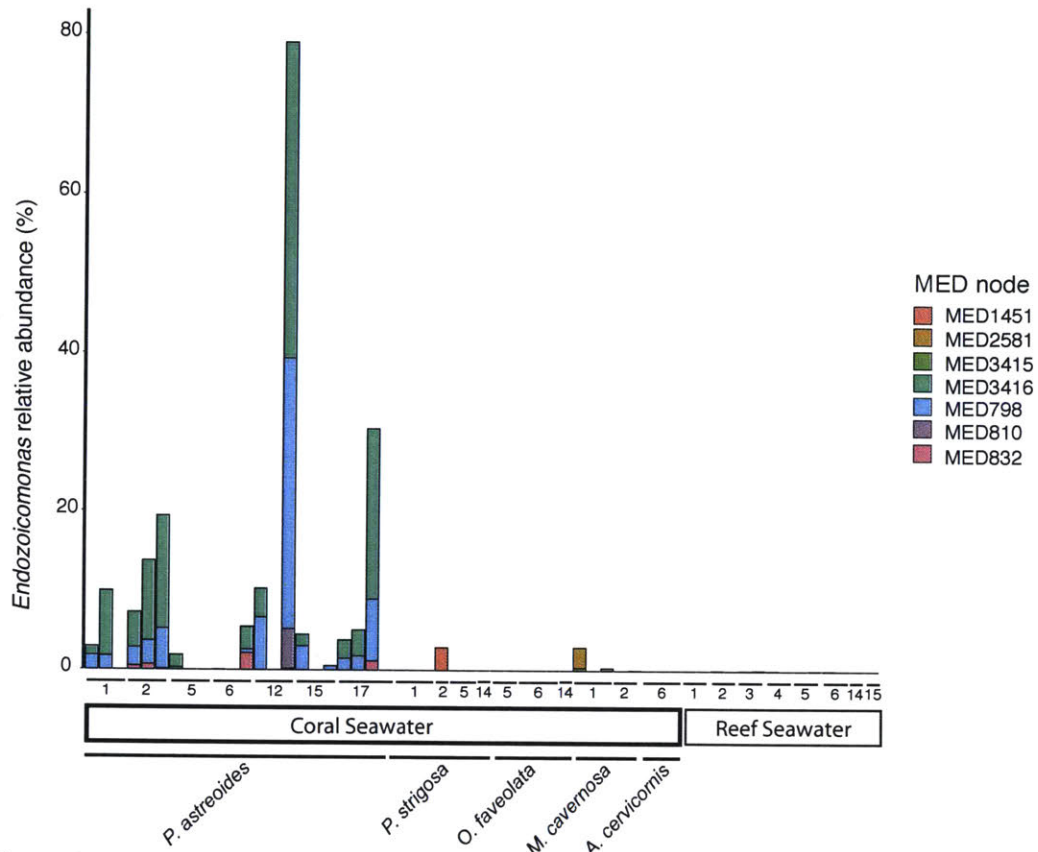
Figure 2. Relative abundances of *Endozoicomonas* minimum entropy decomposition (MED) nodes identified in CSW and RSW samples. The numbers immediately underneath the bars indicate the reef site of collection. The black lines underneath the 'coral seawater' or 'reef seawater' rectangles indicate the coral species from which the surrounding seawater was sampled.

Scaled relative abundance
2
1
0
-1
-2

Reef seawater

Coral seawater

JR6 seawater
JR5 seawater
JR2 seawater
JR4 seawater
A. cervicornis
P. astreoides
O. faveolata
M. cavernosa
P. strigosa
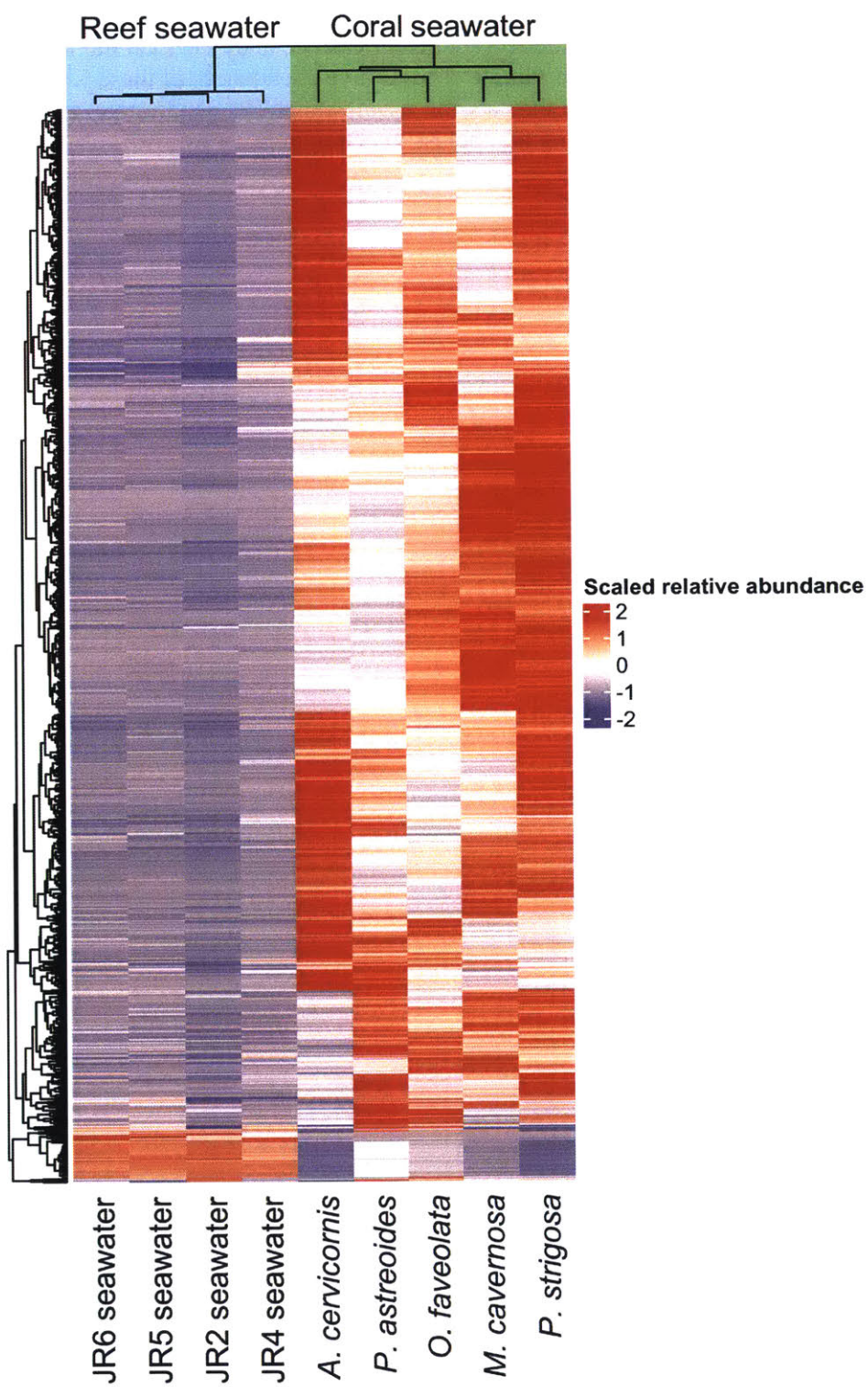
Figure 3. Heatmap displaying the relative abundance of 1058 significantly different KO gene identifiers that were detected between CSW and RSW from Jardines de la Reina. Relative abundances were calculated by dividing the KO counts for each gene by the total number of significantly different KO counts for each sample. The relative abundances of these KOs were scaled using the 10th and 90th quantiles of the data for visualization. The dendrogram reflects hierarchical clustering of the samples using the 'hclust' function in R.
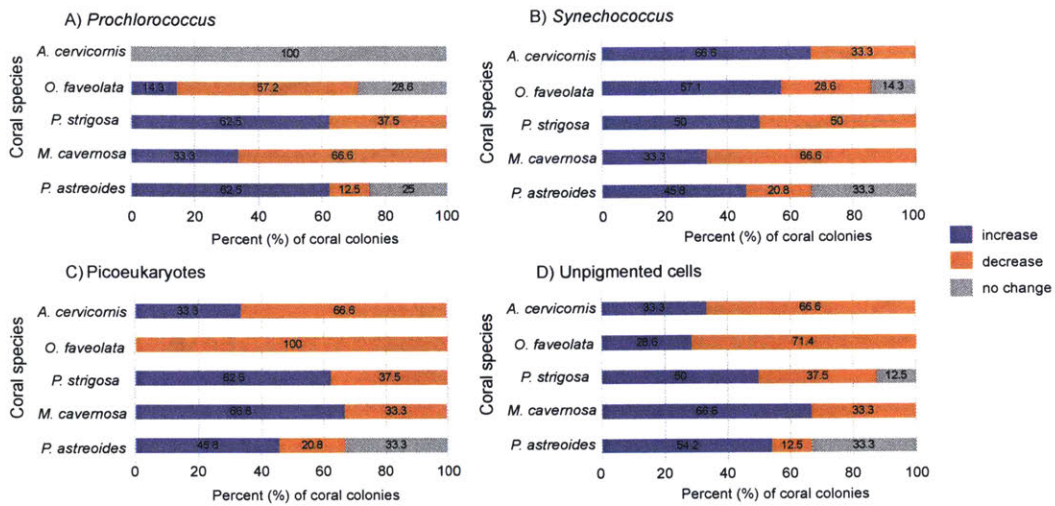
Figure 4. Stacked bar graphs organized by picoplankton functional group in panels A-D depict the percent (%) of coral colonies by coral species where the abundance of cells increased (blue), decreased (orange), or where there was no change as the distance from the colonies increased. The numbers overlaid on the stacked bars reflect the percentage of each category.

125
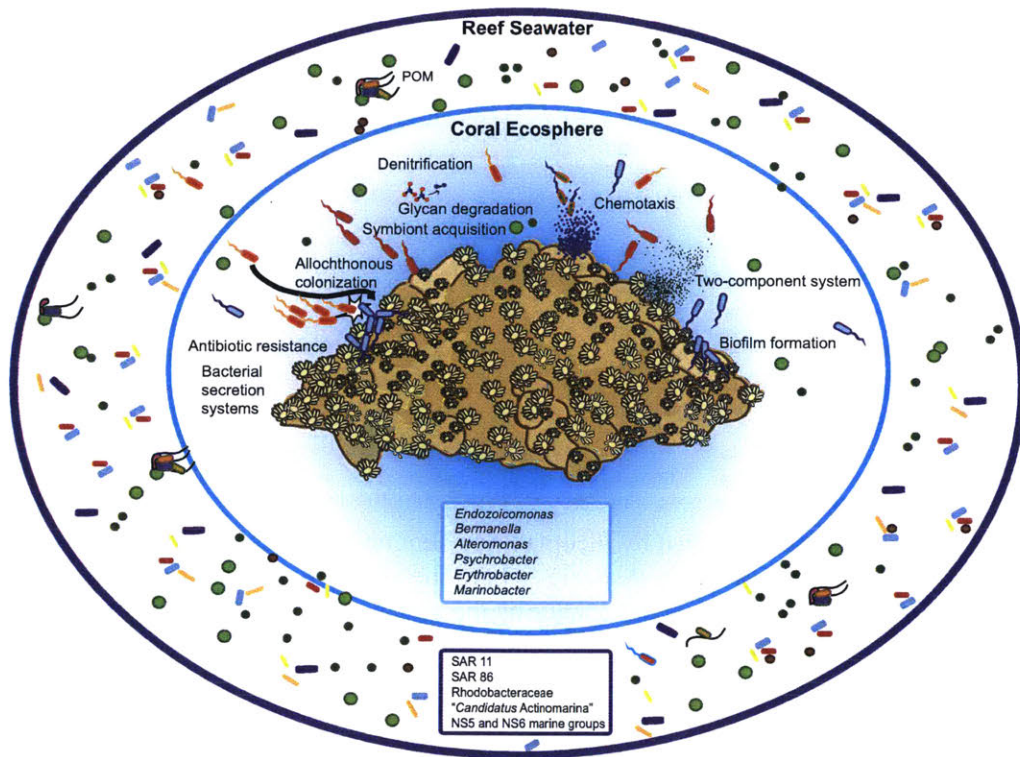
Figure 5. Conceptual diagram of the microbial functions, potential interactions, and enriched bacterial taxa identified within the coral ecosphere. The coral ecosphere and reef-depth seawater hemispheres surround a representative *P. astreoides* coral colony. A light blue line highlights the coral ecosphere boundary and select microbial functions that emerged from this study are illustrated, with microbial cells depicted as colored circles or capsular objects. A darker blue line depicts the reef-depth seawater hemisphere. Enriched bacterial taxa are contained within the boxes within each hemisphere and are based on the within site coral and reef-depth seawater comparisons. The illustrations are not to scale and POM indicates particulate organic matter.

3.11 Supporting Information

Supporting methods

Cell enumeration using flow cytometry
Flow cytometry samples were analyzed at the University of Hawaii using an Altra flow cytometer (Beckman Coulter Life Sciences, Inc, Indianapolis, IN, USA) with a laser excitation wavelength of 488 nm. Unstained and stained (SybrGreen I, Invitrogen™, Waltham, MA, USA) subsamples of each sample were run on the instrument to estimate the concentration of fluorescent picocyanobacteria (*Prochlorococcus* and *Synechococcus*) and picoeukaryotes (unstained) as well as the concentration of unpigmented (stained) cells, respectively. The abundance of unpigmented cells generally serves as a proxy for heterotrophic bacteria within the sample (Marie et al. 1997). Fluorescence spectra were binned, analyzed, and transformed into count data using FlowJo (v. 6.4.7) (FlowJo, LLC) software. The number of cells per ml of seawater was estimated using the original sample volume (1 ml).

To determine overall trends in picoplankton cell abundances over the distance gradient sampled around each colony, we counted the number of instances when cell counts increased, decreased, or had no trend over the distance gradient surrounding each colony. We did this for each species as well as each picoplankton group. Using this metric, the percentage of colonies that had increasing, decreasing, or no trend in cell abundance was determined by species and picoplankton group.

DNA Extractions
Reef seawater (RSW) DNA was extracted from replicate samples taken at each reef. No duplicate samples were collected for the near-coral seawater (CSW) samples. DNA was extracted from the filters using two different DNA extraction protocols (Santoro et al. 2010; Urakawa et al. 2010). The sucrose–EDTA DNA extraction (Santoro et al. 2010) involves lysing the cells that are retained on the filter using a combination of chemical lysis (Sucrose – EDTA lysis buffer), bead-beating, a proteinase-K (25 µl of 20 mg/ml) (PK Solution, Promega, Madison, WI, USA) digestion, and a column-based separation of the DNA from the remaining cellular material. The original filter used during the first DNA extraction method was preserved at -80° C and used again for a second round of extractions using a modified phenol: chloroform: isoamyl alcohol (25:24:1) DNA extraction method (Urakawa et al., 2010). Purified DNA (yielded from the two different extraction methods) was combined for each sample using the Genomic DNA Clean and Concentrator kit (Zymo Research Corporation, Irvine, CA, USA). DNA concentrations were then quantified using the Qubit 2.0 high sensitivity dsDNA assay (ThermoFisher Scientific, Waltham, MA, USA) and inspected for quality using gel electrophoresis (1% Tris-Borate-Ethylenediaminetetracetic acid (TBE) agarose gel) and the HyperLadder™ 1kb marker (Bioline, London, UK).

Fluidigm amplification
DNA extracts were amplified using Fluidigm microfluidic amplification according to Fluidigm protocols. Prior to the first amplification step, 2 ng (1 µl) of each DNA extract was added to 4 µl of a PCR mastermix (Roche High Fidelity Fast Start Kit) in a PCR plate for a total volume of 5 µl. PCR primers were added to a second plate (50 µM each) and diluted to a total volume of 100

μl with the Fluidigm loading reagent and water. Four μl of the sample and 4 μl of the primer were loaded into a primed Fluidigm 48.48 Access Array Integrated Fluidic Circuit (IFC) and the IFC was placed within an AX controller in order to combine the primers with the samples prior to PCR amplification. The Fluidigm Biomark HD PCR machine was used for amplification without imaging. The following amplification steps and cycle numbers were used: 50 °C for 2 minutes (1 cycle); 70 °C for 20 minutes (1 cycle); 95 °C for 10 minutes (1 cycle); 95 °C for 15 seconds, 55 °C for 30 seconds, and 72 °C for 1 minute (10 cycles); 95 °C for 15 seconds, 80 °C for 30 seconds, 60 °C for 30 seconds, and 72 °C for 1 minute (2 cycles); 95 °C for 15 seconds, 55 °C for 30 seconds, and 72 °C for 1 minute (8 cycles), 95 °C for 15 seconds, 80 °C for 30 seconds, 60 °C for 30 seconds, and 72 °C for 1 minute (2 cycles); 95 °C for 15 seconds, 55 °C for 30 seconds, and 72 °C for 1 minute (8 cycles); and 95 °C for 15 seconds, 80 °C for 30 seconds, 60 °C for 30 seconds, and 72 °C for 1 minute (5 cycles).

After the first amplification, 2 μl of Fluidigm Harvest Buffer was loaded into each sample inlet and the AX controller was used to collect the PCR products for each sample. PCR products were then diluted 1:100 in water and 1 μl of the diluted product was amplified using Illumina linkers and barcodes in 20 μl reactions (15 μl of PCR mastermix, 1 μl of diluted PCR product, and 4 μl of Illumina linker barcodes). The PCR reaction conditions included 95 °C for 10 minutes (1 cycle); 95 °C for 15 seconds, 60 °C for 30 seconds, and 72 °C for 1 minute (15 cycles); and an extension step at 72 °C for 3 minutes.

PCR products were quantified and amplicon regions and expected sizes were confirmed using a Fragment Analyzer (Advanced Analytics, Ames, IA). PCR products were pooled into equal ratios, pools were run on a gel for size selection, and bands of the expected size were extracted (Qiagen gel extraction kit). The size and profiles of the pooled and purified PCR products were then checked using a Bioanalyzer (Agilent).

Microbial community sequencing and analysis
Using the program mothur v.1.36.1 (Schloss et al. 2009), forward and reverse reads were united and the locus-specific forward and reverse primers were removed (*make.contigs*). Reads with ambiguous positions or exceeding 275 bp in length were removed (*screen.seqs*), reads matching to unknown, mitochondrial, or eukaryotic sequences were identified (*classify.seqs*, method = 'knn') using the Silva database v119 (Quast et al. 2013) as a reference and removed (*remove.lineages*), and UCHIME (Edgar et al. 2011) was used to identify and remove chimeric reads (based on the command '*chimera.uchime*', reference = self). Reads detected in the DNA extraction and pooling controls were removed from all samples (*remove.seqs*), mock community samples were removed from the dataset prior to read clustering and analyzed separately, and data was subsampled to 8,500 reads per sample. These quality-filtering and processing steps omitted 14 samples and the remaining sequences were used to complete microbial community analysis.

The sequencing error rate (defined as the sum of mismatches to the reference file/ sum of bases within the query) was 0.0027 and was calculated using the '*seq.error*' command for the mock community samples using mothur (Schloss et al. 2009). Subsequent clustering analysis on the two sequenced mock community DNA samples yielded 21 and 17 MED nodes, matching closely with the 21 bacterial strains used to make the synthetic community.

For within site differential enrichment comparisons using DESeq2, geometric means were calculated separately for each MED node because zeros were present in the count data. The command "DESeq" was run with default parameters and a "local" fit smoothed trend-line to estimate gene dispersion. Cook's distance filtering was not applied because normalized count outliers could not be discerned as a result of low sample sizes. After manual inspection of the normalized counts, we determined that this filtering criteria was too conservative for this dataset and this approach has been taken in other studies (Pepe-Ranney and Hall 2015). These within site comparisons were made between all CSW and reef-depth seawater communities at reefs JR 1, 2, 5, and 6 and statistically significant differences (Wald test, Benjamini-Hochberg corrections) between specific contrasts (i.e. *P. astreodies* CSW vs. reef-depth SW) were determined. The R package 'phyloseq' (McMurdie and Holmes 2013) was used to generate bar plots of the relative abundances of the MED nodes.

Due to the nature of this field data, there are layers of environmental variables that contribute to microbial community similarity between samples. We completed nested PERMANOVA (Adonis) tests using the vegan R package (Oksanen et al. 2017) in order to control the permutations over the samples by specific factors. To complete the nested test for CSW collected from *P. strigosa*, *O. faveolata*, *M. cavernosa*, and *A. cervicornis*, the factor of sample type (CSW or RSW) was nested within the factor of reef site (e. g. JR 1). To complete the nested PERMANOVA test for CSW collected from *P. astreoides* as well as RSW, the factor of sample type was nested within the factor of reef site that was then nested within the factor of region (either JR or CAN). Lastly, the nested test for all of the CSW and RSW microbial communities sampled within JR was performed by nesting the factor of coral species (e. g. *O. faveolata*, *P. astreoides*, or RSW) within the factor of reef site. The R^2 value determined by these tests indicates the effect size and shows the percentage of variation explained by each specific factor.

Preparation of coral seawater and benthic seawater DNA for shotgun metagenomic sequencing
Seawater DNA was extracted from the four reef seawater metagenome filters using a modified cetyl-trimethylammonium bromide (CTAB)-phenol: chloroform: isoamyl alcohol extraction that was developed from two existing DNA extraction methods (Zhou et al. 1996; William et al. 2004). One half of each filter was exposed to physical, enzymatic, and chemical disruption via 3 freeze-thaw cycles, bead-beating, and proteinase-k (20 mg/ml) (PK Solution, Promega, Madison, WI, USA) and lysozyme (20 mg/ml) (Pierce™ Lysozyme, Thermo Scientific) digestions. CTAB, an effective surfactant used for purifying DNA in the presence of polysaccharides (Clarke 2009), was added to the sample, followed by a phenol: chloroform (24:1), phenol: chloroform: isoamyl alcohol (25:24:1), phenol: chloroform (24:1) rinsing series. DNA was precipitated using molecular grade isoproponal overnight at -20 °C and the DNA pellet was rinsed with 70% ethanol twice before it was eluted into 50 µl of TE buffer (10 mM Tris-Cl, pH 7.5; 1 mM EDTA). DNA concentrations for these samples were quantified and screened for quality using by using gel electrophoresis prior to sequencing.

After sequencing, the fastq files were demultiplexed, and library adaptors were trimmed from the 3' ends of the reads. Overall, 92,699,608 paired-end reads were generated with an average read number of 18,539,922 (+/- 9,882,964) paired-end reads per sample. The total number of paired-end reads sequenced in reef seawater samples were as follows: 28,934,702 reads for site JR 2; 35, 933, 406 for site JR 4; 27, 444, 129 reads for site JR 5; and 32, 748, 216

129

reads for site JR 6. The total number of paired-end reads obtained from CSW were as follows: 30, 095, 047 reads for *P. astreodies*; 7, 914, 591 reads for *P. strigosa*; 15, 345, 580 reads for *O. faveolata*, 11, 563, 389 reads for *M. cavernosa*, and 27, 781, 001 reads *for A. cervicornis*. DNA fragment size for the pooled *A. cervicornis* seawater samples ranged from 280-700 bp and fragment sizes for the other CSW samples ranged between 80-600 bp in length.

BBTools (Bushnell 2016) was used to quality-filter and prepare the raw metagenomic reads for functional analysis. Remnant sequencing adaptors were removed from the raw forward and reverse reads for each sample using *bbduk.sh*, the BBtools reference adaptors file (adapter.fa), and the following parameters: ktrim=r k=23 mink=11 hdist=1 tpe tbo (Bushnell 2016). Following removal of the adaptor sequences, the *bbduk.sh* script was implemented again to quality-trim the forward and reverse reads using the Phred algorithm and a Q score of 10 (qtrim=rl trimq=10).

After quality-filtering, the Functional Mapping and Analysis Pipeline for metagenomics and metatranscriptomics studies (FMAP) (Kim et al. 2016) was used to determine functional differences between the CSW and RSW metagenomes. The example script (available here: https://qbrc.swmed.edu/FMAP/) was modified to suite this specific comparison. Reef seawater had a higher assignment of quality-filtered reads to KEGG Orthologies (KOs) compared to pooled coral seawater samples, ranging from 6.58 – 11.17 % of unmerged reads with an average read assignment of 8.61 %. Individually, 11.17% of all reads could be annotated for JR 2, 9.39% could be annotated for JR 4, 7.29% could be annotated for JR 5, and 6.58% could be annotated for JR 6.The pooled coral seawater samples had lower overall assignment, ranging from 1.59 – 12.83%, with an average read assignment of 4.48%. Reads from pooled *A. cervicornis* CSW had the highest level of annotation at 12.83%, followed by *P. strigosa* CSW (3.76%), *M. cavernosa* CSW (2.59%), *O. faveolata* CSW (1.66%), and *P. astreoides* CSW (1. 60%). The final files generated by FMAP provide the user with count data for the number of gene hits that are assigned to a specific Kyoto Encyclopedia of Genes and Genomes (KEGG) Ortholog identifier, as well as files revealing which KOs, KEGG pathways, KEGG modules, and KEGG operons are significantly different between the two sample groups (Kruskal – Wallis test, p-value < 0.05, FDR adjusted to control for false positives) .

To analyze the output from this pipeline, the KEGG KO abundance table file and significantly different (p adjusted < 0.05) KOs between the CSW and RSW file were merged using core R functions in R studio. The KO count data was converted into relative abundance using the sum of all genes that could be annotated in order to normalize changes and visually compare differences across samples. We also scaled the relative abundances using the 10th and 90th quantiles of the data to enhance visual comparison between the samples. Changes in gene abundance across the samples were visualized using the R package 'ComplexHeatmap' (Gu et al. 2016) (available from https://github.com/jokergoo/ComplexHeatmap) with scripts that were tailored for our data (script scaffolds were obtained from http://zuguang.de/supplementary/ComplexHeatmap-supplementary1-4/supplS2_scRNASeq/supplS2_scRNAseq.html). The dendrogram reflects hierarchical clustering of the samples using the 'hclust' function in R.
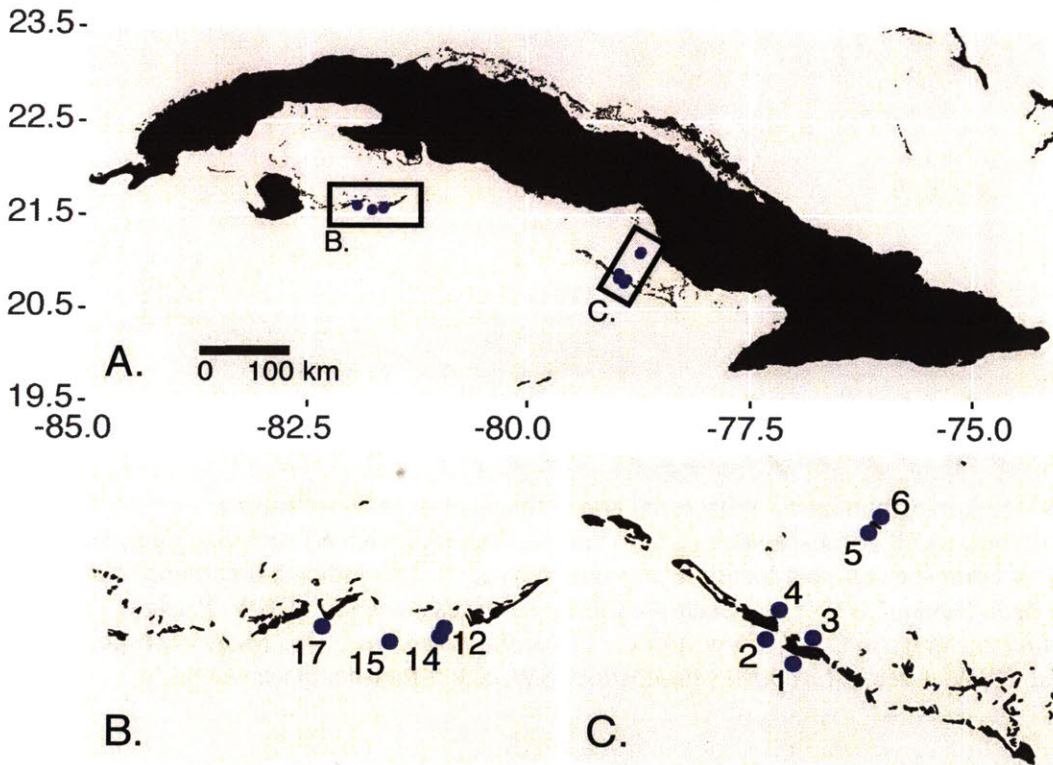
Figure S1. Overview map of the reef locations that were surveyed in this study. A) The location of the reefs relative to the island of Cuba. Reef-systems of Los Canarreos (B.) and Jardines de la Reina (C.) are contained within the black boxes B) Close-up of reef locations in the Canarreos reef-system. C) Close-up of reef locations in the Jardines de la Reina reef-system. The blue circles indicate the location of each reef.
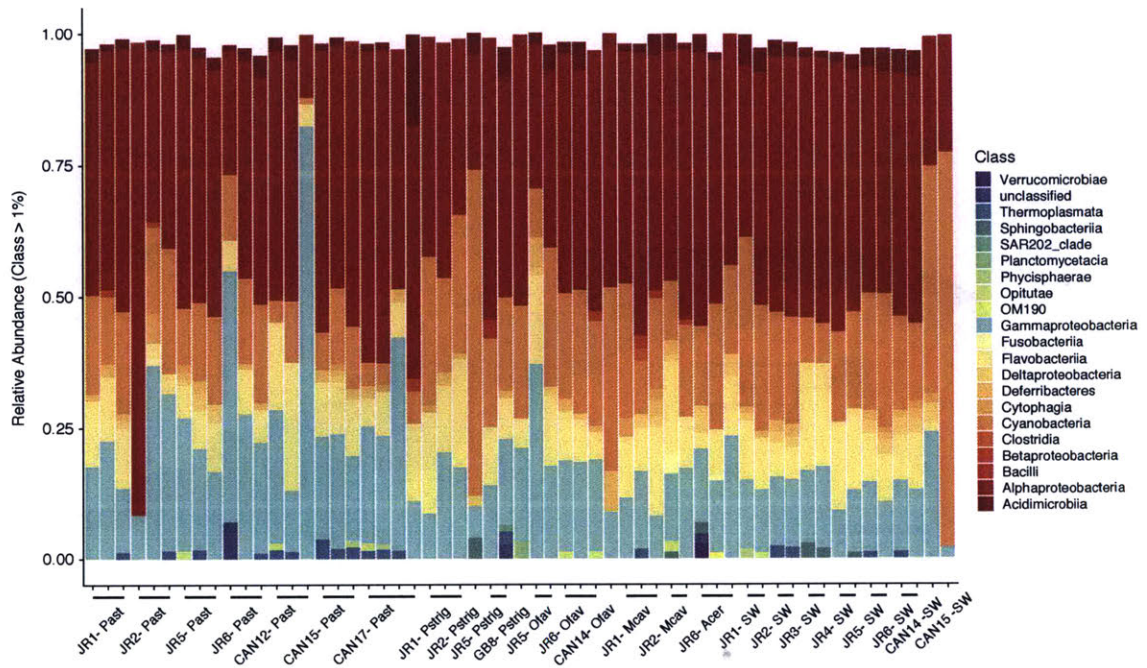
Figure S2. Relative abundance of bacterial and archaeal classes that comprise >1 % of the community across all coral seawater (CSW) and reef seawater (RSW) samples. Samples are grouped by coral species, reef location, and sample type. Colors indicate taxonomic class. JR = Jardines de la Reina, CAN = Los Canarreos. Past = *Porites astreoides* CSW, Pstrig = *Pseudodiploria strigosa* CSW, CSW, Ofav = *Orbicella faveolata* CSW, Mcav = *Montastraea cavernosa* CSW, Acer = *Acropora cervicornis* CSW, SW = reef-depth seawater.
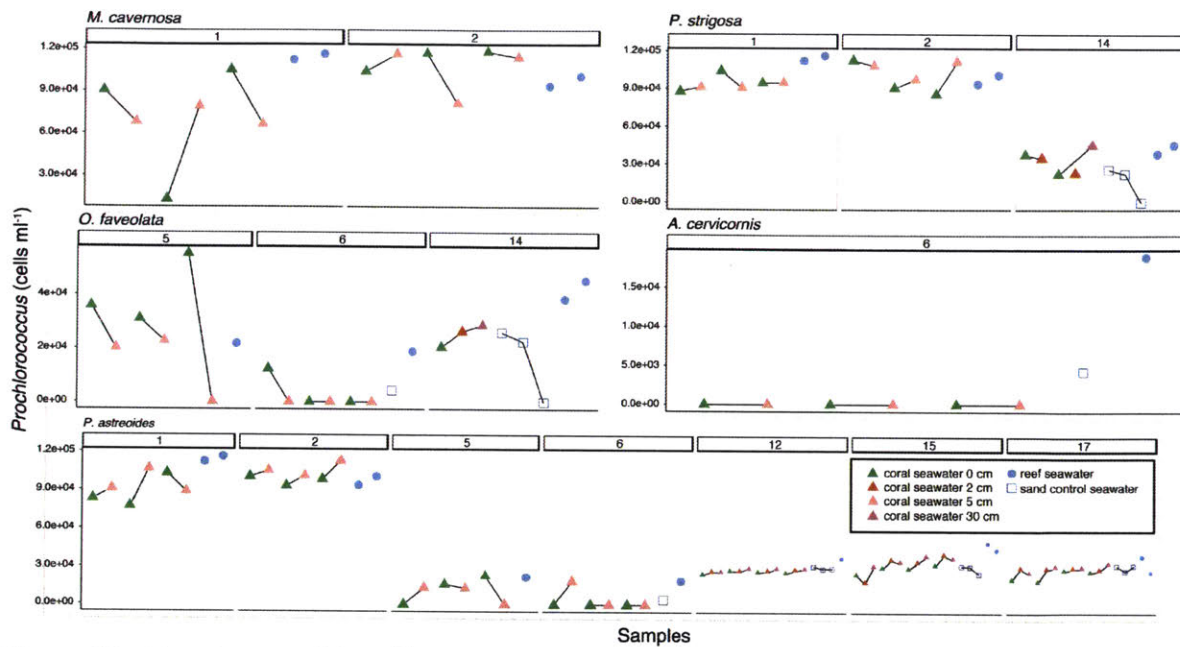
Figure S3. Abundances of *Prochlorococcus* within coral seawater (CSW), reef seawater (RSW), and sand control samples by coral species. The line connects samples that were obtained over a distance gradient from that colony and indicates the direction of the trend. RSW and sand control samples are not colored differently between surface and reef-depth or over the distance gradient sampled from the sand control samples (i.e. sites 12, 14, 15, 17; distances of 0, 2, and 30 cm away from the sand).
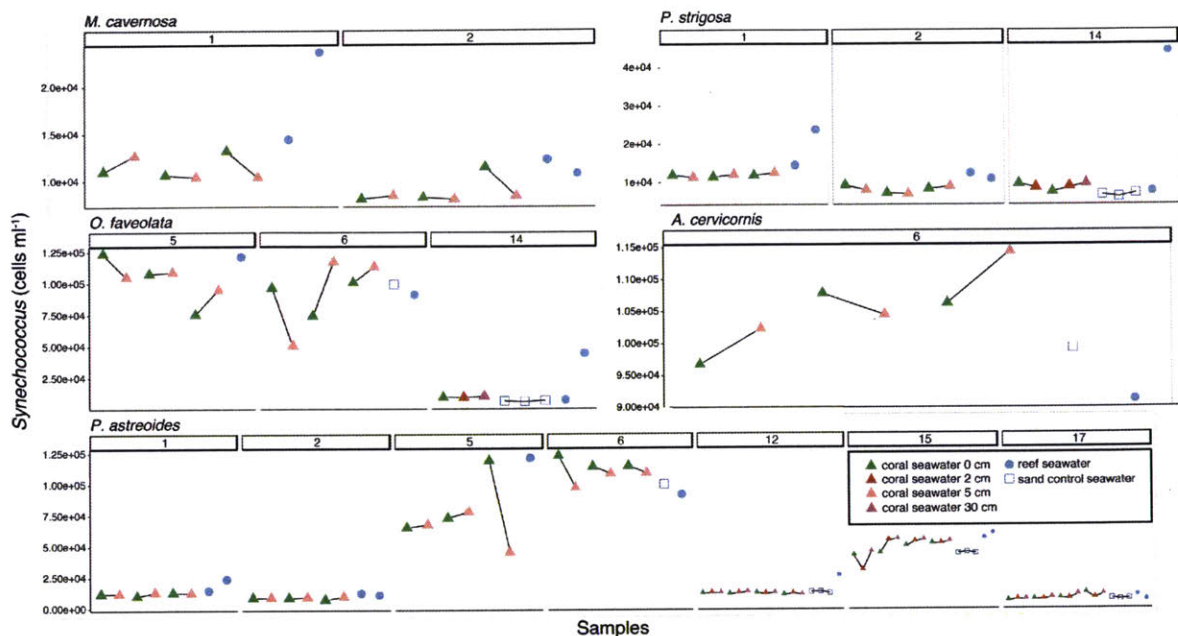
Figure S4. Abundances of *Synechococcus* within coral seawater (CSW), reef seawater (RSW), and sand control samples by coral species. The line connects samples that were obtained over a distance gradient from that colony and indicates the direction of the trend. RSW and sand control samples are not colored differently between surface and reef-depth or over the distance gradient sampled from the sand control samples (i.e. sites 12, 14, 15, 17; distances of 0, 2, and 30 cm away from the sand).

134
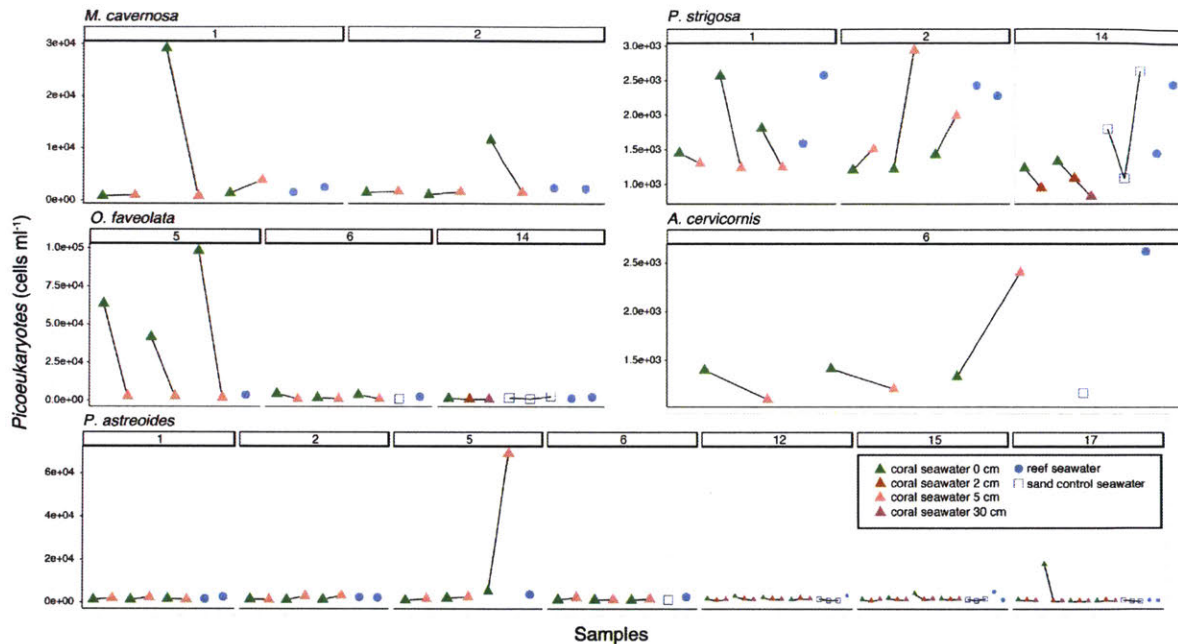
Figure S5. Abundances of picoeukaryotes within coral seawater (CSW), reef seawater (RSW), and sand control samples by coral species. The line connects samples that were obtained over a distance gradient from that colony and indicates the direction of the trend. RSW and sand control samples are not colored differently between surface and reef-depth or over the distance gradient sampled from the sand control samples (i.e. sites 12, 14, 15, 17; distances of 0, 2, and 30 cm away from the sand).
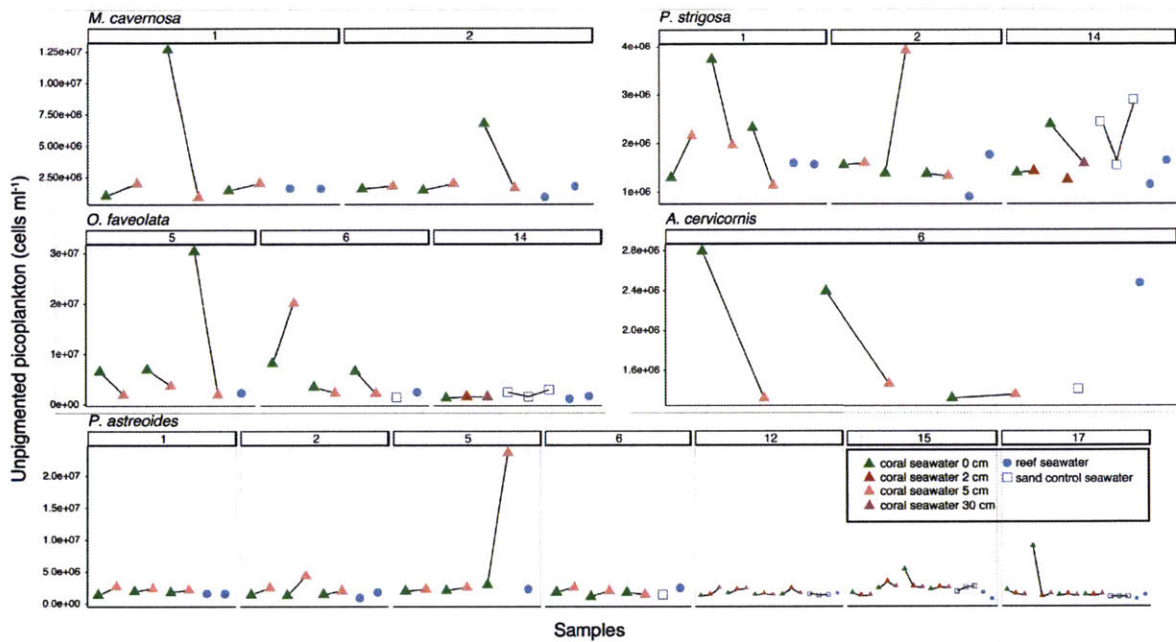
Figure S6. Abundances of unpigmented picoplankton within coral seawater (CSW), reef seawater (RSW), and sand control samples by coral species. The line connects samples that were obtained over a distance gradient from that colony and indicates the direction of the trend. RSW and sand control samples are not colored differently between surface and reef-depth or over the distance gradient sampled from the sand control samples (i.e. sites 12, 14, 15, 17; distances of 0, 2, and 30 cm away from the sand).

Table S1. Significantly enriched and depleted MED nodes detected in Jardines de la Reina (JR) coral seawater (CSW) compared to reef seawater (RSW) by reef site according to paired differential enrichment analysis using DESeq2§.

Site	Coral	MED node	Log2 fold change	Padj*	Taxa
			Enriched in Coral Seawater (CSW)		
JR1	*P. astreoides*	MED2613	21.88	4.28E-04	Gammaproteobacteria, Cellvibrionaceae
		MED3465	16.29	1.31E-03	Gammaproteobacteria, *Marinobacter*
		MED5309	15.55	1.31E-03	Gammaproteobacteria, *Alteromonas*
JR2	*P. astreoides*	MED56	27.54	4.83E-08	Alphaproteobacteria, SAR11 clade, Surface 1
		MED2424	23.61	5.96E-06	Gammaproteobacteria, *Vibrio*
		MED5299	21.23	4.83E-08	Gammaproteobacteria, *Vibrio*
		MED2524	21.18	4.83E-08	Fusobacteria, *Propionigenium*
		MED798	11.4	4.39E-10	Gammaproteobacteria, *Endozoicomonas*
		MED3416	11.32	2.98E-10	Gammaproteobacteria, *Endozoicomonas*
JR5	*P. astreoides*	MED3416	20.8	2.79E-06	Gammaproteobacteria, *Endozoicomonas*
		MED3450	17.7	3.02E-05	Firmicutes, Bacillales, unclassified
		MED5924	16.46	3.02E-05	Gammaproteobacteria, *Psychrobacter*
		MED2954	9.89	2.40E-03	Gammaproteobacteria, *Bermanella*
JR6	*P. astreoides*	MED3416	21.7	2.10E-04	Gammaproteobacteria, *Endozoicomonas*
		MED832	21.21	2.23E-04	Gammaproteobacteria, *Endozoicomonas*
		MED5123	18.93	1.03E-05	Gammaproteobacteria, *Psychrobacter*
		MED615	18.81	2.10E-04	Flavobacteria, *Mesoflavibacter*
		MED29181	18	3.10E-04	Gammaproteobacteria, *Pseudoalteromonas*
		MED20	10.67	2.88E-02	Gracilibacteria, unclassified
		MED2840	10.04	2.52E-02	Gammaproteobacteria, *Idiomarina*
		MED53094	7.03	1.95E-02	Gammaproteobacteria, *Alteromonas*
JR1	*P. strigosa*	MED1993	21.59	7.02E-05	Alphaproteobacteria, SAR11 clade
		MED4248	21.07	9.73E-05	Alphaproteobacteria, SAR11 clade, Surface 2
		MED5392	21.02	9.73E-05	Flavobacteria, unclassified
		MED4380	20.97	9.73E-05	Cyanobacteria, Subsection III, *Leptolyngbya*
		MED2918	20.54	1.39E-04	Gammaproteobacteria, *Pseudoalteromonas*
		MED2355	20.42	1.46E-04	Bacteroidetes, Cytophagia, Order III
		MED2577	20.33	2.12E-05	Gammaproteobacteria, *Marinobacter*
		MED5468	20.3	7.02E-05	Alphaproteobacteria, SAR 116 clade
		MED53091	19.72	4.53E-06	Gammaproteobacteria, *Alteromonas*
		MED34651	19.72	7.81E-06	Gammaproteobacteria, *Marinobacter*
		MED4860	7	4.58E-02	Alphaproteobacteria, SAR11 clade, Surface 1
JR2	*P. strigosa*	MED1408	12.01	1.09E-05	Bacteroidetes, Chitinophagaceae
		MED1451	11.53	5.39E-04	Gammaproteobacteria, *Endozoicomonas*

		MED3451	8.55	1.56E-02	Deltaproteobacteria, OM27 clade
		MED6080	2.86	7.11E-03	Alphaproteobacteria, SAR11, Surface 1 clade
		MED36901	2.63	3.95E-02	Actinobacteria, "*Candidatus* Actinomarina"
JR5	*P. strigosa*	MED2229	23.79	3.13E-10	Alphaproteobacteria, Rhodobacteraceae
		MED1731	23.52	1.50E-05	Bacteroidetes, *Owenweeksia*
		MED3330	22.73	2.16E-08	Alphaproteobacteria, "*Candidatus* Laris"
		MED34501	22.17	5.25E-08	Firmicutes, Bacillales, unclassified
		MED29541	7.97	1.86E-02	Gammaproteobacteria, *Bermanella*
JR1	*M. cavernosa*	MED2581	21.23	1.31E-04	Gammaproteobacteria, *Endozoicomonas*
		MED2356	20.1	3.41E-04	Bacteroidetes, Cytophagia
		MED42	20.05	1.31E-04	Betaproteobacteria, *Variovorax*
		MED1037	19.8	3.99E-04	Firmicutes, Clostridia, Halanaerobiales
		MED53092	19.49	3.02E-06	Gammaproteobacteria, *Alteromonas*
		MED34652	18.11	5.13E-05	Gammaproteobacteria, *Marinobacter*
		MED25771	15	5.30E-03	Gammaproteobacteria, *Marinobacter*
JR5	*O. faveolata*	MED59241	22.98	2.10E-09	Gammaproteobacteria, *Psychrobacter*
		MED2233	11.27	2.44E-03	Alphaproteobacteria, *Erythrobacter*
		MED1437	11.06	7.24E-03	Alphaproteobacteria, *Thalassospira*
		MED5923	10.02	7.24E-03	Gammaproteobacteria, *Psychrobacter*
		MED53093	8.44	1.14E-04	Gammaproteobacteria, *Alteromonas*
JR6	*A. cervicornis*	MED2482	23.32	1.49E-05	Gammaproteobacteria, Piscirickettsiaceae
		MED51231	21.73	1.35E-07	Gammaproteobacteria, *Psychrobacter*
		MED29182	21.53	1.43E-05	Gammaproteobacteria, *Pseudoalteromonas*
		MED1955	21.41	1.69E-04	Deltaproteobacteria, SAR324 Clade (marine group B)
		MED1266	21.37	1.69E-04	Parcubacteria (OD1), unclassified
		MED5184	21.22	4.53E-05	Gammaproteobacteria, OM60 (Nor5) clade
		MED4684	20.85	2.59E-04	Gammaproteobacteria, SAR86 clade
		MED710	20.18	4.57E-04	Verrucomicrobia, *Rubritalea*
		MED6177	1.81	4.87E-02	Alphaproteobacteria, SAR11 clade, Surface 1
		MED6119	1.23	1.69E-04	Cyanobacteria, *Synechococcus*
Depleted in Coral Seawater (CSW)					
JR1	*P. astreoides*	MED4826	-19.52	2.38E-04	Alphaproteobacteria, SAR11 clade, Surface 1
JR5	*P. astreoides*	MED2284	-17.34	3.14E-06	Alphaproteobacteria, Rhodobacteraceae
JR6	*P. astreoides*	MED1929	-15.64	1.37E-02	Bacteroidetes, Chitinophagaceae
JR1	*P. strigosa*	MED48261	-18.84	4.63E-05	Alphaproteobacteria, SAR11 clade, Surface 1
		MED4567	-18.24	1.43E-05	Bacteroidetes, NS9 marine group
		MED2968	-17.95	8.27E-06	Alphaproteobacteria, AEGEAN-169 marine group
		MED6081	-17.1	5.21E-04	Alphaproteobacteria, SAR11 clade, Surface 1
		MED1081	-8.87	3.48E-03	Actinobacteria, "*Candidatus* Actinomarina"

Site	Species	MED node	log₂FC	padjust*	Taxonomy
		MED3565	-8.77	2.93E-02	Alphaproteobacteria, SAR11 clade, Surface 4
		MED521	-8.62	5.25E-04	Alphaproteobacteria, Rhodobacteraceae
		MED5856	-6.67	4.58E-02	Alphaproteobacteria, SAR11 clade, Surface 1
		MED3690	-3.18	4.58E-02	Actinobacteria, "*Candidatus* Actinomarina"
JR2	*P. strigosa*	MED3636	-8.5	7.11E-03	Alphaproteobacteria, AEGEAN-169 marine group
		MED3857	-8.18	1.48E-02	Cyanobacteria, Subsection1, Family1
		MED5355	-7.76	8.24E-03	Bacteroidetes, NS5 marine group
		MED5238	-7.62	3.59E-02	Gammaproteobacteria, SAR86 clade
		MED3368	-7.42	3.95E-02	Gammaproteobacteria, SAR86 clade
		MED3501	-7.37	2.59E-02	Alphaproteobacteria, SAR116 clade
		MED4287	-6.98	3.95E-02	Alphaproteobacteria, SAR11 clade, Surface 2
		MED6124	-3.35	5.41E-03	Alphaproteobacteria, SAR11 clade
JR5	*P. strigosa*	MED2284	-15.4	8.12E-05	Alphaproteobacteria, Rhodobacteraceae
JR1	*M. cavernosa*	MED3620	-21.2	1.20E-07	Deferribacteres, SAR406 clade (Marine group A)
		MED1441	-9.93	1.31E-04	Gammaproteobacteria, SAR86 clade
		MED35651	-8.94	3.42E-02	Alphaproteobacteria, SAR11 clade, Surface 4
		MED5761	-5.8	3.42E-02	Alphaproteobacteria, SAR11 clade, Surface 1
JR6	*A. cervicornis*	MED3341	-8.65	1.36E-04	Gammaproteobacteria, SAR86 clade
		MED52381	-7.62	2.92E-02	Gammaproteobacteria, SAR86 clade

*padjust = adjusted p-value calculated using Benjamini-Hochberg corrections.

§Only site and species combinations with significantly enriched or depleted MED nodes are indicated and taxa are arranged in order from the highest to lowest $\log_2$ fold change within each species and site grouping.

Table S2. Relative abundance (%) of two-component system KEGG Orthologs (KO) across reef seawater (RSW) and coral seawater (CSW) metagenomes.

KO definition	JR 2 RSW	JR 4 RSW	JR 5 RSW	JR 6 RSW	Acer* CSW	Pstrig& CSW	Mcav+ CSW	Ofav# CSW	Past$ CSW
K02477; two-component system, LytTR family, response regulator	0.0070	0.0157	0.0080	0.0072	0.0658	0.0526	0.0366	0.0365	0.0263
K02478; two-component system, LytTR family, sensor kinase [EC:2.7.13.3]	0.0008	0.0029	0.0027	0.0018	0.0321	0.0697	0.0151	0.0133	0.0303
K02481; two-component system, NtrC family, response regulator	0.0108	0.0468	0.0407	0.0331	0.1015	0.0569	0.0452	0.1163	0.0850
K02483; two-component system, OmpR family, response regulator	0.0513	0.1251	0.1487	0.1280	0.3640	0.3529	0.3919	0.3489	0.2286
K02484; two-component system, OmpR family, sensor kinase [EC:2.7.13.3]	0.0025	0.0080	0.0160	0.0168	0.0440	0.1466	0.1055	0.0565	0.0283
pleD; two-component system, cell cycle response regulator [EC:2.7.7.65]	0.0017	0.0132	0.0027	0.0042	0.0131	0.0470	0.1012	0.0399	0.0121
pilH; twitching motility two-component system response regulator	0.0037	0.0055	0.0073	0.0024	0.0131	0.0313	0.0129	0.0166	0.0142
pilS, pehS; two-component system, NtrC family, sensor histidine kinase PilS [EC:2.7.13.3]	0.0004	0.0070	0.0060	0.0030	0.0354	0.0612	0.0194	0.0266	0.0364
cheA; two-component system, chemotaxis family, sensor kinase [EC:2.7.13.3]	0.0178	0.1043	0.0667	0.0625	0.4620	0.3685	0.3273	0.3057	0.1861
cheB; two-component system, chemotaxis family, response regulator [EC:3.1.1.61]	0.0033	0.0333	0.0173	0.0258	0.2554	0.1651	0.2541	0.1130	0.0829
cheY; two-component system, chemotaxis family, response regulator	0.0050	0.0516	0.0167	0.0319	0.1447	0.1181	0.1077	0.0797	0.0748

cheV; two-component system, chemotaxis family, response regulator	0.0037	0.0139	0.0040	0.0042	0.1545	0.0768	0.0431	0.0199	0.0829
phoQ; two-component system, OmpR family, sensor histidine kinase [EC:2.7.13.3]	0.0004	0.0018	0.0047	0.0018	0.0457	0.0470	0.0086	0.0133	0.0344
rstB; two-component system, OmpR family, sensor histidine kinase [EC:2.7.13.3]	0.0004	0.0037	0.0033	0.0012	0.0944	0.0825	0.0323	0.0598	0.0263
cpxA; two-component system, OmpR family, sensor histidine kinase [EC:2.7.13.3]	0.0004	0.0033	0.0047	0.0030	0.0879	0.0370	0.0108	0.0233	0.0263
creC; two-component system, OmpR family, sensor histidine kinase [EC:2.7.13.3]	0.0004	0.0022	0.0020	0.0006	0.0121	0.0455	0.0237	0.0233	0.0061
baeS, smeS; two-component system, OmpR family, sensor histidine kinase BaeS [EC:2.7.13.3]	0.0012	0.0026	0.0060	0.0066	0.0777	0.0455	0.0237	0.0199	0.0222
cusS, copS, silS; two-component system, OmpR family, heavy metal sensor histidine kinase CusS [EC:2.7.13.3]	0.0008	0.0018	0.0040	0.0024	0.0676	0.1138	0.0667	0.0432	0.0324
qseC; two-component system, OmpR family, sensor histidine kinase [EC:2.7.13.3]	0.0017	0.0015	0.0053	0.0054	0.0795	0.1039	0.0711	0.0365	0.0384
kdpD; two-component system, OmpR family, sensor histidine kinase [EC:2.7.13.3]	0.0004	0.0015	0.0040	0.0024	0.0529	0.3088	0.2670	0.1229	0.0384
torS; two-component system, OmpR family, sensor histidine kinase [EC:2.7.13.3]	0.0000	0.0026	0.0027	0.0048	0.0130	0.0100	0.0108	0.0066	0.0061
arcB; two-component system, OmpR family, aerobic respiration control	0.0017	0.0051	0.0060	0.0066	0.0781	0.0228	0.0108	0.0266	0.0303

sensor histidine kinase [EC:2.7.13.3]									
mtrB; two-component system, OmpR family, sensor histidine kinase [EC:2.7.13.3]	0.0012	0.0066	0.0113	0.0108	0.0119	0.0640	0.0301	0.0266	0.0364
phoP; two-component system, OmpR family, response regulator	0.0008	0.0190	0.0093	0.0072	0.0365	0.0256	0.0194	0.0332	0.0243
rstA; two-component system, OmpR family, response regulator	0.0025	0.0048	0.0027	0.0054	0.0735	0.0484	0.0129	0.0332	0.0222
cpxR; two-component system, OmpR family, response regulator	0.0004	0.0099	0.0073	0.0066	0.1146	0.0327	0.0129	0.0332	0.0425
cusR, copR, silR; two-component system, OmpR family, copper resistance phosphate regulon response regulator	0.0021	0.0048	0.0067	0.0024	0.0542	0.0953	0.0560	0.0332	0.0162
qseB; two-component system, OmpR family, response regulator	0.0029	0.0022	0.0020	0.0042	0.0564	0.0526	0.0172	0.0631	0.0162
kdpE; two-component system, OmpR family, KDP operon response regulator	0.0012	0.0026	0.0020	0.0024	0.0214	0.0953	0.1120	0.0399	0.0202
uhpB; two-component system, NarL family, sensor histidine kinase [EC:2.7.13.3]	0.0000	0.0004	0.0000	0.0000	0.0204	0.0398	0.0215	0.0233	0.0324
barA, gacS, varS; two-component system, NarL family, sensor histidine kinase BarA [EC:2.7.13.3]	0.0004	0.0106	0.0067	0.0084	0.2554	0.1537	0.0689	0.1362	0.0850
evgS, bvgS; two-component system, NarL family, sensor histidine kinase EvgS [EC:2.7.13.3]	0.0008	0.0026	0.0040	0.0024	0.0293	0.1138	0.0474	0.0465	0.0101
narL; two-component system, NarL family, nitrate/nitrite response regulator	0.0054	0.0088	0.0060	0.0024	0.0538	0.0341	0.0215	0.0100	0.0222

desR; two-component system, NarL family, response regulator	0.0021	0.0015	0.0040	0.0012	0.0422	0.0441	0.0086	0.0332	0.0162
glrR, qseF; two-component system, NtrC family, response regulator GlrR	0.0050	0.0124	0.0127	0.0156	0.0830	0.0427	0.0258	0.0299	0.0425
senX3; two-component system, OmpR family, sensor histidine kinase [EC:2.7.13.3]	0.0004	0.0059	0.0073	0.0084	0.0155	0.0470	0.0366	0.0133	0.0202
arcA; two-component system, OmpR family, aerobic respiration control protein	0.0012	0.0022	0.0033	0.0048	0.0609	0.0242	0.0108	0.0133	0.0182
desK; two-component system, NarL family, sensor histidine kinase [EC:2.7.13.3]	0.0000	0.0007	0.0033	0.0006	0.0214	0.0228	0.0129	0.0365	0.0040
K07814; putative two-component system response regulator	0.0012	0.0172	0.0167	0.0126	0.0618	0.0470	0.0538	0.0465	0.0182
algZ; two-component system, LytTR family, sensor histidine kinase [EC:2.7.13.3]	0.0004	0.0055	0.0080	0.0036	0.0347	0.0413	0.0172	0.0266	0.0101
algR; two-component system, LytTR family, response regulator	0.0017	0.0066	0.0000	0.0030	0.0451	0.0384	0.0366	0.0332	0.0081
dctD; two-component system, NtrC family, C4-dicarboxylate transport response regulator	0.0004	0.0095	0.0080	0.0054	0.1543	0.0868	0.1249	0.0764	0.0344
algB; two-component system, NtrC family, response regulator	0.0008	0.0004	0.0000	0.0006	0.0122	0.0057	0.0108	0.0033	0.0040
divK; two-component system, cell cycle response regulator	0.0017	0.0062	0.0067	0.0066	0.0173	0.0128	0.0409	0.0266	0.0142
cpdR; two-component system, cell cycle response regulator	0.0004	0.0062	0.0020	0.0042	0.0057	0.0185	0.0237	0.0133	0.0101
cheBR; two-component system, chemotaxis family, CheB/CheR fusion	0.0012	0.0099	0.0073	0.0018	0.0308	0.0726	0.0409	0.0432	0.0162

143

protein [EC:2.1.1.80 3.1.1.61]									
fixL; two-component system, LuxR family, sensor kinase [EC:2.7.13.3]	0.0000	0.0011	0.0013	0.0006	0.0186	0.0413	0.0538	0.0199	0.0162
adeS; two-component system, OmpR family, sensor histidine kinase [EC:2.7.13.3]	0.0000	0.0000	0.0000	0.0000	0.0018	0.0100	0.0172	0.0100	0.0061
pfeS, pirS; two-component system, OmpR family, sensor histidine kinase PfeS [EC:2.7.13.3]	0.0012	0.0007	0.0027	0.0024	0.0287	0.0100	0.0065	0.0166	0.0162
pfeR, pirR; two-component system, OmpR family, response regulator	0.0000	0.0004	0.0000	0.0000	0.0091	0.0057	0.0086	0.0033	0.0101
NIK1, TCSC; osomolarity two-component system, sensor histidine kinase NIK1 [EC:2.7.13.3]	0.0008	0.0007	0.0013	0.0018	0.0023	0.0270	0.0258	0.0233	0.0222
sagS; two-component system, sensor histidine kinase [EC:2.7.13.3]	0.0012	0.0018	0.0007	0.0006	0.0113	0.0185	0.0108	0.0066	0.0142
K20974; two-component system, sensor histidine kinase [EC:2.7.13.3]	0.0000	0.0029	0.0033	0.0072	0.0599	0.0498	0.0194	0.0199	0.0243
K20975; two-component system, sensor histidine kinase [EC:2.7.13.3]	0.0000	0.0018	0.0060	0.0006	0.0851	0.0270	0.0280	0.0332	0.0121
hsbR; two-component system, HptB-dependent secretion and biofilm response regulator	0.0017	0.0055	0.0013	0.0030	0.0303	0.0185	0.0129	0.0399	0.0283

*Acer = *Acropora cervicornis*

&Pstrig = *Psuedoploria strigosa*

+Mcav = *Montastraea cavernosa*

#Ofav = *Orbicella faveolata*

$Past = *Porites astreoides*

Additional Supporting Methods: Seawater volume experiment

Sample collection and processing

Surface seawater samples of different volumes (60 mL, 1.5 L, and 2 L) were collected from two different reef sites in St. John, U.S. Virgin Islands in October 2016. Two 1.5 L samples and four 60 mL samples were collected from the 'Dock' location and two 2 L samples and five 60 mL samples were collected from Tektite reef. These samples were then filtered onto 0.2 µm Supor filters using a peristaltic pump, DNA was extracted using the Sucrose-EDTA lysis method (Santoro et al. 2010), and this DNA was prepared for Fluidigm® amplification of the SSU rRNA gene and 2x250 bp MiSeq Illumina sequencing at the Keck Center for Functional Genomics (University of Illinois, Urbana, IL) using the V4 primer pair 515F-Y (5'-GTGYCAGCMGCCGCGGTAA-3') and 806RB (5'-GGACTACNVGGGTWTCTAAT-3') (Apprill et al. 2015; Parada et al. 2016).

Microbial community analysis using MED and 97% similarity OTU clustering
After sequencing, the SSU rRNA gene amplicon sequences were processed using two different clustering methods: Minimum Entropy Decomposition (MED) (Eren et al. 2015) and 97% similarity Operational Taxonomic Unit (OTU) clustering. These clustering methods were chosen in order to investigate how different sampling volumes impacted the microbial diversity and composition using both methods of clustering. Prior to clustering, sequences were processed using the same pipeline that was used for analyzing the RSW and CSW samples collected in Cuba. MED analysis of these sequences was also conducted using the methods described in the methods section of this manuscript. We used mothur v.1.36.1 (Schloss et al. 2009) and guidance from the mothur MiSeq SOP webpage (https://www.mothur.org/wiki/MiSeq_SOP) to accomplish 97% similarity OTU clustering on this data. We also chose to complete clustering with non-subsampled and subsampled data (4940 sequences subsampled from each sample) to determine how subsampling impacted microbial community composition.

After clustering the sequences into either MED nodes or OTUs (97% similarity), a variety of methods were used to examine if differences in original seawater collection volume contributed to significant differences between the microbial communities. Microbial community composition data obtained using MED was analyzed with PhyloSeq, Vegan, and DESeq2 in the R environment with the same code used for analysis in this manuscript (McMurdie et al. 2013; Love et al. 2014; Oksanen et al. 2017). More specifically, we completed non-metric multidimensional scaling analysis (NMDS), tested differences in microbial community similarity using Permutational Multivariate Analysis of Variance tests using distance matrices (PERMANOVA/ADONIS) (Oksanen et al. 2017), and completed DESeq2 (Love et al. 2014) to test for impacts of different volumes on the microbial communities.

Additional Supporting Results: Seawater volume experiment

Microbial community alpha diversity
The 1.5 and 2 L seawater samples collected at the Dock and Tektite had greater species richness (absolute number of unique MED nodes) than the corresponding 60 mL samples that were collected at both of the sites (Figure S7). A Kruskal-Wallis rank sum test revealed statistically significant (p = 0.04) differences in MED node alpha diversity between the different sampling volumes and sites, but post-hoc pairwise Dunn's tests did not detect significant differences between any of the individual sample types (Figure S8). Individual Kruskal-Wallis rank sum tests were also conducted for samples collected at each specific site (Figures S9 and S10). Both

these tests revealed that there significant differences between the number of observed MED nodes as a factor of seawater sampling volume, although these differences were slight (Figures S9 and S10).

The impact of clustering method (97% OTU or MED) and depth of subsampling on alpha richness of the microbial community was also tested (Figure S11). OTU clustering of the non-subsampled and subsampled sequences yielded the highest alpha richness in samples collected from the Dock site compared to MED clustering that was preformed on the same sequences (Figure S11). In contrast, MED node clustering resulted in the highest alpha diversity in samples collected from the Tektite site (Figure S11). Within each site, larger volume samples had higher species richness, but there was also a larger discrepancy between the un-subsampled and subsampled dataset in comparison to 60 mL samples. A Kruskal-Wallis rank sum test indicated that there were significant differences by clustering type (97% OTU, subsampling OTU, or MED, $p < 0.05$), but no significant differences were found during pairwise post-hoc Dunn's testing with Bonferroni corrections.

Microbial community composition and beta diversity
Non-metric multidimensional scaling analysis (NMDS) revealed that microbial communities collected from the same site were more similar to each other than to samples collected at the other site (Figure S12). Within each site, larger volume samples clustered together, whereas the 60 mL samples were mostly evenly distributed from each other along the NMDS1 axis (Figure S12). An Adonis test revealed that site of collection significantly influenced microbial community composition whereas collection volume did not significantly contribute to these differences (Figure S13).

Lastly, no significant enrichment or depletion of MED nodes was detected between the 60 mL and 1.5 L or 2 L samples in DESeq2 comparisons.
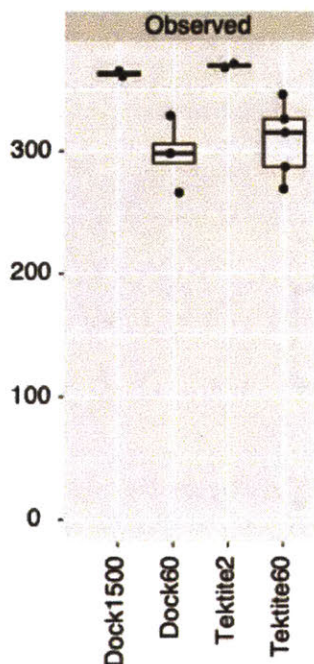
Figure S7. Boxplots of the number of observed MED nodes (absolute count of unique nodes within each sample) for seawater samples collected with different volumes of water. The absolute count of MED nodes is depicted on the y-axis and the sample type is indicated on the x-axis. Dock1500 = 1.5 L samples collected from the Dock, Dock60 = 60 mL samples collected from the Dock, Tektite2 = 2.0 L samples collected from Tektite, Tektite60 = 60 mL samples collected from Tektite. The lower and upper edges of the box correspond to the first and third quartiles and the middle black bar reflects the median. Points that fall outside of the whiskers extend beyond 1.5 X the interquartile range of the box plot.

```
    Kruskal-Wallis rank sum test

data: x and group
Kruskal-Wallis chi-squared = 8.119, df = 3, p-value = 0.04



                        Comparison of x by group
                              (Bonferroni)
Col Mean-|
Row Mean |   Dock1500    Dock60   Tektite2
---------+---------------------------------
  Dock60 |   1.781447
         |     0.2245
         |
Tektite2 |  -0.514259  -2.375262
         |     1.0000     0.0526
         |
Tektite6 |   1.567376  -0.344975   2.182034
         |     0.3511     1.0000     0.0873


List of pairwise comparisons: Z statistic (adjusted p-value)
---------------------------------------
Dock1500 - Dock60    :   1.781447 (0.2245)
Dock1500 - Tektite2  :  -0.514259 (1.0000)
Dock60 - Tektite2    :  -2.375262 (0.0526)
Dock1500 - Tektite60 :   1.567376 (0.3511)
Dock60 - Tektite60   :  -0.344975 (1.0000)
Tektite2 - Tektite60 :   2.182034 (0.0873)

alpha = 0.05
Reject Ho if p <= alpha/2
```

Figure S8. Screenshot of results from a Kruskal-Wallis rank sum test that was performed on MED node alpha diversity within each sample. Both sites (Dock and Tektite) and sampling volumes (60 mL, 1.5 L, and 2 L) were tested.

```
Kruskal-Wallis rank sum test

data: x and group
Kruskal-Wallis chi-squared = 3.5294, df = 1, p-value = 0.06



                              Comparison of x by group
                                   (Bonferroni)
Col Mean-I
Row Mean I         1500
---------+-----------
      60 I    1.878672
         I       0.0301


List of pairwise comparisons: Z statistic (adjusted p-value)
----------------------------------
1500 - 60 :   1.878672 (0.0301)

alpha = 0.05
Reject Ho if p <= alpha/2
```

Figure S9. Screenshot of results from a Kruskal-Wallis rank sum test that was performed on MED node alpha diversity at samples collected from the Dock site.

```
Kruskal-Wallis rank sum test

data: x and group
Kruskal-Wallis chi-squared = 3.75, df = 1, p-value = 0.05



                          Comparison of x by group
                               (Bonferroni)
Col Mean-I
Row Mean I         2
---------+-----------
      60 I    1.936491
         I       0.0264


List of pairwise comparisons: Z statistic (adjusted p-value)
--------------------------
2 - 60 :   1.936491 (0.0264)

alpha = 0.05
Reject Ho if p <= alpha/2
```

Figure S10. Screenshot of results from a Kruskal-Wallis rank sum test that was performed on MED node alpha diversity at samples collected from Tektite.
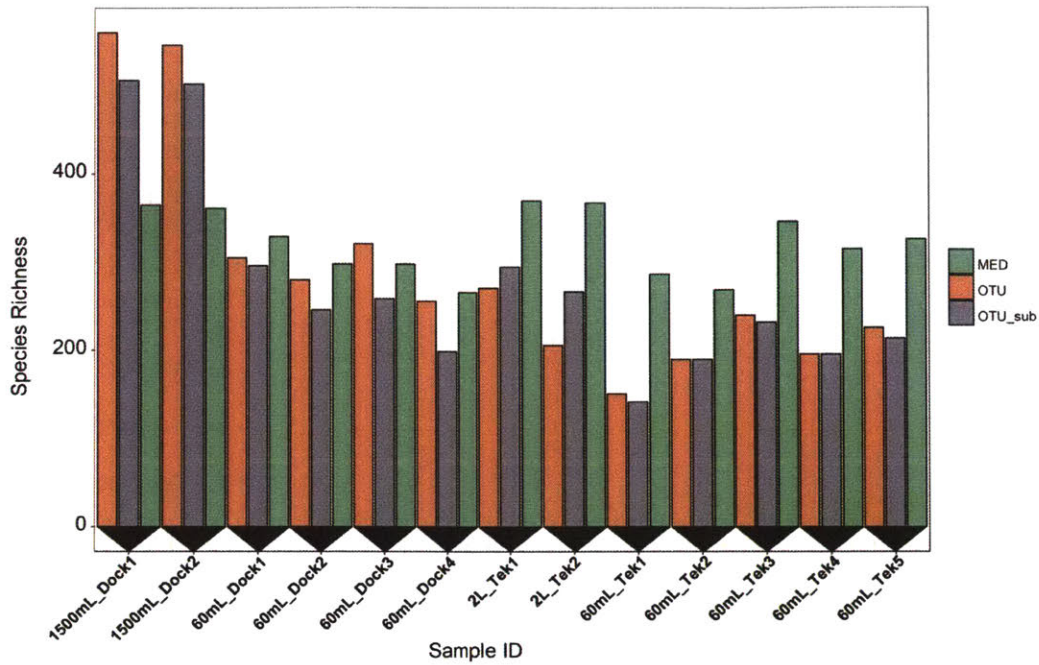
Figure S11. Comparison of microbial species richness (absolute count of unique species within each sample) by sample volume, site, and clustering method. Bars are colored by the clustering method used to analyze the sequences. MED = Minimum Entropy Decomposition, OTU = 97% OTU similarity clustering with no subsampling, and OTU_sub = 97% OTU similarity clustering with subsampling at 4940 sequences per sample.
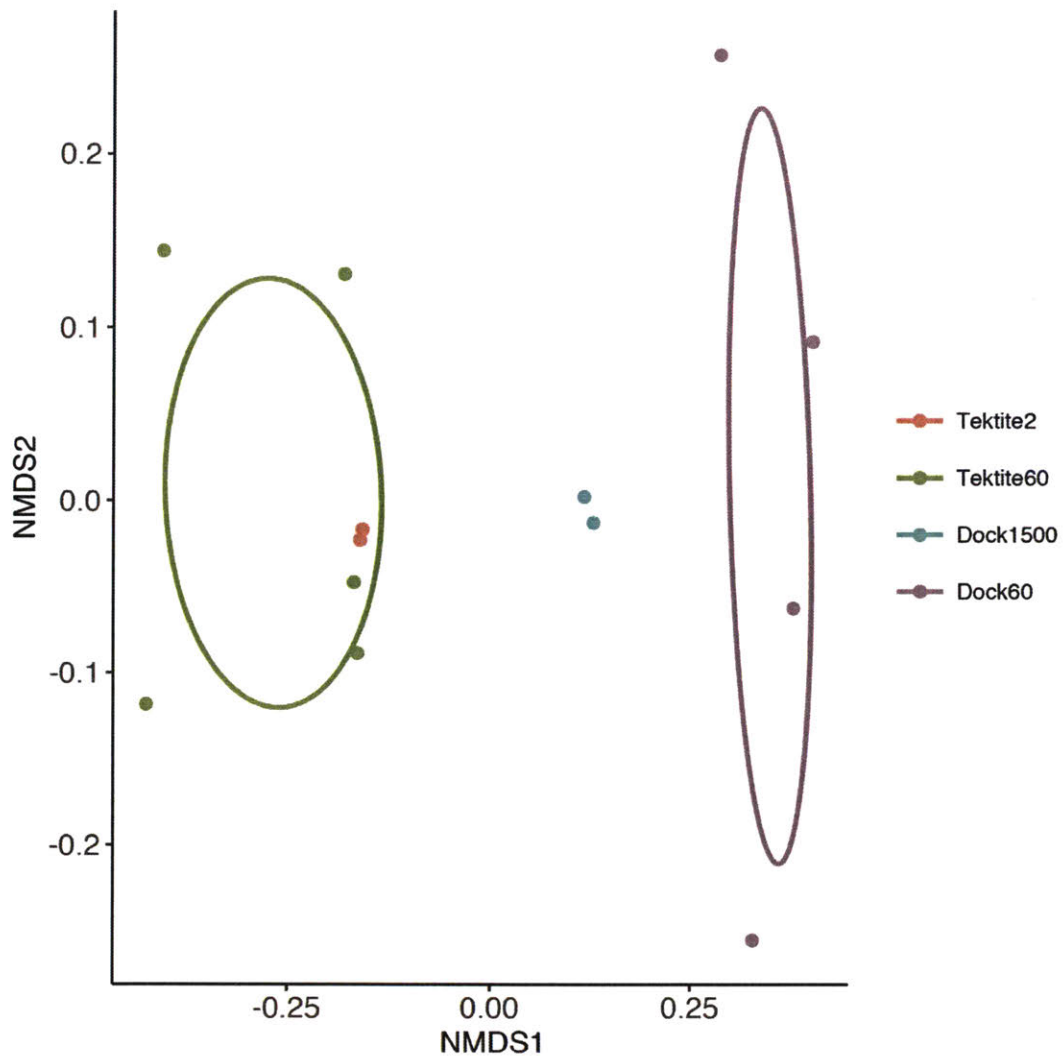
Figure S12. Non-metric multidimensional scaling (NMDS) was performed using a Bray-Curtis dissimilarity matrix that was obtained from square-root transformed microbial community SSU rRNA gene amplicon data. Ellipses are drawn using the group mean and covariance for each species (Eren et al., 2015). Different colors reflect the different sample types (site of collection and seawater volume). 2-D stress: 0.070.

```
Call:
adonis(formula = d_all ~ site + samplingvol, data = df_all)

Permutation: free
Number of permutations: 999

Terms added sequentially (first to last)

            Df SumsOfSqs MeanSqs F.Model     R2 Pr(>F)
site         1   0.41303 0.41303  6.1322 0.36026  0.001 ***
samplingvol  1   0.05990 0.05990  0.8893 0.05224  0.550
Residuals   10   0.67355 0.06735         0.58749
Total       12   1.14648                 1.00000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```
Figure S13. Screenshot of PERMANOVA (Adonis) results from the test that was performed on the square-root transformed Bray-Curtis dissimilarity matrix.

References

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.

Bushnell, B. (2016) BBMap short read aligner.

Clarke, J. (2009) Cetyltrimethyl ammonium bromide (CTAB) DNA miniprep for plant DNA isolation. *Cold Spring Harb Protoc.*

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

Eren, A.M., Morrison, H.G., Lescault, P.J., Reveillaud, J., Vineis, J.H., and Sogin, M.L. (2015) Minimum entropy decomposition: unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences. *ISME J* **9**: 968-979.

Gu, Z., Eils, R., and Schlesner, M. (2016) Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**: 2847-2849.

Kim, J., Kim, M.S., Koh, A.Y., Xie, Y., and Zhan, X. (2016) FMAP: Functional Mapping and Analysis Pipeline for metagenomics and metatranscriptomics studies. *BMC Bioinformatics* **17**: 420.

Love, M.I., Huber, W., and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**: 550.

Marie, D., Partensky, F., Jacquet, S., and Vaulot, D. (1997) Enumeration and cell cycle analysis of natural populations of marine picoplankton by flow cytometry using the nucleic acid stain SYBR Green I. *Appl Environ Microb* **63**: 186-193.

McMurdie, P.J., and Holmes, S. (2013) phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* **8**: e61217.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D. *et al.* (2017) vegan: community ecology package. R package version 2.4-5.

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.

Pepe-Ranney, C., and Hall, E.K. (2015) The effect of carbon subsidies on marine planktonic niche partitioning and recruitment during biofilm assembly. *Front Microbiol* **6**: 703.

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

Santoro, A.E., Casciotti, K.L., and Francis, C.A. (2010) Activity, abundance and diversity of nitrifying archaea and bacteria in the central California Current. *Environ Microbiol* **12**: 1989-2006.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B. *et al.* (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75**: 7537-7541.

Urakawa, H., Martens-Habbena, W., and Stahl, D.A. (2010) High abundance of ammonia-oxidizing Archaea in coastal waters, determined using a modified DNA extraction method. *Appl Environ Microbiol* **76**: 2129-2135.

William, S., Feil, H., and Copeland, A. (2004) Bacterial DNA isolation CTAB protocol bacterial genomic DNA isolation using CTAB materials & reagents.

Zhou, J., Bruns, M.A., and Tiedje, J.M. (1996) DNA recovery from soils of diverse composition. *Appl Environ Microbiol* **62**: 316-322.

Chapter 4

Diel, daily, and spatial variation of coral reef seawater microbial communities

4.1 Abstract

Reef organisms influence the microorganisms within the surrounding seawater, yet the spatial and temporal dynamics of seawater microbial communities located in proximity to corals are rarely investigated. To better understand reef seawater microbial community dynamics over time and space, we collected small-volume seawater samples during the day and night over a 72 hour period from three locations that differed in spatial distance from corals on a shallow reef in St. John, U.S. Virgin Islands: near-coral (within 5 cm of 5 different *Porites astreoides* colonies), reef-depth (within 2 m of each coral colony) and surface seawater (within 1 m from surface). At all timepoints and locations, we quantified abundances of microbial cells, sequenced SSU rRNA genes of bacterial and archaeal communities, and measured inorganic nutrient concentrations. *Prochlorococcus* and *Synechococcus* cells were consistently elevated at night compared to day and these abundances changed over time, corresponding with temperature, nitrite, and silicate concentrations. During the day, bacterial and archaeal alpha diversity was significantly higher in reef-depth and near-coral seawater compared to the surface seawater, signifying that the reef influences the diversity of the seawater microorganisms. At night, alpha diversity decreased across all samples, suggesting that photosynthesis may favor a more taxonomically diverse community. While *Prochlorococcus* and *Synechococcus* exhibited temporal rhythmicity, additional taxa were significantly enriched in reef seawater at night compared to day or in reef-depth compared to surface seawater. Despite some variation, there were no significant differences in microbial community composition, nutrient concentrations, or cell abundances between reef-depth and near-coral seawater. This study demonstrates that temporal variation supersedes small-scale spatial variation in proximity to corals in reef seawater bacterial and archaeal communities. As coral reefs continue to change in composition worldwide, monitoring microbial composition in response to temporal changes and environmental fluctuations will help discern normal variability from longer lasting changes in reef microbial dynamics.

4.2 Introduction

Microorganisms play fundamental roles in coral reef ecosystem nutrient cycling (Sorokin 1973; Atkinson 2010). Microbial processes on coral reefs take place both in the reef benthos and within the water column. In the water column, picocyanobacteria and eukaryotic phytoplankton fix carbon into biomass through the process of photosynthesis. A significant fraction of this photosynthetically fixed carbon is released from cells through grazing, viral lysis, or exudation (reviewed by Thornton 2014). Heterotrophic bacteria in the water column respire organic matter released through all these processes and recycle inorganic nutrients back into the dissolved phase (Azam and Malfatti 2007; Haas et al. 2013; Nakajima et al. 2017). Within the reef benthos, symbiotic dinoflagellates residing within corals translocate photosynthate to the host and corals

use this photosynthate for their own metabolisms, exuding dissolved and particulate organic matter (OM) into the water column (Falkowski et al. 1984; Wild et al. 2004; Tanaka et al. 2009; Tremblay et al. 2012). Macroalgae and other benthic organisms are also sources of dissolved OM (Haas et al. 2011; Haas et al. 2013). Additionally, microorganisms within coral reef sediments mineralize carbon, fix nitrogen, and photosynthesize depending on their location in the sediment and the presence of oxygen (Werner et al. 2006; Werner et al. 2008).

Most of our knowledge of reef seawater microbial community composition and function is obtained from opportunistic field sampling of reef seawater collected across reefs at a single point in time. This sampling approach has contributed knowledge about coral reef seawater microbial dynamics along various environmental and anthropogenic gradients (Dinsdale et al. 2008; Nelson et al. 2011; Kelly et al. 2014), but does not allow for an understanding of resilience, resistance, and/or change in the same communities over time. Currently, coral reef ecosystems are experiencing dramatic shifts in reef composition (Gardner et al. 2003), fish biomass (Valdivia et al. 2017) and nutrient availability (reviewed within D'Angelo and Wiedenmann 2014) due to a variety of human-induced stressors including global climate change, pollution, coastal development, and overfishing. There is also evidence that a corresponding trophic shift is occurring within the microbial loop in coral reef ecosystems, favoring microbial communities that are dominated by more copiotrophic heterotrophic bacteria and potentially pathogenic taxa (Haas et al. 2016). Due to the lack of microbial time-series studies and the observed trophic shifts in coral reef ecosystems, we have a limited understanding of baseline temporal variability of microbial community composition and function over time, further complicating our ability to discern consistent and recurring variability from ecosystem shifts over longer timescales.

On coral reefs, there are significant temporal changes due to the presence of light (and primary production), influences of tides, and diel vertical migration and grazing by zooplankton (Yahel et al. 2005). The presence of light is a major driver of net community metabolism on reefs, leading to primary production during the day and respiration at night (DeCarlo et al. 2017; Silbiger et al. 2018). This metabolic switch also influences pH, oxygen concentrations, and aragonite saturation state within the seawater over a diel cycle (Guadayol et al. 2014; DeCarlo et al. 2017; Silbiger et al. 2018). Additionally, cyanobacterial mats on reefs release a significant amount of dissolved organic carbon into the water column at night (Brocke et al. 2015). Bearing in mind the diel fluctuations in light availability, net community metabolism, and grazing as well as other processes on a reef, we would expect corresponding changes in the microbial community.

Despite these diel fluctuations, there are only a handful of studies that have catalogued changes in microbial communities over the diel cycle. For example, the microbial community diversity in seawater sampled over a reef flat and adjacent to colonies of *Acropora formosa* changed between day and night (Sweet et al. 2010). In contrast, seawater microbial communities sampled adjacent to *Mussimillia braziliensis* showed no significant changes in composition or functional potential over a diurnal cycle spanning 48 hours (Silveira et al. 2017). It is surprising that neither of these studies reported changes in the abundance of photosynthetic picocyanobacteria between day and night, considering the prevalence of these cells on some coral reefs (Charpy et al. 2012) and their diel fluctuations in cell abundance in oligotrophic tropical ocean gyres (Vaulot et al. 1995). A more recent study of seawater microbial communities sampled from forereefs in the Pacific Ocean detected synchronous changes in microbial community composition and function over one diel cycle (Kelly et al. 2019). They

found consistent enrichment of specific taxa during both day and night and more genes indicative of diverse strategies for carbohydrate metabolism and general catabolism at night (Kelly et al. 2019), demonstrating a shift in net metabolism of the collective microbial community over a diel cycle. That being said, no studies have tracked changes in reef seawater microbial community composition over a longer diel time-series, making it difficult to assess consistent diel and daily shifts in microbial biomass and community composition over time.

Reef seawater microbial communities can also vary in cell biomass, community composition, potential function, and growth dynamics based on water depth and proximity to reef organisms (Moriarty et al. 1985; Seymour et al. 2005; Tout et al. 2014). Seawater microbial communities located adjacent to corals are exposed to slightly different environmental and nutrient conditions as a result of exudation of organic matter and mucus from corals (Wild et al. 2004; Haas et al. 2013), local changes in temperature, light availability (Anthony and Hoegh-Guldberg 2003), and water flow close to coral colonies (Shashar et al. 1996). These conditions may impact microbial community composition as well as potential microbial functions in the seawater surrounding corals at the scale of the momentum boundary layer (Tout et al. 2014; Silveira et al. 2017; Ochsenkuhn et al. 2018), a layer of water surrounding the coral that is influenced by coral morphology and micro-currents caused by animal activity within the coral (Shashar et al. 1996). In fact, there is evidence that distinct microbial environments exist within 30 cm surrounding coral colonies in an environment called the coral ecosphere (Weber et al. 2019). For example, coral ecosphere microbial communities were generally enriched with copiotrophic Gammaproteobacteria compared to microbial communities sampled from water >1 meter above the reef (Weber et al. 2019) and this finding corroborated earlier observations of copiotrophic enrichment in the seawater adjacent to corals (Tout et al. 2014). Additionally, the

159

Gammaproteobacteria *Endozoicomonas*, a ubiquitous coral tissue and mucus symbiont, was

enriched within the ecosphere surrounding *P. astreoides,* indicating that the ecosphere

environment may serve as a reservoir for coral symbionts and pathogens (Weber et al. 2019).

Furthermore, environmental conditions in the seawater surrounding corals also change at night

due to decreased oxygen and pH in the diffusive boundary layer (Shashar et al. 1993; Smith et al.

2013), decreased exudation of DOM (Kurihara et al. 2018), and heterotrophic feeding by some

coral species (reviewed within Houlbreque and Ferrier-Pages 2009). Diel changes in these

conditions may lead to diel changes in composition that are unique to coral ecosphere microbial

communities compared to communities sampled from seawater further away.

The present study was designed to compare diel, daily, and spatial variations in microbial

cell abundances, inorganic macronutrient concentrations, and microbial community diversity and

composition over the course of 3 days across three different environments including surface,

reef-depth, and near-coral (5 cm away from individual coral colonies; coral ecosphere) reef

seawater. We hypothesized that overall community composition would change temporally over

diel and daily time-scales. Additionally, we expected that coral ecosphere microbial communities

would be enriched with Gammaproteobacteria compared to reef-depth and surface seawater

communities and that these communities would shift over a diel cycle in relation to potential

changes in environmental conditions close to the coral surface.

4.3 Materials and Methods

Sample collection

Five *Porites astreoides* colonies and a sand patch were selected and marked with flagging tape

by divers on Ram Head reef (18°18'07.3" N, 64°42'14.5" W; 8 m depth in sand) in St. John, U.

S. Virgin Islands. Colonies of various sizes (3 – 16 inches in diameter) from a range of heights above the sea floor (1 – 27 cm) were selected and these colonies were labeled A through E (Figure 1). Additionally, colonies were evenly distributed across the reef in order to minimize location effects (range of 3.6 to 14 meters between each colony). All colonies were located directly next to sand patches based on colony size constraints and the space needed for deployment of the custom-made Coral Ecosphere Sampling Devices (CESD) (Figure 1). Six CESD made out of aluminum strut material were deployed adjacent to each sampling location with sand screws. The last CESD was placed in a wide sand patch with no corals or benthic organisms located in its vicinity and this sampling location was used as a 'no-coral' control. Divers positioned the CESD so that a 60 ml syringe with an attached filter holder could be placed 5 cm away from the middle of the colony (Figure 1). Light and temperature loggers (8K HOBO/PAR loggers; Onset, Wareham, MA) were zip-tied to the end of each CESD and programmed to collect temperature and relative light intensity measurements every 5 minutes over the course of the three day study.

An hour after CESD deployment, scuba divers collected the first set of samples (Day 1, 3:00 pm). Filter holders were pre-loaded with 0.22 μm pore size Supor® filters (Pall Corporation, Ann Arbor, MI, USA) and were contained within sterile Whirl-pack® bags prior to sampling. Divers also descended with acid-washed polyethylene nutrient bottles (30 ml volume) to collect seawater samples for unfiltered inorganic nutrient analysis and flow cytometry. Samples were also collected for analysis of total organic carbon, but are not included in this study because they became contaminated during sample storage. At depth, seawater samples (60 ml) collected for amplicon-based microbial community analyses were conducted at 2 different stationary locations relative to the CESD device (with the exception of collections completed at the sand-patch

location). Reef-depth samples were collected first at the top of the CESD (2 m from the colony) in order to minimize stirring close to the coral ecosphere sampling area (Figure 1). To collect the sample, a diver attached a piece of acid-cleaned Masterflex silicone tubing to connect the end of the filter holder to the mouth of the syringe and then used reverse filtration to pull seawater through the filter. The filter-holder was then placed in an individual Whirl-pack® bag and sealed. After collection of microbial biomass with the syringe, a nutrient sample was collected. After collection of the reef-depth sample, a diver attached the filter holder to the syringe, slowly descended closer to the coral colony, but behind the CESD to maintain sufficient distance from the sampling area and then placed the syringe into the syringe holder located on the horizontal arm of the CESD. As before, the diver first collected the coral ecosphere sample (5 cm from the colony) onto the filter followed by a nutrient sample in the same location (Figure 1). Replicate samples collected for microbial community analyses were collected from both seawater environments surrounding each colony on the first dive, but were not collected on the following dives due to time constraints. Surface seawater samples (< 1 m) were collected using 60 mL syringes at each time point from the dive boat.

This sampling scheme was repeated at approximately 3 am and 3 pm for the next three days, totaling up to 6 sampling time points. Divers sampled each colony and collected samples in the same order (reef-depth followed by coral ecosphere) during all time points. After collection, samples were placed in a cooler equipped with blue-ice packs for the transit from the reef to the lab and then samples were processed immediately. Over the course of sampling, 85 seawater samples for microbial community analyses were collected.

After the last time point, coral tissue was collected from each colony (close to the area where the coral ecosphere seawater was sampled) using a hammer and chisel and the CESD were

removed. Sand was also collected in the location where the sand control CESD device was deployed.

Sample processing

In the laboratory, sterile syringes were used to remove residual seawater trapped within filter holders and then filters were placed into cryovials, flash-frozen in a dry shipper charged with liquid nitrogen, and then transferred into a -20° C freezer.

Seawater collected for flow cytometric analysis was subsampled from unfiltered nutrient samples and preserved with paraformaldehyde (Electron Microscopy Sciences, Allentown, PA) to a final concentration of 1% (by volume). Nutrient, DNA, and flow cytometry samples were shipped frozen back to Woods Hole Oceanographic Institution and ultimately stored at -80 °C prior to analysis. The coral tissue and sand samples were stored in a second dry shipper and ultimately at -80 °C until they were processed.

Macronutrient analysis and flow cytometry

Frozen and unfiltered nutrient samples were analyzed with a continuous segmented flow-system using previously described methods (as in Apprill and Rappé 2011). The concentrations of nitrite + nitrate, nitrite, phosphate, ammonium, and silicate were measured in all of the samples. Nitrate concentrations were obtained by subtracting the nitrite concentration from the nitrite + nitrate measurements for each sample.

Samples collected for flow cytometry were analyzed using colinear analysis (laser excitation wavelength of 488 nm, UV) on an Altra flow cytometer (Beckman Coulter, Pasadena, CA). Unstained subsamples were used to enumerate the abundances of picocyanobacteria (*Prochlorococcus, Synechococcus*) and picoeukaryotes. Stained (Hoechst stain, 1 µg ml^{-1} final concentration) subsamples were analyzed to estimate the abundance of unpigmented cells (an

163

estimate of heterotrophic bacterial abundance) (Marie et al. 1997). FlowJo (v. 6.4.7) software was used to estimate the abundance of each cell type. The abundance of total cells was calculated by adding the cell counts obtained for each of the respective picoplankton classes together for each sample.

DNA extraction, amplification, pooling, and sequencing

DNA was extracted from filters using a sucrose-lysis extraction method and Qiagen spin-columns (Santoro et al. 2010). Control extractions were also completed with unused filters (filters without biomass) in order to account for contamination from the filters or extraction reagents. Lastly, diluted DNA from a synthetic staggered mock community (BEI Resources, Manassas, VA, USA) was used to account for amplification and sequencing errors in downstream microbial community analyses. Coral tissue was removed from the skeleton using air-brushing with autoclaved 1% phosphate-buffered-saline (PBS) solution (Apprill et al. 2016; Weber et al. 2017). The coral tissue slurry was pelleted using a centrifuge and the PBS supernatant was discarded. DNA was extracted from each pellet (300 mg of tissue) using a modified version of the DNeasy DNA extraction kit protocol (Qiagen, Germantown, MD). The lysis buffer in the kit was added to each tube followed by approximately 300 mg of garnet beads (from a MOBIO DNA extraction kit) and 300 mg of Lysing B matrix beads (MP Biomedicals, Solon, OH). The tubes were subjected to a bead-beating step for 15 minutes so that the beads could break up the coral tissue (Weber et al. 2017). After bead-beating, 20 μl of proteinase-k was added to each tube and the samples were incubated with gentle agitation for 10 minutes at 56 °C. After these modifications, the DNeasy protocol (Qiagen) was followed to complete extractions.

Extracts were amplified with barcoded primers targeting the V4 hypervariable region of the bacterial and archaeal small subunit ribosomal RNA gene (Kozich et al. 2013). The forward

primer: 5' TATGGTAATTGTGTGYCAGCMGCCGCGGTAA 3' (Parada et al. 2016) and

reverse primer: 3' AGTCAGTCAGCCGGACTACNVGGGTWTCTAAT 5' (Apprill et al. 2015)

were used, along with the barcodes, to amplify and tag each sample prior to pooling. We used

forward and reverse primers with degeneracies in order to eliminate amplification biases against

Crenarchaeota/ Thaumarchaeota (Parada et al. 2016) and SAR 11 (Apprill et al. 2015). Triplicate

polymerase chain reactions (25 µl volume) were run with 2 µl of DNA template from each

sample using the same barcodes in order to minimize the formation of chimeras during

amplification. The reaction conditions included: a 2 minute hot start at 95 °C followed by 36

cycles of 95 °C for 20 seconds, 55 °C for 15 seconds, and 72 °C for 5 minutes. The final

extension step was 72 °C for 10 minutes. Triplicate barcoded amplicons were pooled and

screened using gel electrophoresis to assess quality and assess amplicon size. Amplicons were

purified using the MinElute Gel Extraction Kit (Qiagen) and pooled to form the sequencing

library. The library was sequenced (paired-end 2x250 bp) at the Georgia Genomics and

Bioinformatics Core with a Miseq (Illumina, San Diego, CA) sequencer and raw sequence reads

are available at the NCBI Sequence Read Archive under BioProject # PRJNA550343.

Microbial community analyses

Raw sequences were quality-filtered and grouped into amplicon sequence variants (ASVs) using

DADA2 (Callahan et al. 2016). Reads were filtered (maxN = 0, truncQ=2, rm.phix = TRUE, and

maxEE = 2), trimmed, and dereplicated. The DADA2 algorithm was used to infer the number of

different ASVs (8357 distinct ASVs), paired reads were merged, an ASV table was constructed,

and chimeras were removed (1% of all ASVs). Taxonomy was assigned to each ASV using the

Silva v.132 reference database (Quast et al. 2013). Mock communities were used to assess the

performance of the program as well as sequencing error rates. DADA2 inferred 15, 17, and 17

strains within the mock community (compared to the 20 expected stains present at different concentrations within the staggered community) and 13, 14, and 14 of the strains were exact matches to the expected sequences from the mock community reference file. Sequence recovery is slightly lower than expected, but is comparable to normal performance of DADA2 on this staggered mock community (Callahan et al. 2016).

The R packages Phyloseq (McMurdie and Holmes 2013), Vegan (Oksanen et al. 2017), DESeq2 (Love et al. 2014), and ggplot2 (Wickham 2016) were used for downstream analyses of the microbial community. Sequences were not subsampled, but samples with less than 1000 reads (2 samples) were removed. In addition, ASVs identifying as chloroplasts were removed. Sequences representing ASVs that identified as "NA" at the Phylum level were checked using the SINA aligner and classifier (v.1.2.11) (Pruesse et al. 2012) and then removed if not identified as bacteria or archaea at 70% similarity. The average number of reads across all seawater samples used in microbial community analyses was 58,398 ($\pm$ 32,184 standard deviation) with a range of 11,502 – 206,689 reads. The average number of reads in coral tissue samples was 38,096 ($\pm$23,854) with a range of 11,538 – 59,437 reads. DNA extraction control communities were initially inspected and then removed because they fell out as outliers compared to the highly similar seawater microbial communities. Taxonomic bar plots, metrics of alpha diversity (observed richness of ASVs), and boxplots of alpha diversity were made and calculated using Phyloseq. Constrained analysis of principal coordinates (CAP) based on Bray – Curtis dissimilarity was completed (using 'capscale' in Vegan) and variance partitioning was used to identify which of the measured environmental parameters significantly ($p < 0.01$) contributed to shifts in the microbial community composition over time. Permutational Multivariate Analysis of Variance using distance matrices (PERMANOVA/Adonis) tests identified categorical factors

that significantly (p<0.05) contributed to similarity between the microbial communities. DESeq2 was used to identify differentially abundant ASVs between day and night as well as reef-associated (reef-depth and coral ecosphere) compared to surface microbial communities (using the "local" fitType parameter to estimate gene dispersion). Lastly, the Rhythmicity Analysis Incorporating Non-parametric methods (RAIN) R package was used to identify ASVs that experienced rhythmic change in relative abundance over a period of 24 hours (Thaben and Westermark 2014). This analysis was completed separately for reef-depth and coral ecosphere seawater and the input ASV matrix was center log-ratio transformed and detrended following previous methods (Hu et al. 2018). Only ASVs with significant p-values (p < 0.05) after adaptive Benjamini-Hochberg correction were reported to control for false recovery rates (Benjamini and Hochberg 2000).

Statistical analyses

A Principal Coordinates Analysis (PCA) was completed to summarize changes in picoplankton abundances, inorganic nutrient concentrations, and relative light and temperature information collected from the HOBO loggers and reduce the dimensionality of this data. Separate PCAs were also generated using samples collected during either day or night to observe trends specific to these times. Kruskal-Wallis rank sums tests were used to test for significant differences (p < 0.05) in alpha diversity between the different sample groupings. Pairwise post-hoc Dunn's tests with Bonferonni corrections were used to identify which groups were significantly different from each other. These tests were also used to test for significant differences in picoplankton cell abundance over time, between day and night samples, between coral ecosphere and reef-depth samples, and between samples collected surrounding different colonies/ sand. Project data can be accessed using the BCO-DMO repository: 775229 (DOI: 10.1575/1912/bco-dmo.775229.1).

167

4.4 Results

Picoplankton abundances and inorganic macronutrient concentrations

Picoplankton abundances did not significantly differ between coral ecosphere and reef-depth

seawater, but changed between day and night and over the course of the three-day study,

regardless of proximity to corals (Figure 2, Table 1, see S1 for statistical details).

Prochlorococcus and *Synechoccocus* populations showed the strongest diel signal with

abundances increasing significantly at night, compared to abundances measured the previous day

(Figure 2A and B). *Prochlorococcus* consistently doubled at night relative to day, but decreased

throughout the study (Figure 2A). Over the course of the study, day and night abundances of

Synechoccocus and picoeukaryotes increased significantly compared to their initial abundances

(Figure 2B and C, Table 1, S1). The abundances of unpigmented cells were generally similar

over time (Figure 2D). Additionally, there was more spatial variability in the abundances of

Prochlorococcus, *Synechoccocus*, and unpigmented cells compared to the variability in

picoeukaryote abundances sampled across the reef at each time point. There were no significant

trends in picoplankton abundance by colony/sand (A-E, F) or by sampling distance away from

each colony (reef-depth vs. coral ecosphere) (Table 1, S1).

 The concentrations of inorganic macronutrients were low and generally similar between

day and night as well as over time (Figures 3, S1, and S2; Table 1, S1). That being said, silicate

concentrations changed significantly over the course of the study, with the biggest decrease

occurring between day 1 and night 1 (Figure 3B). Ammonium concentrations were significantly

higher in ecosphere seawater compared to reef-depth seawater and were variable, but

concentrations of the other macronutrients did not vary with spatial distance from the coral

colonies (Figure S1, Table 1, S1).

Principal Coordinates Analysis illustrated the diel signal of *Prochlorococcus* and

Synechoccocus abundances, and increased temperature and relative light during the day (Figure

4). The spatial distance from coral colonies had little influence on the measured parameters

(Figure 4A). In addition, there was a correlation between nitrate concentrations and

picoeukaryotic cell abundances (Figure 4A). The PCA completed on samples collected during

the night revealed correlations between *Prochlorococcus* cell abundance with temperature and

Synechococcus cell abundance with nitrite and silicate concentrations (Figure 4C). Day and night

samples from day 3 were each correlated with picoeukaryotes (Figure 4A, B).

Microbial community alpha diversity

The observed bacterial and archaeal richness, the number of unique ASVs identified in each

sample via bacterial and archaeal SSU rRNA gene sequencing, from reef-depth and coral

ecosphere seawater communities during the day was significantly higher than surface seawater

during both day and night as well as over time (Figure 5). Reef-depth seawater had significantly

higher richness during the day compared to night (Figure 5, Table 1, S1). Additionally, richness

during the day in reef-depth and coral ecosphere samples was more variable compared to

communities surveyed at night (Figure 5). Richness was mostly similar across reef-depth or coral

ecosphere microbial communities collected during the same time, although there was more

variability in samples collected on Day 3 (Figure 5). Overall, there was more variable richness in

reef-depth and coral ecosphere seawater compared to surface seawater microbial communities

(Figure 5). There were no differences in bacterial and archaeal richness between coral

ecosphere/reef-depth and sand ecosphere/reef-depth seawater (Table1, Supp. Table 1).

169

Microbial community composition

Reef seawater bacterial and archaeal compositions, assessed using SSU rRNA gene sequencing, were highly similar, but consistently shifted between day and night as well as over time (Figure 6). There were no significant differences in microbial community composition between coral ecosphere/reef-depth and sand ecosphere/reef-depth control seawater. Constrained Analysis of Principal Coordinates (CAP) of the Bray – Curtis dissimilarity matrix and variance partitioning of quantitative environmental variables revealed that temperature (ANOVA, p = 0.023), *Prochlorococcus* abundance (ANOVA, p = 0.001), and *Synechococcus* abundance (ANOVA, p = 0.008) significantly explained shifts in microbial community composition over time (Figure 6). In addition, communities sampled at the same time were most similar to each other, but also ordinated more broadly by day or night (Figure 6). A PERMANOVA (Adonis) test conducted on the Bray – Curtis dissimilarity matrix supported this observation and revealed that the categorical factors of time (day or night), colony sampled, time-series, distance, and the interactions between time and colony and colony and time-series significantly explained microbial community dissimilarity (Table 2).

In terms of taxonomic composition, the average relative abundance of sequences identified as *Synechococcus* CC9902 was higher (30%) than *Prochlorococcus marinus* (MIT9313) (13%) across the time-series (Figure 7A). The average relative abundance of *Prochlorococcus marinus* (MIT9313) was higher at night compared to the day (1.2 times higher). Additionally, the relative abundance of *Synechococcus* CC9902 increased both at night relative to day (1.2 times higher) and over the entire study, aligning with the observed changes in cell abundances of these two groups (Figure 7A). Flavobacteriales and SAR11 sequences

170

(average relative abundances of 5% and 12%, respectively) were detected across samples, with

SAR11 clade 1a sequences being more abundant than clade 1b (Figure 7). Rhodobacterales

HIMB11 sequences were absent during day 1 and night 1, but were detected during subsequent

days and nights at low relative abundances (Figure 7). Sequences identifying as Cellvibrionales

OM60 (NOR5) clade were also detected sporadically and at low relative abundances during all

sampling time points. *Endozoicomonas* and *Vibrio* sequences were detected within a majority of

the coral ecosphere and reef-depth samples at very low average relative abundances (0.11% and

0.09%, respectively). The coral tissue microbial communities were mostly dominated by

Endozoicomonas (average relative abundance of 54%, Figure 7B). Colony C had a more diverse

composition compared to the other colonies and *Vibrio* was detected at a low relative abundance

of 0.5% in this colony (Figure 7B). No ASVs were shared between coral ecosphere or reef-depth

Differential enrichment of taxa

Differential enrichment tests revealed that there were 28 significantly enriched ASVs identified

to the level of genus (p<0.05) between day (9) and night (19) reef-depth and coral ecosphere

seawater bacterial and archaeal communities (Figure 8A, see S2 for ASV sequences).

Interestingly, there were 3 *Synechococcus* CC9902 ASVs that were enriched during the day,

whereas 3 *Synechococcus* CC9902 and 4 *Prochlorococcus marinus* (MIT9313) ASVs were

enriched at night (Figure 8A). Gammaproteobacteria *Marinobacterium, Litoricola* (2), and

Alcinovorax within the order Oceanospirillales were enriched at night (Figure 8A). OM60

(NOR5) clade, *Staphylococcus*, NS4, NS5, and NS2b marine group, HIMB, and "*Candidatus*

Puniceispirillum" ASVs were also enriched in samples collected at night (Figure 8A). In

contrast, *Enterovibrio*, SAR11 clade 1a, and *Marinoscillum* ASVs were only enriched during the

day and the fold changes were higher (Figure 8A).

171

Differential enrichment tests were also conducted between benthic reef seawater

(including reef-depth and coral ecosphere communities) and surface reef seawater (Figure 8B).

There were 20 significantly enriched ASVs identified at the level of genus in bottom reef

seawater compared to surface reef seawater (7 enriched ASVs) (Figure 8B). Overall, ASVs

identified as *Prochlorococcus marinus* (MIT9313), SAR 11 clades 1a and 1b, NS4 marine group,

and "*Candidatus* Actinomarina" were more enriched in bottom reef seawater (Figure 8B). No

ASVs were differentially enriched between coral ecosphere and reef-depth seawater microbial

communities by both day and night or tested individually by either day or night.

Taxa exhibiting diel and daily rhythms

Ten ASVs exhibited significant rhythmicity in reef-depth and coral ecosphere seawater (Table

3). *Prochlorococcus marinus* (MIT9313) and *Synechococcus* ASVs accounted for a majority

(80%) of the taxa that underwent significant synchronous changes in abundance over 24 hours

across both seawater environments (Table 3, see S2 for ASV sequences). Interestingly,

Prochlorococcus marinus (MIT9313) ASVs experienced changes in abundance over a phase of

24 hours compared to *Synechococcus* ASVs which experienced a phase of 12 hours. Aside from

Prochlorococcus marinus (MIT9313) and *Synechococcus*, ASVs identified as *Cyanobium* PCC-

6307 exhibited rhythmic patterns in reef-depth and coral ecosphere seawater. An ASV

identifying within the Pirellulaceae family only experienced rhythmicity in reef-depth seawater

(Table 3).

4.5 Discussion

This study examined diel and daily variation in reef seawater microbial communities and

inorganic nutrient concentrations and also investigated how these factors changed over three

spatial scales: near-coral, reef-depth and surface waters. Overall, the marine cyanobacteria

Prochlorococcus and *Synechococcus* exhibited consistent diel patterns in cell abundance.

Bacterial and archaeal alpha diversity was higher in reef-depth and coral ecosphere seawater

compared to surface seawater and diversity decreased at night. Bacterial and archaeal community

composition of coral ecosphere and reef seawater microbial communities were compositionally

similar, but there were consistent changes in the relative abundances of picocyanobacteria and

differential enrichment of select taxa between day and night (in reef-depth and ecosphere

seawater) as well as between surface and reef-depth seawater. Lastly, there were several

Prochlorococcus and *Synechococcus* amplicon sequence variants (ASVs) in reef-depth and coral

ecosphere seawater that exhibited significant rhythmicity over time.

Diel and daily shifts in Synechococcus *and* Prochlorococcus *abundances*

Synechococcus and *Prochlorococcus* cell abundances were the strongest drivers of microbial

community change across this reef seawater time-series and increased at night relative to day.

Patterns of picocyanobacterial cell division and cell-cycling have been comprehensively

documented in pelagic and oligotrophic regions of the Pacific and Atlantic (Vaulot et al. 1995;

Binder and DuRand 2002; Agawin and Agusti 2005; Ribalet et al. 2015), but not in tropical and

coastal coral reef ecosystems. Our observations of increasing abundances of *Synechococcus* and

Prochlorococcus during the night can be explained by the synchronization between light

irradiance, the cell cycle (Jacquet et al. 2001) and circadian rhythms in *Synechococcus*

populations (Cohen and Golden 2015). *Prochlorococcus* do not have true circadian rhythms, but

do oscillate over a diel cycle with the influence of environmental cues (Cohen et al. 2015). Cell

abundances for both populations were higher at night compared to their daytime abundances, but

the populations were not monitored continuously between these times so the peak cell

abundances could not be estimated. Our data correspond with previously reported diel patterns in cell division of *Synechococcus* and *Prochlorococcus* measured in the eastern equatorial Pacific (Binder et al. 2002), but not with patterns of *Prochlorococcus* abundance in the North Pacific, where populations reached peak abundance during the day in warmer water (Ribalet et al. 2015). The decrease in cyanobacterial abundance during the day could be attributed to physical movement of cells, grazing activities, and/or viral lysis (Binder et al. 2002), but we did not measure these factors here.

Over the course of three days, the abundance of *Prochlorococcus* decreased whereas the abundance of *Synechococcus* increased significantly and there were no correlations between picoplankton abundance and inorganic macronutrient concentrations when all time points were analyzed together. However, when samples were only compared by either day or night, *Prochlorococcus* abundances were correlated with temperature while *Synechococcus* abundances were correlated with nitrite and silicate concentrations at night, aligning with previous observations (DuRand et al. 2001; Cox et al. 2006; Ribalet et al. 2015).

Diel shifts in microbial community composition and diversity

Population fluctuations of *Synechococcus* and *Prochlorococcus* were also evident in the sequencing data and significantly explained compositional shifts in these seawater communities over time. Changes in cell abundances corresponded with changes in the relative abundance of these two groups, demonstrating coherence between flow cytometry and 16S rRNA gene amplicon sequencing in this case. Additionally, decreases in observed community richness at night may partly reflect cell division dynamics in *Synechococcus* and *Prochlorococcus*. Logically, as *Synechococcus* and *Prochlorococcus* populations divide and increase in abundance, the overall diversity of the community decreases because the picocyanobacteria comprise more

of that community. That being said, this trend did not change when we temporarily removed *Prochlorococcus* and *Synechococcus* ASVs from diversity analysis, demonstrating that overall changes in observed diversity occurred across the entire community and not just in the dominant picocyanobacterial populations.

While the community compositions were highly similar, differential enrichment tests of these reef seawater communities (reef-depth and coral ecosphere) between day and night revealed ASVs exhibiting diel enrichment. Excluding the consistent diel changes in *Synechococcus* and *Prochlorococcus*, more ASVs within the orders Oceanospirillales, Flavobacteriales, Puniceispirillales, and Cellvibrionales were enriched at night. During the day, a few ASVs within the same orders of Puniceispirillales and Flavobacteriales were also enriched in addition to Vibrionales, the SAR 11 clade, and Marinoscillum. To our knowledge, this is the first report of differentially enriched taxa between day and night in coral reef seawater, but some of these trends have been reported in other marine microbial communities. Gammaproteobacteria exhibited higher activity at night (Ruiz-González et al. 2012) in the Mediterranean, corresponding with potential dissolved organic matter release from grazing zooplankton. In our study, it is possible that enrichment of Oceanospirillales also indicates enhanced grazing on the reef at night. SAR11 clade bacteria were enriched during the day in reef seawater, aligning with decreases in SAR11 relative abundances at night in the English Channel (Gilbert et al. 2010) and up-regulated gene transcription in SAR11 during the day in the North Pacific Subtropical Gyre (Ottesen et al. 2014). These trends are likely partly explained by the dependence of SAR11 on sunlight for driving proteorhodopsin activity (Giovannoni et al. 2005; Lami and Kirchman 2014). *Heterogeneity in bacterial and archaeal diversity and composition across the reef*

The observed richness of bacterial and archaeal ASVs was higher in seawater collected from reef-depth and coral ecosphere seawater relative to surface seawater during both day and night, demonstrating spatial heterogeneity in the microbial community over the water column and elevated microbial diversity at depth. Enhanced diversity closer to the reef may reflect a wider variety of microbial niches, increased nutrient availability, as well as less photoinhibition compared to surface seawater. Taxa enriched in reef-depth and coral ecosphere seawater compared to surface seawater bacterial and archaeal communities included *Prochlorococcus*, SAR11 (clades 1a and 1b), the NS4 marine group, and "*Candidatus* Actinomarina," taxa that are associated with oligotrophic environments and that are commonly detected in reef seawater.

Prochlorococcus ASVs were identified as *Prochlorococcus marinus* (MIT9313), a low-light adapted *Prochlorococcus* ecotype. We suspect that there are more *Prochlorococcus* ecotypes in reef seawater, but did not have the resolution to capture this diversity by comparing differences in the v4 region of the 16S rRNA gene because ecotypes can differ by <1% (reviewed by Coleman et al. 2006). Compared to other *Prochlorococcus* ecotypes, MIT9313 is usually found at the base of the euphotic zone and has the genetic capability of using and reducing nitrite as a source of nitrogen (Rocap et al. 2003; Tolonen et al. 2006). *Prochlorococcus* could be advecting onto the reef from offshore currents, but if this were the case, we would expect *Prochlorococcus* to be evenly distributed across the shallow water column. We consistently detected this ecotype at 7 m depth as opposed to 100 m in the open ocean, demonstrating that reef-depth seawater has unique attributes that may select for the growth of this ecotype or that surface reef seawater is less hospitable to *Prochlorococcus* because of photoinhibition (Soitamo et al. 2017).

Like *Prochlorococcus*, heterotrophic bacteria within the SAR11 clade are abundant in oligotrophic marine environments and coral reef seawater (Morris et al. 2002; Tout et al. 2014). In fact, coral reef exudates from *P. astreoides* stimulate SAR11 growth rates and there is evidence that *P. astreoides* also grazes on these cells in mesocosm experiments (McNally et al. 2017). Enrichment of SAR11 within reef-depth and coral ecosphere seawater suggests that substrates that accumulate within the reef benthic boundary layer could contribute to the presence or growth of SAR11 in reef seawater.

NS4 marine group bacteria within the phylum Bacteroidetes are commonly detected in marine microbial communities (Alonso et al. 2007; Becker et al. 2017; Milici et al. 2017) and exhibit seasonality in some environments (Alonso-Saez et al. 2015). Bacteria within the "*Ca.* Actinomarina" are very small and their distribution tracks with abundances of *Synechococcus* (Ghai et al. 2013). "*Ca.* Actinomarina" have also been identified in coral reef seawater and coral mucus previously (Apprill et al. 2016). We have shown that both these groups are enriched in reef-depth seawater, expanding our knowledge about the biogeography of these taxa and their potential association with the reef.

Coral ecosphere and reef-depth seawater bacterial and archaeal communities had similar compositions and observed ASV richness even though sample type (reef-depth or coral ecosphere) influenced community similarity as revealed by the Adonis test. We expected to observe taxonomic differences between these two seawater environments as well as enrichment of Gammaproteobacteria within the coral ecosphere seawater based on previous observations (Tout et al. 2014; Weber et al. 2019), but did not observe these trends. This could be for several reasons including that *P. astreoides* does not influence coral ecosphere bacterial and archaeal communities to the same degree that other coral species do. For example, *P. astreoides'*

ecosphere bacterial and archaeal communities sampled in Cuba were more similar to reef-depth

seawater microbial communities compared to ecosphere communities sampled surrounding other

Caribbean coral species including *P. astreoides, Orbicella faveolata, Montastrea cavernosa,*

Pseudodiploria strigosa, and *Acropora cervicornis* (Weber et al. 2019). The second potential

explanation is that associations between Gammaproteobacteria and *P. astreoides* could be site-

specific and depend on local environmental conditions and/or on the health state of the coral

colony, similar to observations that have been made about common coral-associated

microorganisms (Hernandez-Agreda et al. 2018; Weber et al. 2019). For instance, the coral

colonies sampled within Cuba by Weber et al. (2019) were located on more continuous reef

structure with higher coral cover compared to the corals sampled in this study, which were

located adjacent to sand patches. Differences in coral density (potentially influencing the source,

supply, and detection of potential coral-associates or pathogens) or hydrodynamics could play a

role in the differences reported here. It is interesting to note that the tissue microbiomes of *P.*

astreoides were mostly dominated by *Endozoicomonas* bacteria whereas *Endozoicomonas* were

present but cryptic (average relative abundance of 0.11%) in coral ecosphere and reef-depth

seawater. These findings demonstrate that interactions between corals and planktonic

microorganisms may be nuanced and depend on factors that have been previously unexplored,

necessitating more research in this area.

4.6 Conclusions

High-resolution sampling of surface, reef-depth, and coral ecosphere seawater microbial

communities revealed several novel observations of microbial dynamics in the seawater on

tropical coral reefs. Even though reef seawater bacterial and archaeal communities were highly

similar over three days, there were consistent increases in *Prochlorococcus* and *Synechococcus* cell abundances at night and changes over time that corresponded with changes in temperature and increases in nitrite and silicate. We also identified diel patterns in bacterial and archaeal diversity, enrichment of different taxa by day and night, and specific taxa exhibiting rhythmic population fluctuations in reef-depth and coral ecosphere seawater. Bacterial and archaeal alpha diversity was higher in reef-depth seawater compared to surface seawater, suggesting that there may be enhanced microbial niches close to the reef, a hypothesis held widely, but rarely observed. Coral ecosphere and reef-depth seawater bacterial and archaeal communities had similar compositions, suggesting that colony or site-specific conditions may influence the outcome of coral-microbial interactions within the coral ecosphere. Comparatively, temporal changes superseded spatial differences in terms of influence on the seawater microbial community, emphasizing the need for a more comprehensive understanding of how these communities change over short term (tidal cycle to days to weeks) and longer term (seasonal to annual) time-scales. Overall, these findings demonstrate the small-scale population dynamics that take place over a diel cycle and the relative influence of temporal compared to spatial changes on microbial communities sampled across one reef.

At the scale of the reef, understanding variability in microbial composition as well as dominant forcings on these communities is essential for describing baseline temporal and spatial dynamics in productive, diverse, and sensitive coral reef ecosystems. As coral reefs continue to decline, these changes may not only impact coral reef health and the composition of microbial communities on reefs, but also the variability of microbial population fluctuations.

4.7 Acknowledgements

4.8 References

Agawin, N.S.R., and Agusti, S. (2005) *Prochlorococcus* and *Synechococcus* cells in the Central Atlantic Ocean: distribution, growth and mortality (grazing) rates. *Vie Et Milieu* **55**: 165-175.

Alonso, C., Warnecke, F., Amann, R., and Pernthaler, J. (2007) High local and global diversity of Flavobacteria in marine plankton. *Environ Microbiol* **9**: 1253-1266.

Alonso-Saez, L., Diaz-Perez, L., and Moran, X.A. (2015) The hidden seasonality of the rare biosphere in coastal marine bacterioplankton. *Environ Microbiol* **17**: 3766-3780.

Anthony, K.R.N., and Hoegh-Guldberg, O. (2003) Variation in coral photosynthesis, respiration and growth characteristics in contrasting light microhabitats: an analogue to plants in forest gaps and understoreys? *Functional Ecology* **17**: 246-259.

Apprill, A., and Rappé, M.S. (2011) Response of the microbial community to coral spawning in lagoon and reef flat environments of Hawaii, USA. *Aquat Microb Ecol* **62**: 251-266.

Apprill, A., Weber, L.G., and Santoro, A.E. (2016) Distinguishing between microbial habitats unravels ecological complexity in coral microbiomes. *mSystems* **1**.

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.

Atkinson, M.J. (2010) Biogeochemistry of nutrients. In *Coral reefs an ecosystem in transition.* Dubinsky, Z., and Stambler, N. (eds): Springer Science & Business Media, pp. 199-206.

Azam, F., and Malfatti, F. (2007) Microbial structuring of marine ecosystems. *Nat Rev Microbiol* **5**: 782-791.

Becker, C., Hughen, K., Mincer, T.J., Ossolinski, J., Weber, L., and Apprill, A. (2017) Impact of prawn farming effluent on coral reef water nutrients and microorganisms. *Aquaculture Environment Interactions* **9**: 331-346.

Benjamini, Y., and Hochberg, Y. (2000) On the adaptive control of the false discovery rate in multiple testing with independent statistics. *Journal of Educational and Behavioral Statistics* **25**: 60-83.

Binder, B.J., and DuRand, M.D. (2002) Diel cycles in surface waters of the equatorial Pacific. *Deep-Sea Res II* **49**: 2601-2617.

Brocke, H.J., Wenzhoefer, F., de Beer, D., Mueller, B., van Duyl, F.C., and Nugues, M.M. (2015) High dissolved organic carbon release by benthic cyanobacterial mats in a Caribbean reef ecosystem. *Sci Rep* **5**: 8852.

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

Charpy, L., Casareto, B.E., Langlade, M.J., and Suzuki, Y. (2012) Cyanobacteria in Coral Reef Ecosystems: A Review. *Journal of Marine Biology* **2012**: 1-9.

Cohen, S.E., and Golden, S.S. (2015) Circadian Rhythms in Cyanobacteria. *Microbiol Mol Biol Rev* **79**: 373-385.

Coleman, M., Sullivan, M.B., Martiny, A.C., Steglich, C., Barry, K., DeLong, E.F., and Chisholm, S.W. (2006) Genomic islands and the ecology and evolution of *Prochlorococcus*. *Science* **311**: 1768-1770.

Cox, E.F., Ribes, M., and Kinzie III, R.A. (2006) Temporal and spatial scaling of planktonic responses to nutrient inputs into a subtropical embayment. *Mar Ecol Prog Ser* **324**: 19-35.

D'Angelo, C., and Wiedenmann, J. (2014) Impacts of nutrient enrichment on coral reefs: new perspectives and implications for coastal management and reef survival. *Current Opinion in Environmental Sustainability* **7**: 82-93.

DeCarlo, T.M., Cohen, A.L., Wong, G.T.F., Shiah, F.-K., Lentz, S.J., Davis, K.A. *et al.* (2017) Community production modulates coral reef pH and the sensitivity of ecosystem calcification to ocean acidification. *Journal of Geophysical Research: Oceans* **122**: 745-761.

Dinsdale, E.A., Pantos, O., Smriga, S., Edwards, R.A., Angly, F., Wegley, L. *et al.* (2008) Microbial ecology of four coral atolls in the Northern Line Islands. *PLoS One* **3**: e1584.

DuRand, M.D., Olson, R.J., and Chisholm, S.W. (2001) Phytoplankton population dynamics at the Bermuda Atlantic Time-series station in the Sargasso Sea. *Deep-Sea Res Pt II* **48**: 1983-2003.

Falkowski, P.G., Dubinsky, Z., Muscatine, L., and Porter, J.W. (1984) Light and the bioenergetics of a symbiotic coral. *BioScience* **34**: 705 - 709.

Gardner, T.A., Cote, I.M., Gill, J.A., Grant, A., and Watkinson, A.R. (2003) Long - term region wide declines in Caribbean corals. *Science* **301**: 958-960.

Ghai, R., Mizuno, C.M., Picazo, A., Camacho, A., and Rodriguez-Valera, F. (2013) Metagenomics uncovers a new group of low GC and ultra-small marine Actinobacteria. *Sci Rep* **3**: 2471.

Gilbert, J.A., Field, D., Swift, P., Thomas, S., Cummings, D., Temperton, B. *et al.* (2010) The taxonomic and functional diversity of microbes at a temperate coastal site: a 'multi-omic' study of seasonal and diel temporal variation. *PLoS One* **5**: e15545.

Giovannoni, S.J., Bibbs, L., Cho, J.C., Stapels, M.D., Desiderio, R., Vergin, K.L. *et al.* (2005) Proteorhodopsin in the ubiquitous marine bacterium SAR11. *Nature* **438**: 82-85.

Guadayol, O., Silbiger, N.J., Donahue, M.J., and Thomas, F.I. (2014) Patterns in temporal variability of temperature, oxygen and pH along an environmental gradient in a coral reef. *PLoS One* **9**: e85213.

Haas, A.F., Nelson, C.E., Wegley Kelly, L., Carlson, C.A., Rohwer, F., Leichter, J.J. *et al.* (2011) Effects of coral reef benthic primary producers on dissolved organic carbon and microbial activity. *PLoS One* **6**: e27973.

Haas, A.F., Nelson, C.E., Rohwer, F., Wegley-Kelly, L., Quistad, S.D., Carlson, C.A. *et al.* (2013) Influence of coral and algal exudates on microbially mediated reef metabolism. *PeerJ* **1**: e108.

Haas, A.F., Fairoz, M.F., Kelly, L.W., Nelson, C.E., Dinsdale, E.A., Edwards, R.A. *et al.* (2016) Global microbialization of coral reefs. *Nat Microbiol* **1**: 16042.

Hernandez-Agreda, A., Leggat, W., Bongaerts, P., Herrera, C., and Ainsworth, T.D. (2018) Rethinking the coral microbiome: simplicity exists within a diverse microbial biosphere. *mBio* **9**: e00812-00818.

Houlbreque, F., and Ferrier-Pages, C. (2009) Heterotrophy in tropical scleractinian corals. *Biol Rev Camb Philos Soc* **84**: 1-17.

Hu, S.K., Connell, P.E., Mesrop, L.Y., and Caron, D.A. (2018) A Hard Day's Night: Diel Shifts in Microbial Eukaryotic Activity in the North Pacific Subtropical Gyre. *Front Mar Sci* **5**.

Jacquet, S., Partensky, F., Marie, D., Casotti, R., and Vaulot, D. (2001) Cell cycle regulation by light in Prochlorococcus strains. *Appl Environ Microbiol* **67**: 782-790.

Kelly, L.W., Williams, G.J., Barott, K.L., Carlson, C.A., Dinsdale, E.A., Edwards, R.A. *et al.* (2014) Local genomic adaptation of coral reef-associated microbiomes to gradients of natural variability and anthropogenic stressors. *Proc Natl Acad Sci U S A* **111**: 10227-10232.

Kelly, L.W., Nelson, C.E., Haas, A.F., Naliboff, D.S., Calhoun, S., Carlson, C.A. *et al.* (2019) Diel population and functional synchrony of microbial communities on coral reefs. *Nat Commun* **10**: 1691.

Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K., and Schloss, P.D. (2013) Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol* **79**: 5112-5120.

Kurihara, H., Ikeda, N., and Umezawa, Y. (2018) Diurnal and seasonal variation of particle and dissolved organic matter release by the coral Acropora tenuis. *PeerJ* **6**: e5728.

Lami, R., and Kirchman, D.L. (2014) Diurnal expression of SAR11 proteorhodopsin and 16S rRNA genes in coastal North Atlantic waters. *Aquat Microb Ecol* **73**: 185-194.

Love, M.I., Huber, W., and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**: 550.

Marie, D., Partensky, F., Jacquet, S., and Vaulot, D. (1997) Enumeration and cell cycle analysis of natural populations of marine picoplankton by flow cytometry using the nucleic acid stain SYBR Green I. *Appl Environ Microb* **63**: 186-193.

McMurdie, P.J., and Holmes, S. (2013) phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* **8**: e61217.

McNally, S.P., Parsons, R.J., Santoro, A.E., and Apprill, A. (2017) Multifaceted impacts of the stony coral Porites astreoides on picoplankton abundance and community composition. *Limnol Oceanogr* **62**: 217-234.

Milici, M., Vital, M., Tomasch, J., Badewien, T.H., Giebel, H.-A., Plumeier, I. *et al.* (2017) Diversity and community composition of particle-associated and free-living bacteria in mesopelagic and bathypelagic Southern Ocean water masses: Evidence of dispersal limitation in the Bransfield Strait. *Limnol Oceanogr* **62**: 1080-1095.

Moriarty, D.J.W., Pollard, P.C., and Hunt, W.G. (1985) Temporal and spatial variation in bacterial production in the water column over a coral reef. *Marine Biology* **85**: 285-292.

Morris, R.M., Rappe, M.S., Connon, S.A., Vergin, K.L., Siebold, W.A., Carlson, C.A., and Giovannoni, S.J. (2002) SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* **420**: 806-809.

Nakajima, R., Tanaka, Y., Guillemette, R., and Kurihara, H. (2017) Effects of coral-derived organic matter on the growth of bacterioplankton and heterotrophic nanoflagellates. *Coral Reefs* **36**: 1171-1179.

Nelson, C.E., Alldredge, A.L., McCliment, E.A., Amaral-Zettler, L.A., and Carlson, C.A. (2011) Depleted dissolved organic carbon and distinct bacterial communities in the water column of a rapid-flushing coral reef ecosystem. *ISME J* **5**: 1374-1387.

Ochsenkuhn, M.A., Schmitt-Kopplin, P., Harir, M., and Amin, S.A. (2018) Coral metabolite gradients affect microbial community structures and act as a disease cue. *Commun Biol* **1**: 184.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D. *et al.* (2017) vegan: community ecology package. R package version 2.4-5.

Ottesen, E.A., Young, C.R., Gifford, S.M., Eppley, J.M., Marin III, R., Schuster, S.C. *et al.* (2014) Multispecies diel transcriptional oscillations in open ocean heterotrophic bacterial assemblages. *Science* **435**: 207 - 212.

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.

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

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

Ribalet, F., Swalwell, J., Clayton, S., Jimenez, V., Sudek, S., Lin, Y. *et al.* (2015) Light-driven synchrony of *Prochlorococcus* growth and mortality in the subtropical Pacific gyre. *Proc Natl Acad Sci U S A* **112**: 8008-8012.

Rocap, G., Larimer, F.W., Lamerdin, J., Malfatti, S., Chain, P., Ahlgren, N.A. *et al.* (2003) Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* **424**: 1042-1047.

Ruiz-González, C., Lefort, T., Massana, R., Simó, R., and Gasol, J.M. (2012) Diel changes in bulk and single-cell bacterial heterotrophic activity in winter surface waters of the northwestern Mediterranean Sea. *Limnol Oceanogr* **57**: 29-42.

Santoro, A.E., Casciotti, K.L., and Francis, C.A. (2010) Activity, abundance and diversity of nitrifying archaea and bacteria in the central California Current. *Environ Microbiol* **12**: 1989-2006.

Seymour, J.R., Patten, N., Bourne, D.G., and Mitchell, J.G. (2005) Spatial dynamics of virus-like particles and heterotrophic bacteria within a shallow coral reef ecosystem. *Mar Ecol Prog Ser* **288**: 1-8.

Shashar, N., Cohen, Y., and Loya, Y. (1993) Extreme diel fluctuations of oxygen in diffusive boundary layers surrounding stony corals. *Biological Bulletin* **185**: 455 - 461.

Shashar, N., Kinane, S., Jokiel, P.L., and Patterson, M.R. (1996) Hydromechanical boundary layers over a coral reef. *J Exp Mar Biol Ecol* **199**: 17-28.

Silbiger, N.J., Nelson, C.E., Remple, K., Sevilla, J.K., Quinlan, Z.A., Putnam, H.M. *et al.* (2018) Nutrient pollution disrupts key ecosystem functions on coral reefs. *Proc Biol Sci* **285**.

Silveira, C.B., Gregoracci, G.B., Coutinho, F.H., Silva, G.G.Z., Haggerty, J.M., de Oliveira, L.S. *et al.* (2017) Bacterial community associated with the reef coral *Mussismilia braziliensis's* momentum boundary layer over a diel cycle. *Front Microbiol* **8**: 784.

Smith, J.E., Price, N.N., Nelson, C.E., and Haas, A.F. (2013) Coupled changes in oxygen concentration and pH caused by metabolism of benthic coral reef organisms. *Marine Biology* **160**: 2437-2447.

Soitamo, A., Havurinne, V., and Tyystjarvi, E. (2017) Photoinhibition in marine picocyanobacteria. *Physiol Plant* **161**: 97-108.

Sorokin, Y.I. (1973) Trophical role of bacteria in the ecosystem of the coral reef. *Nature* **242**: 415-418.

Sweet, M.J., Croquer, A., and Bythell, J.C. (2010) Temporal and spatial patterns in waterborne bacterial communities of an island reef system. *Aquat Microb Ecol* **61**: 1-11.

Tanaka, Y., Miyajima, T., Umezawa, Y., Hayashibara, T., Ogawa, H., and Koike, I. (2009) Net release of dissolved organic matter by the scleractinian coral *Acropora pulchra*. *J Exp Mar Biol Ecol* **377**: 101-106.

Thaben, P.F., and Westermark, P.O. (2014) Detecting rhythms in time series with RAIN. *J Biol Rhythms* **29**: 391-400.

Thornton, D.C.O. (2014) Dissolved organic matter (DOM) release by phytoplankton in the contemporary and future ocean. *European Journal of Phycology* **49**: 20-46.

Tolonen, A.C., Aach, J., Lindell, D., Johnson, Z.I., Rector, T., Steen, R. *et al.* (2006) Global gene expression of Prochlorococcus ecotypes in response to changes in nitrogen availability. *Mol Syst Biol* **2**: 53.

Tout, J., Jeffries, T.C., Webster, N.S., Stocker, R., Ralph, P.J., and Seymour, J.R. (2014) Variability in microbial community composition and function between different niches within a coral reef. *Microb Ecol* **67**: 540-552.

Tremblay, P., Grover, R., Maguer, J.F., Legendre, L., and Ferrier-Pages, C. (2012) Autotrophic carbon budget in coral tissue: a new 13C-based model of photosynthate translocation. *J Exp Biol* **215**: 1384-1393.

Valdivia, A., Cox, C.E., and Bruno, J.F. (2017) Predatory fish depletion and recovery potential on Caribbean reefs. *Science Advances* **3**: e1601303.

Vaulot, D., Marie, D., Olson, R.J., and Chisholm, S.W. (1995) Growth of Prochlorococcus, a photosynthetic prokaryote, in the Equatorial Pacific Ocean. *Science* **268**: 1480-1482.

Weber, L., DeForce, E., and Apprill, A. (2017) Optimization of DNA extraction for advancing coral microbiota investigations. *Microbiome* **5**.

Weber, L., Gonzalez-Díaz, P., Armenteros, M., and Apprill, A. (2019) The coral ecosphere: A unique coral reef habitat that fosters coral-microbial interactions. *Limnol Oceanogr*.

Werner, U., Blazejak, A., Bird, P., Eickert, G., Schoon, R., Abed, R.M.M. *et al.* (2008) Microbial photosynthesis in coral reef sediments (Heron Reef, Australia). *Estuarine, Coastal and Shelf Science* **76**: 876-888.

Werner, U., Bird, P., Wild, C., Ferdelman, T., Polerecky, L., Eickert, G. *et al.* (2006) Spatial patterns of aerobic and anaerobic mineralization rates and oxygen penetration dynamics in coral reef sediments. *Mar Ecol Prog Ser* **309**: 93-105.

Wickham, H. (2016) ggplot2: elegant graphics for data analysis. New York: Springer-Verlag.

Wild, C., Rasheed, M., Werner, U., Franke, U., Johnstone, R., and Huettel, M. (2004) Degradation and mineralization of coral mucus in reef environments. *Mar Ecol Prog Ser* **267**: 159-171.

Yahel, R., Yahel, G., Jaffe, J.S., and Genin, A. (2005) Diel pattern with abrupt crepuscular changes of zooplankton over a coral reef. *Limnol Oceanogr* **50**: 930-944.

4.9 Tables

Table 1. Summary table of statistical tests.

Parameter	Contrasts	Global test results	Post – hoc tests
Prochlorococcus abundances	Time*	Kruskal-Wallis rank sum test $Chi^2 = 57.70$, df = 5, p = 0	Dunn's test with Bonferroni corrections; adjusted p-value < 0.05
	Distance from coral	Kruskal-Wallis rank sum test $Chi^2 = 0.12$, df = 1, p = 0.73	-
	Colony	Kruskal-Wallis rank sum test $Chi^2 = 1.1728$, df = 5, p = 0.9475	-
Synechococcus abundances	Time*	Kruskal-Wallis rank sum test $Chi^2 = 54.90$, df = 5, p = 0	Dunn's test with Bonferroni corrections; adjusted p-value < 0.05
	Distance from coral	Kruskal-Wallis rank sum test $Chi^2 = 0.16$, df = 1, p = 0.69	-
	Colony	Kruskal-Wallis rank sum test $Chi^2 = 2.9031$, df = 5, p = 0.71	-
Picoeukaryote abundances	Time*	Kruskal-Wallis rank sum test $Chi^2 = 56.21$, df = 5, p = 0	Dunn's test with Bonferroni corrections; adjusted p-value < 0.05
	Distance from coral	Kruskal-Wallis rank sum test $Chi^2 = 0.82$, df = 1, p = 0.37	-
	Colony	Kruskal-Wallis rank sum test $Chi^2 = 3.2404$, df = 5, p = 0.663	-
Unpigmented cell abundances	Time*	Kruskal-Wallis rank sum test $Chi^2 = 24.65$, df = 5, p = 0	Dunn's test with Bonferroni corrections; adjusted p-value < 0.05
	Distance from coral	Kruskal-Wallis rank sum test $Chi^2 = 3.51$, df = 1, p = 0.06	-
	Colony	Kruskal-Wallis rank sum test $Chi^2 = 8.392$, df = 5, p = 0.1359	-
Phosphate concentrations	Time*	Kruskal-Wallis rank sum test $Chi^2 = 34.79$, df = 5, p = 0	Dunn's test with Bonferroni corrections; adjusted p-value < 0.05
	Distance from coral	Kruskal-Wallis rank sum test $Chi^2 = 1.22$, df = 1, p = 0.27	-
	Colony	Kruskal-Wallis rank sum test $Chi^2 = 2.3406$, df = 5, p = 0.8003	-
Silicate concentrations	Time*	Kruskal-Wallis rank sum test $Chi^2 = 45.67$, df = 5, p = 0	Dunn's test with Bonferroni corrections; adjusted p-value < 0.05
	Distance from coral	Kruskal-Wallis rank sum test $Chi^2 = 0.47$, df = 1, p = 0.49	-
	Colony	Kruskal-Wallis rank sum test $Chi^2 = 1.7091$, df = 5, p = 0.89	-
Ammonium concentrations	Time*	Kruskal-Wallis rank sum test $Chi^2 = 22.97$, df = 5, p = 0	Dunn's test with Bonferroni corrections; adjusted p-value < 0.05
	Distance from coral *	Kruskal-Wallis rank sum test $Chi^2 = 6.65$, df = 1, p = 0.01	Dunn's test with Bonferroni corrections; adjusted p-value < 0.05
	Colony	Kruskal-Wallis rank sum test $Chi^2 = 5.4575$, df = 5, p = 0.3626	-
Nitrate concentrations	Time*	ANOVA $F(5,62) = 7.756$, p = 1.02e-05	Tukey multiple comparisons of means; z-statistic = -2.578529 adjusted p-value = 0.005
	Distance from coral	ANOVA $F(1,66) = 0.4073$, p = 0.52	-
	Colony	ANOVA $F(5,62) = 0.9202$, p = 0.4740	-
	Time	ANOVA	-

Nitrite concentrations		$F_{(5,62)} = 1.1664$, p = 0.3359	
	Distance from coral	ANOVA $F_{(1,66)} = 0.0073$, p = 0.9322	-
	Colony	ANOVA $F_{(5,62)} = 1.143$, p = 0.3475	-
Observed Bacterial and Archaeal richness	Time and distance from coral*	Kruskal-Wallis rank sum test $Chi^2 = 20.002$, df = 3, p = 0.0001696	Dunn's test with Bonferroni corrections; adjusted p-value < 0.05

Table 2. Results of Permanova (ADONIS) test examining factors influencing amplicon-based reef-depth and coral ecosphere seawater microbial community dissimilarity, using 999 permutations.

Factor	DF[s]	Sums of Squares	MeanSqs[s]	F model	R[2]	Pr(>F)[p]
Time (Day or Night)	1	0.11460	0.114598	11.2979	0.07275	0.000999
Colony	5	0.16125	0.032251	3.1795	0.10236	0.001998
Time-series*	4	0.59465	0.148663	14.6563	0.37749	0.000999
Distance&	1	0.04023	0.040227	3.9658	0.02554	0.004995
Time: Colony	4	0.06572	0.016430	1.6198	0.04172	0.095904
Colony: Timeseries	17	0.32497	0.019116	1.8846	0.20629	0.003996
Residuals	27	0.27387	0.010143		0.17385	
Total	59	1.57529			1.00000	

* The six timepoints were sampled at 3 pm and 3 am over three days

& Reef-depth (2 m) or coral ecosphere (5 cm) sampling distances from the coral colony

[s]DF= degrees of freedom

[s]MeanSqs = mean squares

[p]Pr(>F) = permutational p-values using pseudo-F ratios

Table 3. Amplicon sequence variants (ASVs) that displayed rhythmic fluctuations in relative abundance over a period of 24 hours in reef-depth and coral ecosphere seawater.

Taxonomy	ASV #	Environment	Phase	p-value
Prochlorococcus marinus (MIT9313)	ASV2	reef-depth, ecosphere	24	2.62E-08, 2.49E-07
Prochlorococcus marinus (MIT9313)	ASV6	reef-depth, ecosphere	24	8.78E-07, 5.71E-05
Synechococcus CC9902	ASV43	reef-depth, ecosphere	12	9.17E-08, 1.14E-04
Prochlorococcus marinus (MIT9313)	ASV49	reef-depth, ecosphere	24	2.49E-07, 3.57E-06
Cyanobium PCC-6307	ASV61	reef-depth, ecosphere	12	8.78E-07, 1.27E-06
Prochlorococcus marinus (MIT9313)	ASV80	reef-depth	24	1.82E-06
Prochlorococcus marinus (MIT9313)	ASV128	reef-depth	24	2.70E-05
Synechococcus CC9902	ASV139	reef-depth	12	2.00E-03
Planctomycetacia, Pirellulaceae	ASV149	reef-depth	12	4.00E-04
Synechococcus CC9902	ASV159	reef-depth, ecosphere	12	8.97E-06, 5.87E-04

4.10 Figures
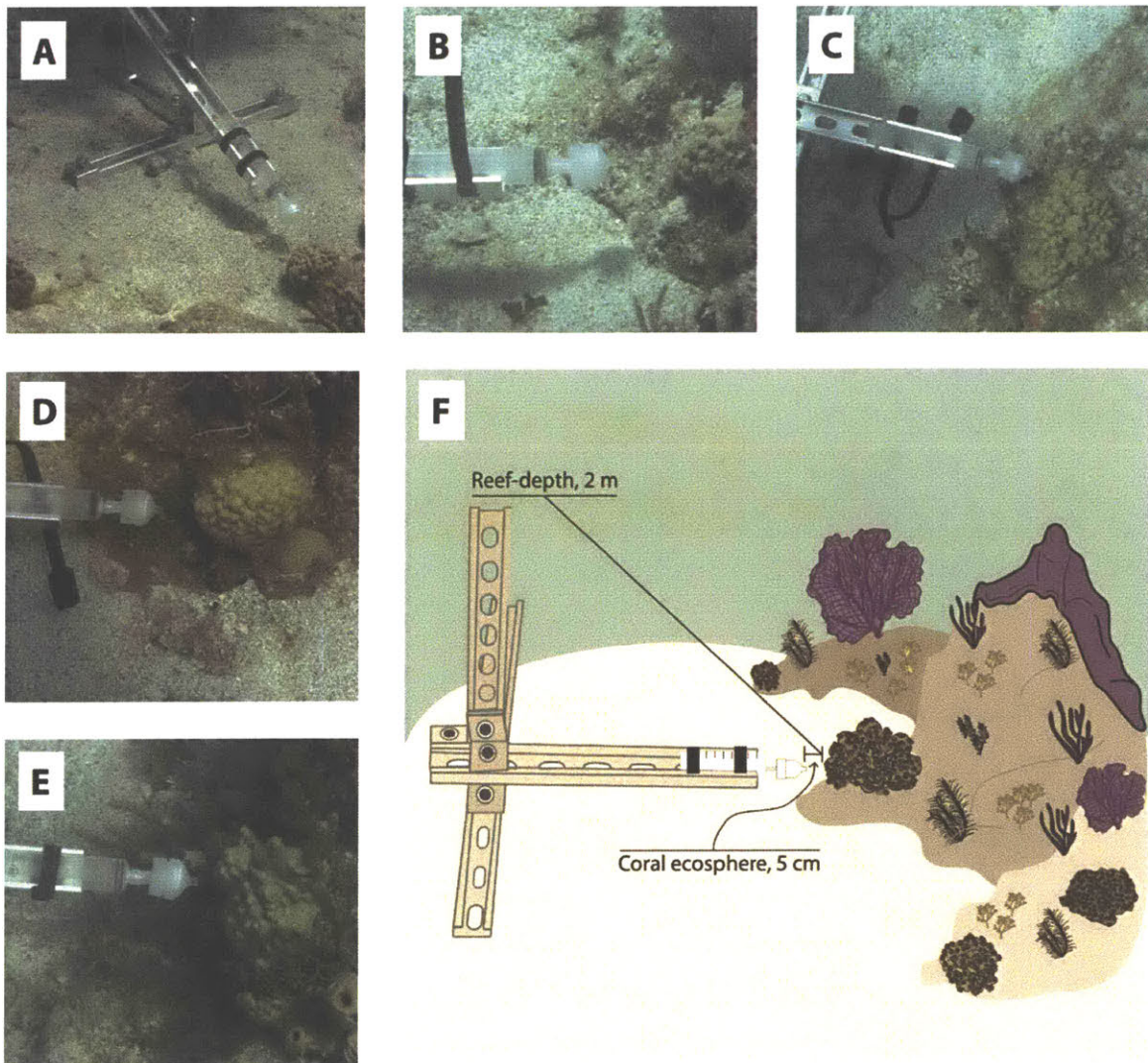
Figure 1. Photographs of the selected *Porites astreoides* coral colonies located adjacent to deployed coral ecosphere sampling devices (CESD) (A-E) and a sampling diagram detailing seawater sampling locations from coral ecosphere and reef-depths (F). Diagram is not drawn to scale.
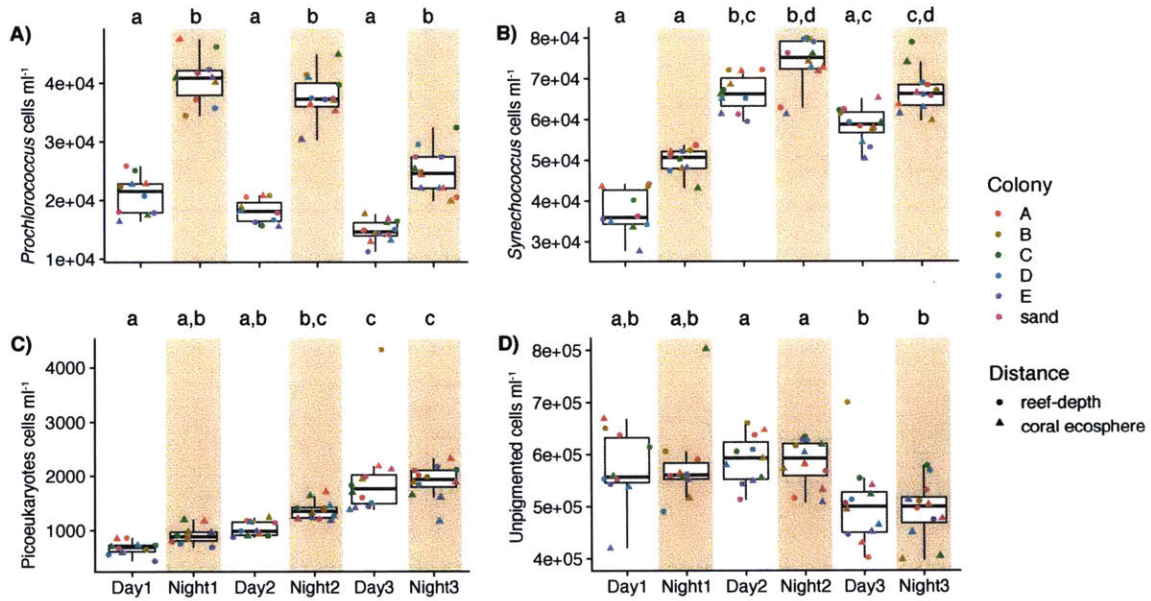
Figure 2. Comparison of picoplankton cell abundance over three consecutive days for A) *Prochlorococcus*, B) *Synechococcus*, C) picoeukaryotes, and D) unpigmented cells (heterotrophic bacteria and archaea). Each point represents a sample. Point shape corresponds to sampling distance from the coral and point color reflects the colony adjacent to where sampling was conducted. Gray shading indicates samples collected at night. Lower and upper edges of the boxplot correspond to the first and third quartiles, the whiskers extend to the largest or smallest value at 1.5 times the interquartile, and the black bar across the box represents the median. Boxplots with different letters denote significantly different cell abundances (Kruskal-Wallis Rank sum test, Dunn's test with Bonferroni corrections, p < 0.05).
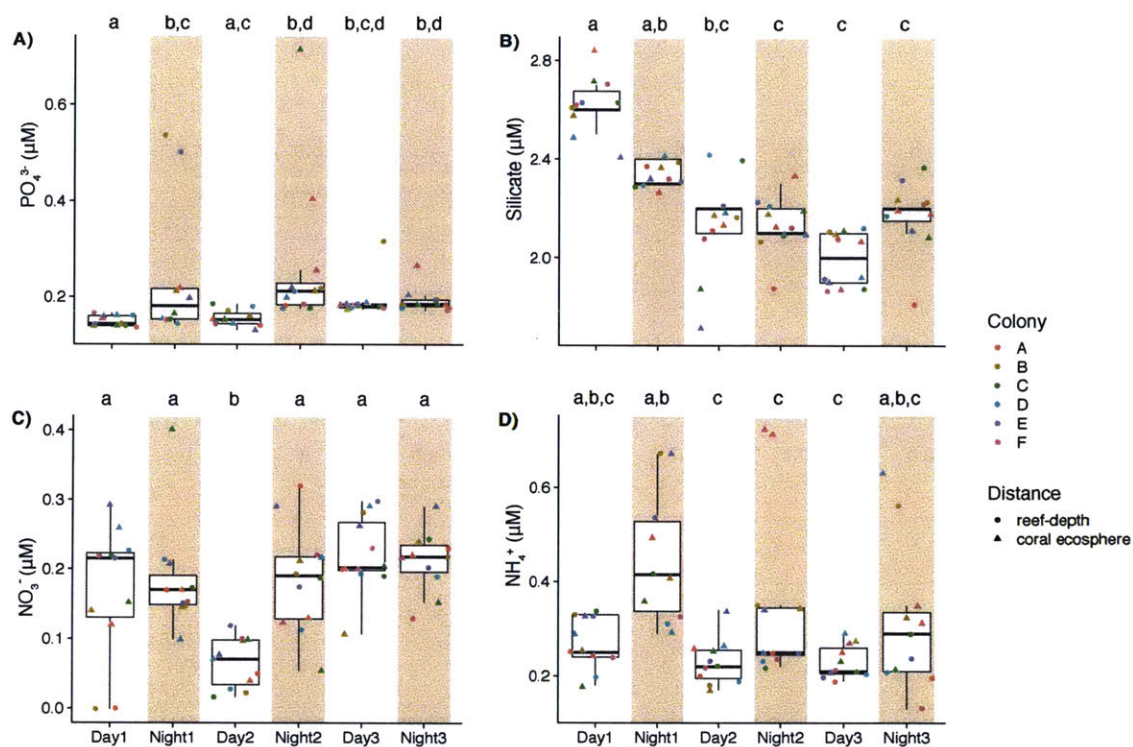
Figure 3. Comparison of inorganic nutrient concentrations over three days for A) phosphate (PO_4^{3-}), B) silicate, C) nitrate (NO_3^-), and D) ammonium (NH_4^+). Each point represents a sample. Point shape corresponds to sampling distance from the coral and point color reflects the colony adjacent to where sampling was conducted. Gray shading indicates samples collected at night. Lower and upper edges of the boxplot correspond to the first and third quartiles, the whiskers extend to the largest or smallest value at 1.5 times the interquartile, and the black bar across the box represents the median. Boxplots with different letters denote significantly different concentrations (ANOVA with Tukey's HSD test or Kruskal-Wallis Rank sum test, Dunn's test with Bonferroni corrections, $p < 0.05$).
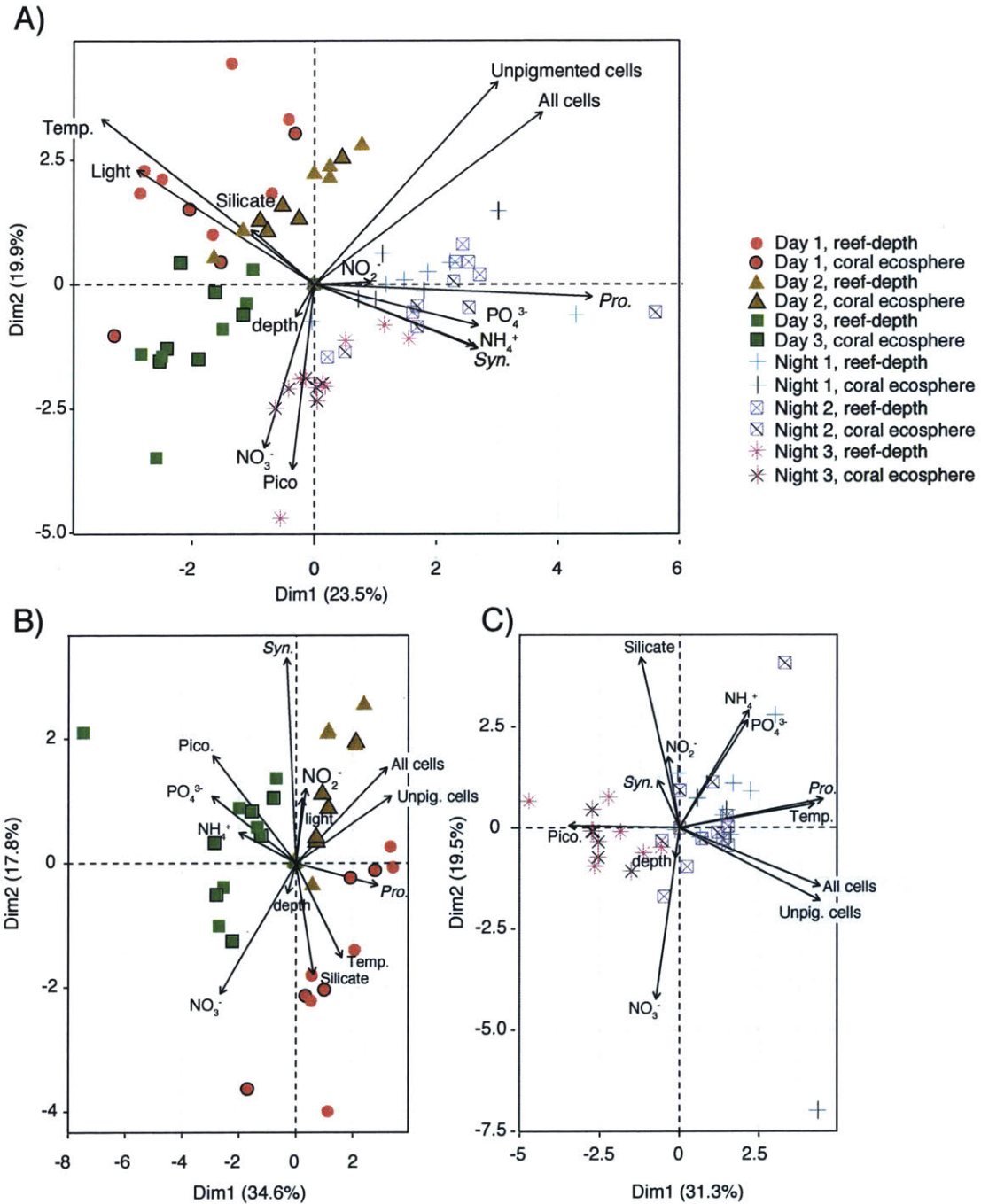
Figure 4. Principal Coordinates Analysis (PCA) biplots of picoplankton abundances, relative light, temperature, sampling depth, and inorganic macronutrient concentrations across A) all time-points, B) samples collected during the day, and C) samples collected during the night. Point color and shape reflect the day and time of sampling. Symbols outlined in black or with 1-2 black lines indicate ecosphere samples. Abbreviations are as follows: *Syn.* = *Synechococcus* cell

abundances, *Pro.* = *Prochlorococcus* cell abundances, Pico. = picoeukaryotes, Unpig. cells = unpigmented cells, and temp. = temperature.
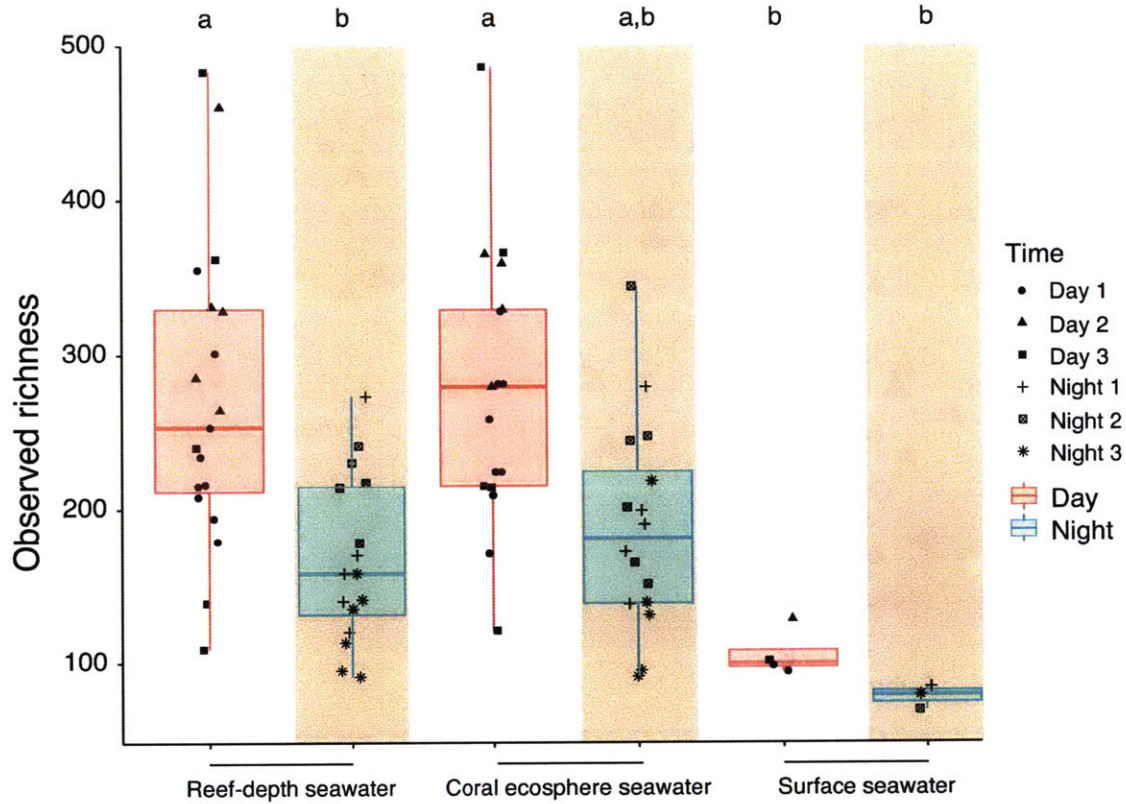
Figure 5. Observed bacterial and archaeal community richness by sample type and time, based on SSU rRNA gene sequences grouped into ASVs. Each point represents a sample. The date of sampling is denoted by the shape and the point color reflects day or night. Gray shading indicates samples collected at night. Lower and upper edges of the boxplot correspond to the first and third quartiles, the whiskers extend to the largest or smallest value at 1.5 times the interquartile, and the black bar across the box represents the median. Boxplots with different letters denote significantly different observed richness (Kruskal-Wallis Rank Sum test, Dunn's test with Bonferroni corrections, $p < 0.05$).
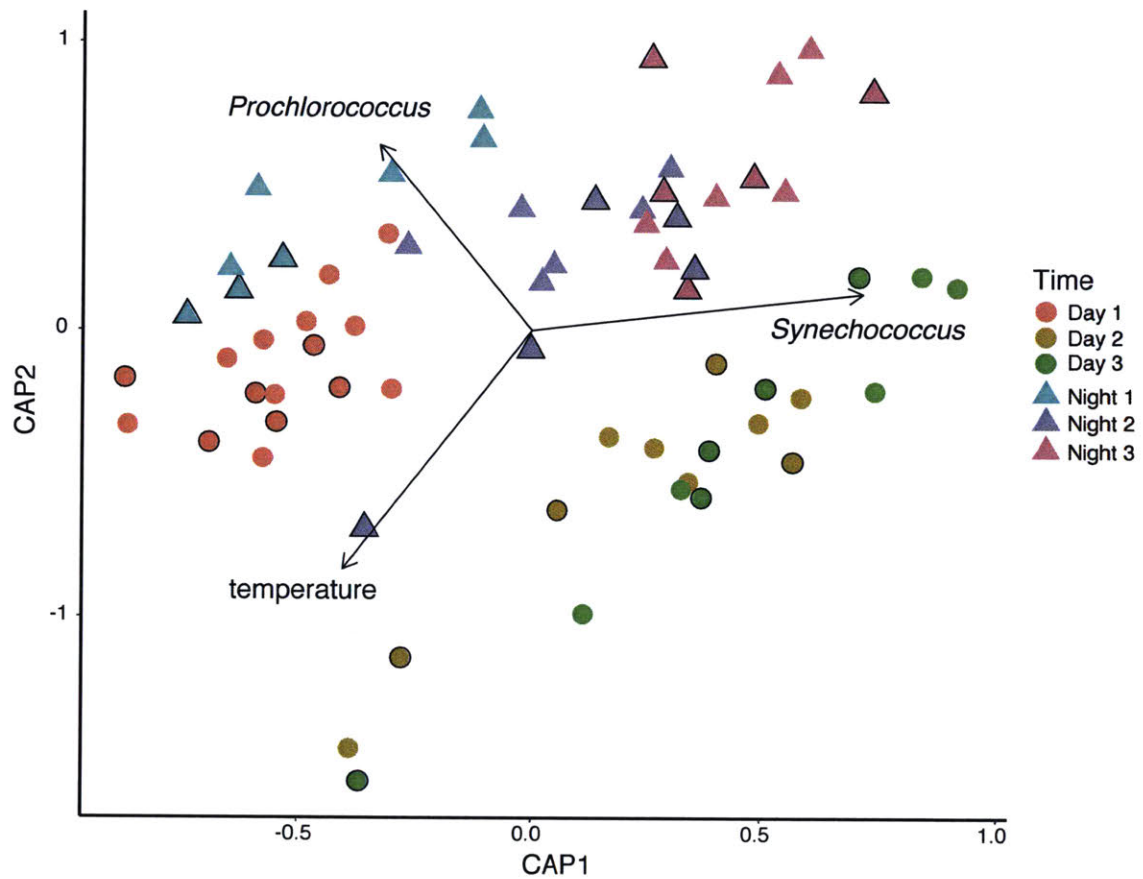
Figure 6. Constrained Analysis of Principal Coordinates (CAP) of bacterial and archaeal SSU rRNA genes sequenced from obtained from reef-depth and coral ecosphere samples compared using the Bray – Curtis index. Significant environmental variables (ANOVA, $p < 0.05$) are overlaid on the plot as vectors and labeled. Symbol color and shape indicate the time of sampling. Symbols outlined in black indicate coral ecosphere samples.

A)

Taxonomy
- Pseudomonadales, Moraxellaceae: *Acinetobacter*
- Actinobacteria: *"Candidatus* Actinomarina"
- Alphaproteobacteria, SAR 116: *"Candidatus* Puniceispirillum"
- Alphaproteobacteria, SAR 11: Clade 1a
- Alphaproteobacteria, SAR 11: Clade 1b
- Alphaproteobacteria, Rhodobacterales: HIMB11
- Bacteroidetes, Cyclobacteriaceae: *Marinoscillum*
- Flavobacteriales: NS2b marine group
- Flavobacteriales: NS4 marine group
- Flavobacteriales: NS5 marine group
- Gammaproteobacteria, Cellvibrionales: OM60(NOR5) clade
- Cyanobacteria, Synechococcales: *Prochlorococcus* MIT9313
- Gammaproteobacteria, Alteromonadales: *Pseudoalteromonas*
- Cyanobacteria, Synechococcales: *Synechococcus* CC9902
- Cyanobacteria, Nostocales: *Trichodesmium* IMS101

B)

Taxonomy
- Planctomycetes, Pirellulales: *Blastopirellula*
- Planctomycetes, Pirellulales: *Bythopirellula*
- Thaumarchaeota, Nitrosopumilales: *"Candidatus* Nitrosopumilus"
- Gammaproteobacteria, Oceanospirillales: *Endozoicomonas*
- Cyanobacteria, Nostocales: *Geitlerinema* PCC-7105
- Planctomycetes, Pirellulales: Pir4 lineage
- Planctomycetes, Pirellulales: *Rhodopirellula*
- Alphaproteobacteria, Rhodobacterales: *Ruegeria*
- Spirochaetes: Spirochaeta 2
- Firmicutes, Bacillales: *Staphylococcus*
- Acidobacteria, Thermoanaerobaculales: Subgroup_10
- Actinobacteria, Microtrichales: Sva0996 marine group
- Cyanobacteria, Synechococcales: *Synechococcus* CC9902
- Gammaproteobacteria: *Vibrio*

Figure 7. Relative abundance of bacterial and archaeal sequences from SSU rRNA gene sequencing, that comprise >1% of the genus-type level community composition in A) reef-depth (RD), coral ecosphere (Eco) and surface seawater samples (*) and B) coral tissue samples. Colors indicate the taxonomic grouping at the genus-type level. Color legends are specific for each plot.
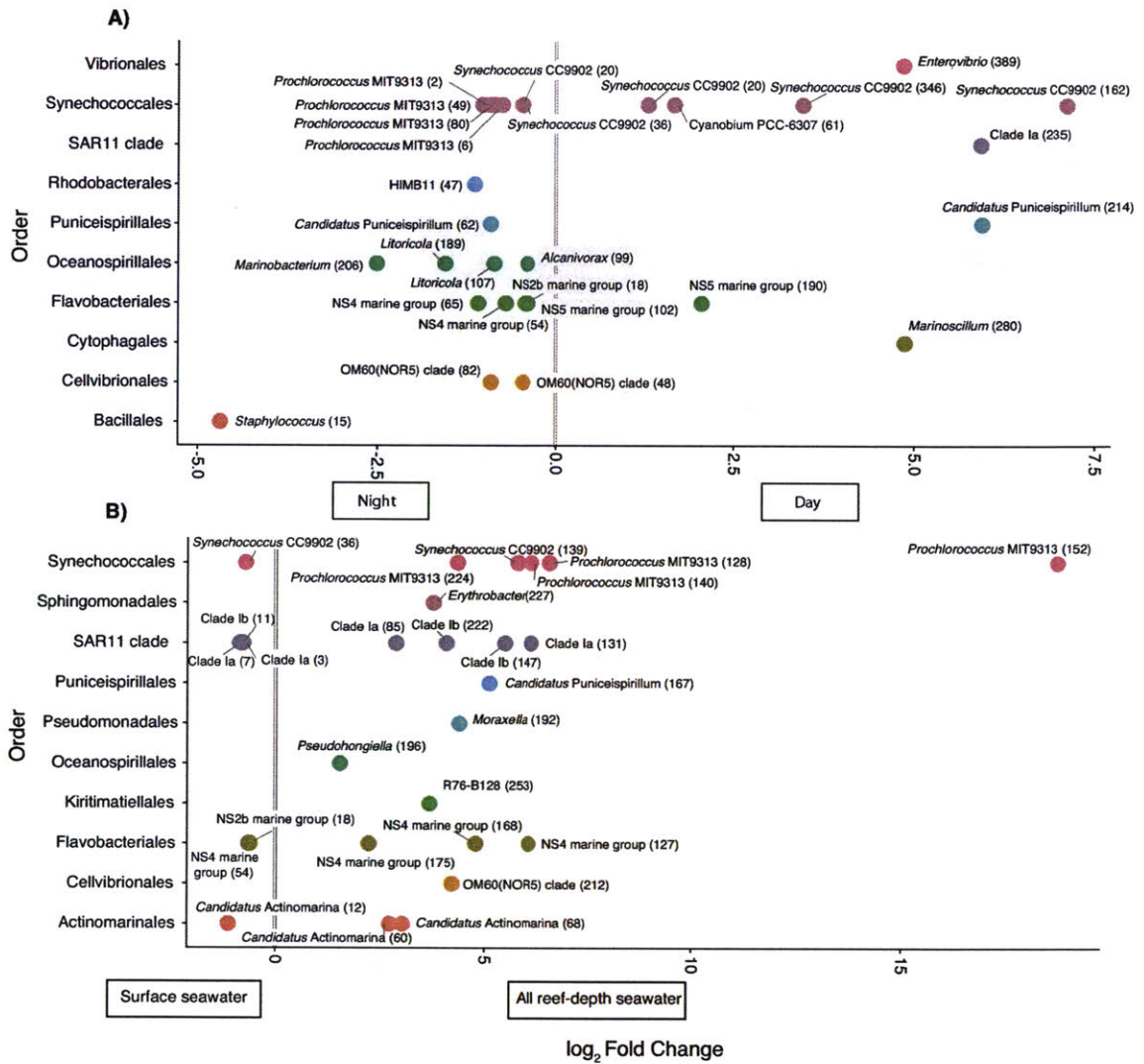
Figure 8. Significantly differential enrichment of bacterial and archaeal SSU rRNA gene amplicon sequence variants (ASVs) across A) reef-depth and coral ecosphere samples collected during the day or night and between B) all reef-depth seawater (reef-depth and ecosphere samples) and surface seawater. Each point represents an individual ASV labeled at the genus-type taxonomic level and the color reflects the order. Only ASVs matched at the genus level are plotted. The number in parentheses indicates the unique number assigned to each ASV sequence.

4.11 Supporting Information
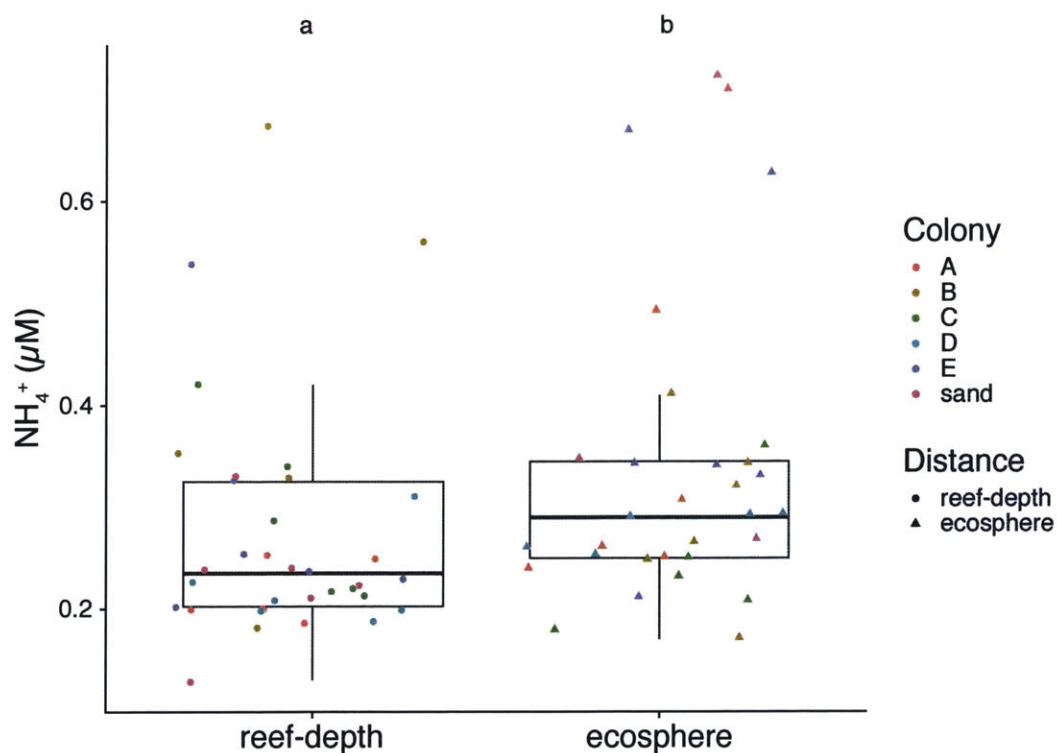
Figure S1. Comparison of ammonium (NH4+) concentrations between reef-depth and ecosphere seawater samples. Each point represents a sample. Point shape corresponds to sampling distance from the coral and point color reflects the colony adjacent to where sampling was conducted. Gray shading indicates samples collected at night. Lower and upper edges of the boxplot correspond to the first and third quartiles, the whiskers extend to the largest or smallest value at 1.5 times the interquartile, and the black bar across the box represents the median. Boxplots with different letters denote significantly different concentrations (Kruskal-Wallis Rank sum test, Dunn's test with Bonferroni corrections, p < 0.05).
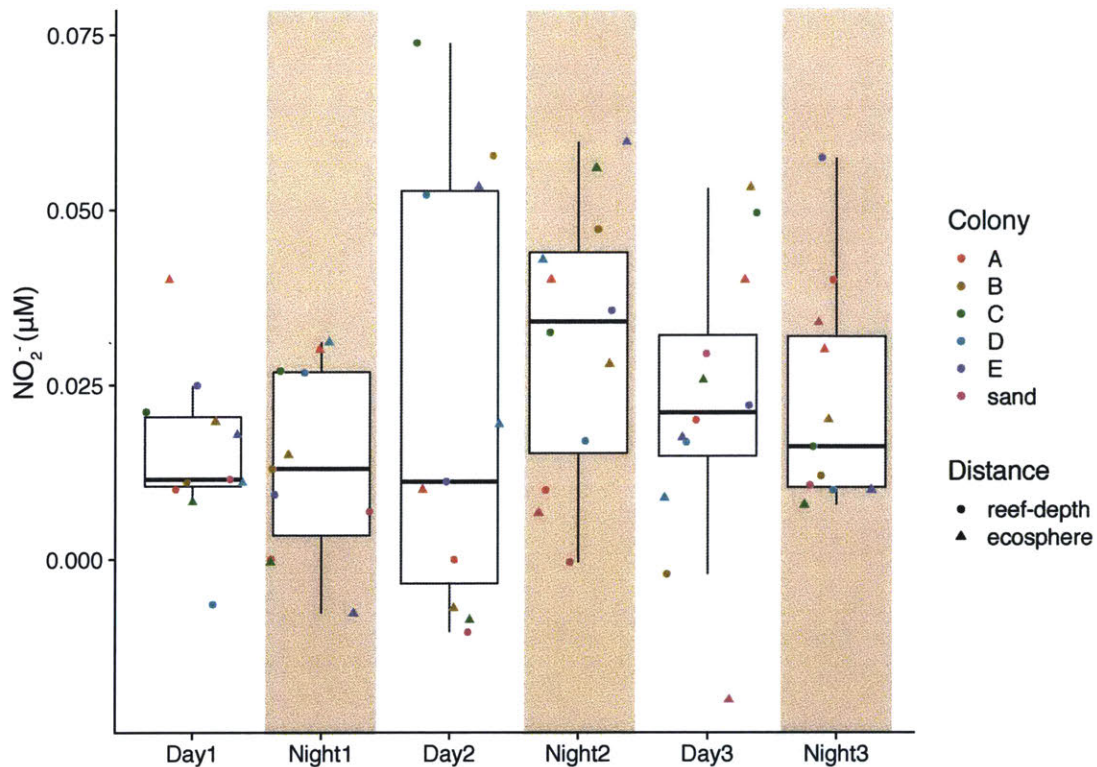
Figure S2. Comparison of nitrite (NO_2^-) concentrations over three days. Each point represents a sample. Point shape corresponds to sampling distance from the coral and point color reflects the colony adjacent to where sampling was conducted. Gray shading indicates samples collected at night. Lower and upper edges of the boxplot correspond to the first and third quartiles, the whiskers extend to the largest or smallest value at 1.5 times the interquartile, and the black bar across the box represents the median. Boxplots with different letters denote significantly different concentrations (ANOVA with Tukey's HSD test, $p < 0.05$).

Chapter 5

Microbial signatures of protected and impacted Northern Caribbean reefs: changes from Cuba to the Florida Keys

5.1 Originality-Significance Statement

Microorganisms play important roles in the biogeochemistry of coral reefs, but our understanding of reef microbial biogeochemistry in the context of reef health and protection is lacking. Here we comprehensively analyzed reef seawater microbial biogeochemistry in a highly protected 'crown jewel' Cuban reef-system and compared these features with more impacted reef-systems within Cuba and the Florida Keys. We demonstrate that protected and offshore Cuban reefs exhibit higher microbial alpha diversity, elevated community similarity, and nutrient-based signatures of oligotrophy compared to more impacted and nearshore reefs. This work suggests that offshore and highly protected reefs harbor distinct microbial biogeochemical signatures. Additionally, this knowledge may aid resource managers as they strive to protect and restore Caribbean coral reefs during a significant time of habitat and climate-based change.

5.2 Summary

There are few baseline reef-systems available for understanding the microbiology of healthy coral reefs and their surrounding seawater. Here, we examined the seawater microbial ecology of 25 Northern Caribbean reefs varying in human impact and protection in Cuba and the Florida Keys, USA, by measuring nutrient concentrations, microbial abundances, and respiration rates as well as sequencing bacterial and archaeal amplicons and community functional genes. Overall, seawater microbial composition and biogeochemistry were influenced by reef location and hydrography. Seawater from the highly protected 'crown jewel' offshore reefs in Jardines de la Reina, Cuba had low concentrations of nutrients and organic carbon, abundant *Prochlorococcus*, and high microbial community alpha diversity. Seawater from the less protected system of Los Canarreos, Cuba had elevated microbial community beta diversity whereas waters from the most impacted nearshore reefs in the Florida Keys contained high organic carbon and nitrogen concentrations and potential microbial functions characteristic of microbialized reefs. Each reef-system had distinct microbial signatures and within this context, we propose that the protection and offshore nature of Jardines de la Reina may preserve the oligotrophic paradigm and the metabolic dependence of the community on primary production by picocyanobacteria.

5.3 Introduction

Caribbean coral reefs have undergone dramatic changes over the past 35 years. The collective impacts of climate change, overfishing, and coastal development have caused shifts in functioning and energy transfer in coral reef ecosystems and these changes have been documented at the level of macro-organisms (Carpenter 1988; Gardner et al. 2003; Miller et al. 2009; Valdivia et al. 2017). In contrast, the impacts of these stressors on reef microbial communities have not been comprehensively documented because molecular techniques for characterizing uncultivated microbes were unavailable prior to the widespread decline of Caribbean coral reefs. This has led to critical gaps in our understanding of how microorganisms, the smallest and most abundant members of Caribbean coral reefs, have changed in abundance, composition, and function alongside broader ecosystem changes.

Global studies have shown that reefs harbor distinct microbial taxa and genomic adaptations compared to cells found in off-reef waters (Nelson et al. 2011; Kelly et al. 2014), suggesting that unique microbial processes occur on coral reefs. Additionally, human impacts (overfishing, pollution) may lead to shifts in reef trophic structure that favor microbial growth (Jackson et al. 2001). On coral reefs, this process of 'microbialization' begins when grazing pressure on algae is lessened due to the removal of herbivorous fish and sea urchins (Hughes et al. 2007; McDole et al. 2012; Haas et al. 2016). Removal of grazers leads to increases in macroalgae (Hughes et al. 2007). More macroalgae may then lead to increases in the standing stock of dissolved organic carbon (DOC) within the water column, increases in the abundance, respiration, and virulence/pathogenicity of heterotrophic microbes, and a net draw down of DOC (Haas et al. 2016). This mechanistic model is referred to as the DDAM (DOC, disease, algae, and microbes) model and has been suggested as one of the invisible causes for the global

degradation of coral reefs (Barott and Rohwer 2012; Haas et al. 2016). Despite the attention dedicated to understanding the microbiology of declining coral reef ecosystems, there are still numerous unknowns surrounding the microbiology supporting healthier coral reefs, especially within the Caribbean, which harbor distinct and less diverse corals communities compared to Indo-Pacific reefs.

Jardines de la Reina (JR) is a protected reef-system in Cuba that may provide useful insight into the microbial ecology of relatively healthy Caribbean coral reefs. The reefs of JR were historically protected from human activities due to their remote nature and are now further protected because maritime traffic, fishing, and recreational diving and tourism are limited within the boundary of Marine National Park (est. in 1996) that encapsulates most of the archipelago. Additionally, Cuba does not currently have large-scale industrialized agriculture or extensive development along most of its coastline (Galford et al. 2018; González-Díaz et al. 2018), minimizing the degree to which nutrient run-off and sedimentation may impact the surrounding waters. The Ana Maria Gulf, referred to here as the JR gulf, spans the inner sea between the island of Cuba and is populated by small mangrove keys, extensive seagrass meadows, and unvegetated sea beds. These features within the JR gulf have likely reduced pollution as well as human-induced sedimentation and eutrophication (Galford et al. 2018; González-Díaz et al. 2018). Together, the protection of JR from human impacts as well as the ecological services provided by mangrove and seagrass biomes (Mumby et al. 2004; Guannel et al. 2016) have likely buffered JR from direct human-induced stressors that plague other reefs in the Caribbean. As a result, this reef-system is regarded as a 'crown jewel' because it supports some of the highest fish biomass (including top predators like sharks and groupers) and coral cover in the Caribbean (Valdivia et al. 2017; González-Díaz et al. 2018).

The more impacted reef-system of Los Canarreos (CAN), Cuba lies ~230 km to the west of JR. CAN encompasses three important keys that have less stringent protection compared to JR: Cayo Largo is an Ecological Reserve and the Rosario and Cantiles Keys are Faunal Refuges. Due to increased accessibility, reefs within CAN are more impacted by humans through subsistence and illegal fishing, tourism, and the diving industry compared to the remote and protected reefs within JR. Fishing has resulted in overexploitation of important finfish and invertebrates in most of Cuba's waters with the exception of the central area within JR (Baisre 2017). The proximity of CAN to JR and the higher degree of human impact present an opportunity to examine the differences in biogeochemistry and microbiology between these two Cuban reef-systems.

The reef-system of the Florida Keys (FK) is located in close proximity to JR and CAN, but has experienced more anthropogenic impacts relative to the Cuban reef-systems. Reefs within FK (spanning ~570 km) are situated close to developed land within FK and South Florida and development activities have influenced the water quality in Floridian waters (Lapointe and Clark 1992; Lapointe et al. 2004). Additionally, FK hosts 2-3M tourists annually (Leeworthy et al. 2010), and many of these visitors engage in water activities such as boating, fishing, and scuba diving. The health of these reefs has been declining precipitously since they were first studied: algal phase shifts, eutrophication of the water column with decreases in water quality, pollution, high prevalence and spread of coral diseases, and loss of coral cover have afflicted these reefs (Szmant and Forrester 1996; Lapointe et al. 2004; Precht et al. 2016). Additionally, commercial and recreational fishing have overexploited over 50 species of predatory fish within FK (Ault et al. 1998; Ault et al. 2006). To combat these stressors, FK was designated as a national marine sanctuary in 1990 and separated into distinct marine zones. Fishing and

harvesting of any marine life are prohibited in only a small portion of these zones and public access to the reefs for recreational fishing and diving is allowed in most areas.

We designed this study to identify field-based microbial signatures of the protected and relatively healthy reef-system of Jardines de la Reina and to describe how biogeochemistry and reef water microbial communities change along a gradient of human impact. We expected to observe microbialized reefs within the Florida Keys compared to the reef-systems in Cuba. Additionally, we hypothesized that there would be small-scale differences in biogeochemistry and microbial community composition within each reef-system because we included reefs along potential hydrographic gradients and different distances from land in order to obtain an understanding of this variability.

5.4 Results

Water sampling and reef surveys were conducted at reef locations across the three reef-systems: Jardines de la Reina, Cuba (JR; 6 reefs), Los Canarreos, Cuba (CAN; 13 reefs – reef surveys were conducted at 5 reefs) and the Florida Keys, USA (FK; 6 reefs) (Figure 1, Table S1). The sampled reefs were grouped into 5 different subregions -- JR offshore, JR gulf (Ana Maria Gulf), CAN, FK offshore and FK nearshore – *a priori* to capture spatial, environmental, and anthropogenic (when applicable) gradients across each reef-system. At each reef, divers surveyed the reef composition and then sampled surface (<1 m) and reef-depth seawater (within 1 m of the reef).

Reef composition

By reef-system, average living coral cover was significantly higher (ANOVA, $F_{(2,14)} = 4.89$, $p = 0.025$) in JR and FK compared to CAN and was similar between JR and FK (Tukey's multiple

comparisons of means, adjusted p-value <0.05; Figure 2). In contrast, average total algal cover was significantly higher (ANOVA, $F_{(2,14)} = 5.82$, $p = 0.014$) in CAN compared to JR and similar to the algal cover in FK (Tukey's multiple comparisons of means, adjusted p-value <0.05; Figure 2). Reef composition also varied locally within each reef-system (Figure 2). Within JR, the offshore forereefs (1 and 2) had average coral and algal covers of 27.4% and 52.5%, respectively (Figure 2, Table S2). Site 5 within the JR gulf had 55.3% coral cover, the highest measured in this study. In CAN, the average coral cover among the five surveyed reefs was 5.4% and the algal cover was 85.3% (Figure 2, Table S2). Of the six reefs surveyed in the Florida Keys, the nearshore site (25) had the highest macroalgal cover (26.0%) as well as the highest total cover of algae (94.7%) out of all the Florida sites and the lowest live coral cover (1.9%) out of all of the reefs surveyed (Figure 2, Table S2).

Macronutrients

The concentrations of organic and inorganic macronutrients were measured at all sites in seawater collected from surface and reef-depths. The sampling location of reef-depth seawater varied across our sites ranging from an average depth of 5.1 m (0.75 – 16 m range) in JR, 5.2 m (1 – 14 m) in CAN, and 3.8 m (1 – 6 m) in FK (Table S1, Supporting Information Methods). Concentrations of total organic carbon (TOC) (ANOVA, $F_{(7,27)} = 78.19$, $p < 0.05$), total organic nitrogen (TON) (ANOVA, $F_{(7,74)} = 53.44$, $p < 0.05$), and dissolved inorganic nitrogen (nitrite, nitrate, and ammonium) (ANOVA, $F_{(7,74)} = 4.21$, $p < 0.05$) were significantly higher within the JR gulf as well as nearshore FK compared to the offshore JR reefs (Tukey's multiple comparisons of means, adjusted p-value < 0.05; Figure 3, Figure S1). Nitrate, nitrite, and ammonium were barely detectable in offshore JR reef seawater (Figure 3, Figure S1). Concentrations of nitrite (ANOVA, $F_{(7,75)} = 7.38$, $p < 0.05$) and nitrate (ANOVA, $F_{(7,76)} = 3.39$, $p < 0.05$) were significantly higher

in CAN and FK compared to JR (Tukey's multiple comparisons of means, $p < 0.05$; Figure S1). Concentrations of ammonium were similarly low across most of the reef locations and depths (Figure S1). Concentrations of TOC, TON, and silicate (ANOVA, $F(7,76) = 14.11$, $p < 0.05$) were significantly higher (Tukey's multiple comparisons of means, $p < 0.05$) in FK nearshore seawater compared to seawater from other reefs (Figure 3, Figure S1). Lastly, nutrient concentrations between surface and reef-depth seawater within each subregion were not significantly different (ANOVA, $p > 0.05$; Figure 3, Figure S1).

Microbial abundances and carbon contributions

Across reef-systems, *Prochlorococcus* abundances were significantly higher (Kruskal-Wallis, Chi^2 = 18.33, df = 4, $p = 0.001$) within JR offshore reef-depth and surface seawater compared to JR gulf reef-depth seawater, CAN surface and reef-depth seawater, and nearshore FK seawater, but not significantly different from abundances in FK offshore reef-depth and surface seawater (Conover's post-hoc test, adjusted p-value < 0.05, Figure 4A). In fact, *Prochlorococcus* was significantly more abundant (approximately six times higher) in offshore JR (surface and reef-depth) as well as JR gulf surface seawater compared to JR gulf reef-depth seawater (Fig 4A). *Synechococcus* abundances followed the opposite pattern and were on average six-fold higher at sites located within the JR gulf compared to the offshore JR reefs, but not significantly different because of variability between reefs (Figure 4C). The abundance of unpigmented cells, generally heterotrophic bacteria and archaea, was mostly similar across all reefs and reef-systems, but elevated within reef-depth gulf seawater (JR5 and 6) and highest within nearshore FK seawater (Figure 4E). Picoeukaryotic cell abundances were more similar between reef-systems (Figure S2A). Calculations of carbon biomass demonstrated that *Synechococcus* contributed carbon

213

biomass to all regions, with up to 12.5 μg of carbon l^{-1} in JR gulf seawater (Figure 4D). *Prochlorococcus* contributed up to 3 μg of carbon l^{-1} in offshore JR reef seawater (Figure 4B).

Phytoplankton

Chlorophyll *a* concentrations were generally low (ranging from 0.053 to 0.337 μg l^{-1}) (Figure S3), but changes in phytoplankton community composition were observed across reef-systems. Phytoplankton community assemblages from Cuban reefs were dominated by cyanobacteria, comprising average relative abundances of 39% (JR) and 29% (CAN) of the phytoplankton community (Figure 5). In contrast, FK reef seawater had a greater representation of the 11 other measured phytoplankton functional classes with significantly less cyanobacteria (14 +/- 8%; ANOVA, $F(2,25) = 15.98$, $p = 3.38E-5$; Tukey multiple comparisons of means, $p < 0.05$) (Figure 5). The relative abundance of diatoms was significantly higher (ANOVA, $F(2,25) = 4.032$, $p = 0.030$) in FK (16%) compared to JR (5%) (Tukey multiple comparisons of means, $p = 0.026$) (Figure 5).

Microbial alpha diversity

Offshore reef seawater in JR had the highest microbial alpha diversity (measured here as microbial richness), as indicated by number of minimum entropy decomposed (MED) nodes of bacterial and archaeal SSU rRNA gene amplicons (median of 333.5 (range: 185 - 359) MED nodes) (Figure 6). Offshore JR seawater had significantly (Kruskal-Wallis, $Chi^2 = 21.41$, $df = 4$, $p = 2.6E-4$) higher alpha diversity compared to CAN (Dunn's test, adjusted $p = 0.0001$; Figure 6) as well as FK nearshore reef seawater (Dunn's test, adjusted $p = 0.042$; Figure 6). Offshore FK seawater had the next highest median alpha diversity (275.5 (range: 141 - 330) MED nodes), followed by FK nearshore reef seawater (median alpha diversity 202 (range: 101 - 261) MED nodes). Microbial alpha diversity in CAN reef seawater had the lowest median richness of 140

(range: 58 - 336) MED nodes (Figure 6). The variation in microbial alpha diversity between reefs within CAN as well as FK was larger compared to JR, with the largest range encountered in CAN (58 - 360 MED nodes) (Figure 6).

Microbial community composition

A nested permutational multivariate analysis of variance (PERMANOVA; Adonis) test on the Bray-Curtis dissimilarity index of reef seawater bacterial and archaeal SSU rRNA gene amplicons grouped into MED nodes indicated that region (reef-system; JR, CAN, or FK), subregion, reef location, and sampling depth influenced the composition of reef seawater communities, with reef location contributing the most to variation in community dissimilarity (Table 1). Additionally, a non-metric multidimensional scaling analysis (NMDS) corroborated these results (Figure 7A). At a broader scale, microbial communities collected from the same reef-system and subregion were more similar to each other (Figure 7A and 7B). In the NMDS, all CAN seawater microbial communities were ordinated in the positive plane of the y-axis (NMDS2) and separated from JR and FK microbial communities (Figure 7A). Surprisingly, there was high similarity in community composition between sites 22 and 23 in the FK and the JR offshore forereefs (JR 1 and 2). Microbial community dispersion was higher and more variable in seawater collected from CAN and FK offshore compared to JR offshore and JR gulf, indicating higher beta diversity across these subregions (Figure 7B).

Within the NMDS, microbial communities sampled within JR ordinated together by location; microbial compositions from reefs JR 1 and 2 were more similar to each other than to the other communities sampled from sites located within the JR gulf (Figs. 1D and 7). Compared to the other reef-systems, all communities from JR grouped closer together and had less variance in community composition relative to CAN and FK microbial communities (Figure 7). The

215

pattern of ordinating by general geographic location was not as evident in microbial communities

collected from CAN and FK (Figure 7A). Microbial community composition from site 25, one of

the reefs closest to Summerland Key, was more dissimilar from the other FK microbial

communities (Figs. 1B and 7A). Environmental variables were fitted to the NMDS ordination

using vector fitting ('envfit' function) and this procedure indicated that picoeukaryote abundance

($R^2 = 0.11$, p = 0.040) and nitrite ($R^2 = 0.33$, p = 0.001) and silicate ($R^2 = 0.16$, p = 0.010)

concentrations were significantly correlated with the ordination of microbial communities in the

NMDS.

Regionally specific microbial taxa

Reef seawater microbial community composition, assessed using bacterial and archaeal SSU

rRNA gene amplicons, showed some variability at the level of Phylum (Figure S4).

Cyanobacteria were more abundant within CAN (29.7 ± 18.3%; mean and standard deviation)

compared to JR (18.5 ± 5.8%) and FK (13.2 ± 5.4%) (Figure S4). *Prochlorococcus* sequences

were 99.2 to 99.6% identical to MIT9313, a low-light ecotype of *Prochlorococcus*, and this was

the only ecotype identified within the amplicon-based survey. Bacteroidetes was most

represented in FK reef seawater, with an average relative abundance of 16.8 ± 8.3%, and less

abundant in JR (12.7 ± 5.7%) and in CAN (8.6 ± 6.4%) (Figure S4). Verrucomicrobia also

followed the same trend as Bacteroidetes and had the highest relative abundance in FK reef

seawater (2.0 ± 1.4%) compared to JR (0.4 ± 0.3%) and CAN (0.6 ± 1.6%) (Figure S4).

Euryarchaeota were detected on nearly all reefs, with average relative abundances of 1.0 ± 0.7%

in JR, 0.7 ± 0.8% in CAN, and 1.3 ± 1.8% in FK (Figure S4).

Enrichment comparisons of specific taxa within reef-depth seawater collected from the

most biogeochemically distinct reef-systems of JR and FK revealed that 44 discrete MED nodes

216

were differentially abundant (p-adjusted <0.05, Benjamini-Hochberg correction for multiple

testing) (Figure 8). The 34 enriched taxa in JR belonged to microbial groups typically found in

reef seawater environments. Alphaproteobacteria comprised 29% of the enriched MED nodes,

including the SAR116 clade, Surface 1 and 2 groups within the SAR11 clade, and

Rhodobacteraceae (Figure 8). Cyanobacteria accounted for 20.6% of reads enriched within JR

seawater, with most of the representative MED node sequences identifying as *Synechococcus*.

Lastly, while the prevalence of Archaea was low across the entire dataset, MED nodes affiliated

with Marine Groups II and III within the Thermoplasmata were significantly enriched in JR

communities (5.8% of enriched sequences) (Figure 8). MED nodes significantly depleted in JR

and enriched in FK reef-depth seawater were mostly comprised of Bacteroidetes (50%),

Alphaproteobacteria (20%), and Verrucomicrobia (20%) (Figure 8). More specifically, MED

nodes affiliated with *Formosa* and *Coraliomargarita* were enriched within FK seawater

compared to seawater collected from JR (Figure 8). All of these MED nodes were present across

the dataset at low relative abundances (Table S3).

Functional differences between Jardines de la Reina and the Florida Keys

Metagenomic sequencing and analysis of the whole microbial community (eukaryotes, bacteria,

archaea, and DNA viruses) in reef-depth seawater from JR and FK resulted in 163 significantly

different functional genes (Figure 9). These genes were grouped into KEGG modules as well as

metabolic pathways. JR metagenomes were enriched in photosynthesis and nitrogen metabolism

pathways (Figure S7) and the KEGG modules of nitrate assimilation, assimilatory nitrate

reduction, the capsular polysaccharide transport system, and the NAD(P)H: quinone

oxidoreductase enzyme (for chloroplasts and cyanobacteria) (Table 2). Metabolic pathways

enriched in FK included fructose and mannose metabolism, pentose and glucoranate

217

interconversions, lipopolysaccharide biosynthesis, toluene degradation, valine, leucine and

isoleucine biosynthesis, as well as the microbial metabolisms in diverse environments category

(including degradation and metabolism of xenobiotics, and energy metabolism of diverse

compounds) and pathway coverage ranged from 0.02 to 0.1 (Table 2). The KEGG modules that

were enriched in FK included fumarate reductase and the degradation step of benzene to catechol

involved in benzene degradation with module coverages ranging from 0.2 to 1 (Table 2).

Community respiration rates

Water column community respiration was determined by monitoring oxygen through time in

dark incubations. Most of the reefs (81%) had positive community respiration rates that ranged

from 0.3 to 16.7 μmol of O_2 consumed $l^{-1} d^{-1}$ (Figure 10). The highest respiration rate of 16.7

μmol $O_2 l^{-1} d^{-1}$ was measured in offshore FK seawater collected from site 21. Negative

respiration rates, implying net oxygen production, were observed in seawater collected from JR

4, CAN 9, and sites FK 23 and 24 (Figure 10). These values ranged from 0.3-6.9 μmol of O_2

produced, with the highest O_2 production at site FK 23 (Figure 10).

Across reef-system relationships between microbial diversity, microbial abundances, and coral

cover

Relationships were examined across the measured parameters, and for brevity only those with

significant results are reported. There was a significant negative regression ($R^2 = 0.33$, p = 0.010,

Figure 11A) between microbial community alpha diversity and heterotrophic cell abundance

across JR and the FK, with less microbial alpha diversity and slightly higher heterotrophic

abundance in FK nearshore reef seawater. However, this regression was not significant when

seawater from CAN was included ($R^2 = -0.02$, p = 0.49, Figure 11B). We also detected a

significant positive regression between the abundance of picocyanobacteria (summation of

218

Prochlorococcus and *Synechococcus* cell abundances) and coral cover across the reef-systems ($R^2 = 0.54$, p = 0.001, Figure S5).

5.5 Discussion

This study compared reef seawater biogeochemistry and microbial communities between protected and impacted Northern Caribbean reefs with the goal of deciphering distinct microbial features. We found that Jardines de la Reina is an oligotrophic reef-system characterized by taxonomically diverse microbial communities with high community similarity and abundant picocyanobacterial biomass, whereas Los Canarreos and the Florida Keys reefs experience more spatial variability in reef seawater microbial community alpha diversity and composition. Furthermore, the spatial variability within Los Canarreos reefs may be driven by release of nutrients from nearby wetlands (the Zapata Swamp, see below) and the hydrodynamic regime created by the complex array of cays and channels. The variability in the Florida Keys may be mostly impacted by increased concentrations of organic and inorganic macronutrients, higher productivity, and/or terrestrial sources of sediments from developed land. The nearshore reefs in the Florida Keys exhibited a few signs of microbialization, but this process was not as evident on the Florida Keys offshore reefs surveyed in this study. The microbial regimes observed across the reef-systems appear to be determined by the intersection of local anthropogenic impact as well as oceanographic processes.

Biogeochemical and microbial features of Jardines de la Reina

A majority of the macronutrient concentrations were low or barely detectable in Jardines de la Reina and are similar to concentrations measured in other oligotrophic systems including the Sargasso Sea, the North Subtropical Pacific Gyre, the Red Sea, and other reefs in the Caribbean

and Pacific, suggesting rapid turnover of these nutrient pools by microorganisms (Lewis 1977; Westrum and Meyers 1978; Sorokin 1995; Karl et al. 1996; Dore et al. 2008). Organic carbon concentrations in JR (especially at JR 1 and 2) were similar to concentrations reported from a reef-crest in Grand Cayman (Westrum et al. 1978).

Nutrient dynamics across JR are likely influenced by differences in hydrodynamics between offshore forereefs and patch reefs within the JR gulf. The forereefs are flushed with pelagic, oligotrophic seawater that is carried to them by the Caribbean current, whereas the patch reefs within the gulf are influenced by productive mangrove forests and seagrass meadows that have less contact with the open ocean. Entrainment of nutrients from these productive biomes within the gulf and tidal flushing of these nutrients onto the forereefs are likely important processes that influence primary productivity, microbial diversity and metabolism, and grazing of cells by the reef community in Jardines de la Reina.

Picocyanobacterial abundances in Jardines de la Reina were similar to abundances observed within oligotrophic open-ocean environments (DuRand et al. 2001; Zinser et al. 2006; Charpy et al. 2012), but were two orders of magnitude higher than abundances detected in seawater from Pacific reefs (Charpy et al. 2012). Furthermore, reef seawater collected from the offshore forereefs in JR had high abundances of *Prochlorococcus* whereas there was a shift to high, but variable abundances of *Synechococcus* in seawater collected from within the JR gulf. This negative relationship between *Prochlorococcus* and *Synechococcus* has been observed previously and tracks with increased macronutrient concentrations and proximity to land (Cox et al. 2006; Yeo et al. 2013). Additionally, the ratio of picocyanobacteria to unpigmented cells was very similar between offshore and gulf reefs in JR, potentially indicating similar nutrient or grazing controls on both populations. We expected that this change in the nutrient regime would

220

select for different ecotypes of *Prochlorococcus*, but all the sequences that identified as *Prochlorococcus* were similar to MIT9313, a low-light adapted ecotype (Rocap et al. 2003). In our study, there was a discrepancy between the trends observed in *Prochlorococcus*, *Synechoccocus*, and unpigmented cell abundances determined using flow cytometry and amplicon-based relative abundances. For example, cell counts for *Prochlorococcus* and *Synechoccocus* were two orders of magnitude lower than unpigmented cells, yet they still comprised a large portion of the bacterial and archaeal community based on relative abundances generated from amplicon-based community analyses. This discrepancy likely arose because amplicon-based sequencing data is not quantitative and cannot be directly compared with cell abundance data, as has been observed previously (reviewed within Martiny et al. 2009).

Despite their prevalence on reefs, the ecological roles of *Prochlorococcus* and *Synechococcus* within reef microbial communities have only been investigated in a few cases (Charpy et al. 2012; McDole Somera et al. 2016). These picocyanobacteria are some of the most important primary producers in reef seawater and they are directly and indirectly grazed by single-celled eukaryotic heterotrophs, mixotrophic plankton, and reef organisms like corals and sponges, effectively linking photosynthetically fixed carbon from the water column to animals on the reef (Sorokin 1995; Ferrier-Pages and Gattuso 1998; Bertilsson et al. 2005; Patten et al. 2011; Charpy et al. 2012; McNally et al. 2017). The high prevalence of *Prochlorococcus* and *Synechococcus* observed in this study demonstrates that picoyanobacterial dynamics on reefs should be explored further from energetic as well as community network perspectives.

Reef seawater from Jardines de la Reina had higher microbial alpha diversity and smaller beta diversity compared to seawater from CAN and FK. There was also a negative relationship between microbial alpha diversity and heterotrophic bacterial abundance between JR and FK,

indicating a potential trade-off between community alpha diversity and biomass across the different reef-systems. The consistent supply of oligotrophic seawater from the Caribbean current to JR forereefs likely enhances niche partitioning within microbial communities and leads to higher alpha diversity. The hydrodynamic regime likely contributes to the high microbial community similarity across this reef-system through mixing processes. On the opposite end of the spectrum, in more disturbed and/or nutrient-rich environments within Los Canarreos or the Florida Keys, microbial alpha diversity tends to be lower or the beta diversity is higher and more variable, suggesting that disturbances on these reefs favor active growth of fewer dominant microorganisms that outcompete other cells within the population for resources (Kearns et al. 2016; Reese and Dunn 2018). Additionally, genes indicative of photosynthesis and nitrogen metabolism were enriched in JR compared to FK, indicating the importance of photosynthesis and nitrogen acquisition in oligotrophic waters. Fewer genes were significantly enriched in JR compared to FK as well, suggesting a higher degree of functional redundancy and homogeneity across the more taxonomically diverse microbial communities in JR. The links between microbial alpha diversity and functional diversity continue to be debated (Louca et al. 2018), but our findings demonstrate that alpha diversity, in the context of reef microbial communities surveyed in JR, CAN, and FK, may be a meaningful feature of protected reefs.

Potential influence of the nutrients from wetlands within Los Canarreos

Reef seawater microbial beta diversity was higher and more variable in Los Canarreos compared to communities from Jardines de la Reina and the Florida Keys. In contrast, there was less variance in the inorganic and organic macronutrient concentrations, picocyanobacterial abundances, and phytoplankton community compositions across Los Canarreos. Overall, CAN reefs were less oligotrophic than the forereefs in Jardines de la Reina and the phytoplankton

community was mostly comprised of eukaryotic phytoplankton including diatoms and

dinoflagellates, suggesting episodic instances of high water-column productivity on these reefs.

Additionally, the productivity of seawater microbial communities in CAN could be stimulated by

nutrients and organic matter released from the Zapata Swamp, an extensive wetland (Galford et

al. 2018) that is located ~60 km from this reef-system.

Elevated nutrients near land in the Florida Keys

Nearshore reefs in the Florida Keys had the highest organic carbon, total organic nitrogen, and

silicate concentrations compared to all the other reefs in this study. In fact, the total organic

carbon and nitrogen concentrations were 2-3 times higher on the nearshore reefs compared to the

offshore reefs, on par with other observations within the Florida Keys (Szmant et al. 1996;

Briceno and Boyer 2015; Apprill et al. 2016). Terrestrial run-off and sediment intrusion are

likely partially responsible for the high TOC, TON, and silicate concentrations on these

nearshore reefs, but we cannot definitively discern the relative contributions of sediment vs.

biological productivity because we did not measure sediment load. Despite the elevated organic

carbon concentrations, community respiration rates were not higher, but more variable than rates

measured in reef seawater from JR and FK.

The most notable differences in reef seawater microbial community composition between

JR and FK included the decrease and absence of *Prochlorococcus* cells on the FK offshore and

nearshore reefs, increase in the relative composition of Bacteroidetes, and detection of

Roseibaccilus and *Coraliomargarita*, both members of the Verrumicrobia phylum, across all FK

reefs. Bacteroidetes have been associated with marine particles and detected at high relative

abundances following phytoplankton blooms (Pinhassi et al. 2004; Teeling et al. 2012).

Furthermore, Bacteroidetes can degrade high molecular weight polymers as well as synthesize

223

adhesion proteins for attaching to particles (Fernandez-Gomez et al. 2013). Verrumicrobia are also particle-associated, although they can be free-living, and typically recovered from terrestrial soils (Bergmann et al. 2011; Freitas et al. 2012). That being said, Verrumicrobia are also detected ubiquitously in seawater and at high relative abundances in coastal marine environments (Freitas et al. 2012). Higher abundances of Bacteroidetes and Verrumicrobia suggest that there are more particles in FK seawater compared to JR and CAN which corresponds with higher nutrient availability and shifts in phytoplankton community composition.

In the Florida Keys, we observed higher total chlorophyll *a* concentrations and a phytoplankton population mostly comprised of diatoms, dinoflagellates, and haptophytes. The increased macronutrient concentrations likely enhance the growth of larger eukaryotic phytoplankton and select against the growth of microbial cells that are not tolerant of higher nutrient conditions. Additionally, there were more diverse functional metabolic pathways enriched in FK compared to JR, in agreement with the premise that microbial communities living in environments with more substrates available will have the functional capability to use the available nutrients. Furthermore, genes involved in the pentose – phosphate pathway have been positively correlated with algal cover on microbialized reefs (Haas et al. 2016) and we detected enrichment of this pathway (pentose and glucoranate interconversions) in FK. Microorganisms using the pentose-phosphate pathway can potentially catabolize more diverse carbon sources, including carbohydrates released by algae (Haas and Wild 2010), and this strategy has been shown to provide a selective advantage to microorganisms that need to grow faster than their competitors (Haas et al. 2016). We did not detect significant enrichment of virulence-associated or pathogenic genes in seawater from the Florida Keys compared to Jardines de la Reina which is contrary to other studies that have observed an increase in the abundance of these genes with reef

degradation or increased human impact (Dinsdale et al. 2008; Bruce et al. 2012; Kelly et al. 2012; Kelly et al. 2014; Moreira et al. 2015).

Revisiting the microbialization hypothesis in the context of different reef regimes

We hypothesized there would be significant increases in the abundance of heterotrophic bacteria, enhanced community respiration, higher concentrations of inorganic and organic macronutrients, and shifts from coral to algal dominance on the reefs along the gradient of human impact. However, we did not observe significant changes in most of these parameters. Overall, hydrogeography and subregion were the largest influences contributing to reef similarity. Offshore reefs in both JR and the FK were oligotrophic, had high abundances of picocyanobacteria, high microbial alpha diversity, and more constrained microbial beta diversity, although the magnitudes of the contrasts were different within each reef-system. The only reefs that supported some of the predictions of the DDAM model were the nearshore reefs in the Florida Keys. These two nearshore reefs had significantly higher concentrations of organic macronutrients, very low abundances of *Prochlorococcus*, and significant enrichment of particle-associated and copiotrophic microbial taxa. Our observations indicate that the process of microbialization on reefs may be more nuanced and that there are additional aspects of hydrogeography that impact these processes, resulting in different reef regimes. In this study, we surveyed a spectrum of reef regimes across JR, CAN, and FK, but we recognize that some subregions (e.g. JR offshore, JR gulf, FK nearshore) have fewer data points compared to the other categories due to sampling limitations and that care needs to be taken when interpreting the statistical differences between subregions. That being said, measurements from these locations were similar within each subregion and are likely representative of the environmental conditions. However, future studies would benefit from collecting samples at a higher spatial or temporal

resolution in order to unravel the process of microbialization on reefs. Additionally, there are other examples of reefs that are subjected to high loads of organic and inorganic nutrients as well as pollution, like Vardero reef in Colombia (Pizarro et al. 2017) or reefs subjected to upwelling events (Leichter et al. 2003; Stuhldreier et al. 2015). Comparisons of microbial community dynamics between these drastically different reef regimes like Vardero reef with Jardines de la Reina would extend our knowledge of how microbial communities contribute to energy cycling and reef health.

Relating and applying back to the reef

Coral and algal coverage varied locally (also observed by Caballero Aragón et al. 2019), but did not change drastically across reef-systems, indicating that these metrics may not be the most immediate and sensitive measure of reef health. Additionally, our observations of coral and algal cover are in agreement with another study examining coral diversity and cover on reef-systems surrounding the island of Cuba (González-Díaz et al. 2018). Furthermore, coral cover on JR reefs was lower than the historical baseline of ~50% cover in the Caribbean (Gardner et al. 2003) and there were observations of bleaching and coral disease, indicating that even the remote reefs of Jardines de la Reina are impacted by environmental change and disease (Ferrer et al. 2016; González-Díaz et al. 2018). In addition to coral cover, other aspects of reef composition, including taxonomic or functional compositions of corals, macroalgae and turf algae, and macro-invertebrates, can serve as important metrics of reef health (Smith et al. 2016). Fish abundances and diversity are also used as metrics for reef health and other studies have found that abundances of commercially valuable and larger fish are higher on some reefs located within the JR National Park (Pina-Amargos et al. 2014; Valdivia et al. 2017). That being said, we are

226

lacking an understanding of how the diversity and abundance of fish correlate with reef

biogeochemistry and microbial ecology and this should be addressed by future studies.

Reef microbial ecology may instead be a more immediate and sensitive measure of reef

health than coral cover or vertebrate abundance. A growing body of research has introduced the

concept of using microorganisms as bioindicators on reefs (reviewed within Glasl et al. 2017) as

well as to predict changes in environmental conditions (Glasl et al. 2019) and the research

presented here builds upon this knowledge. We have demonstrated that the microbial signatures

of high alpha diversity, high community similarity, and high prevalence of *Prochlorococcus* may

be important indicators for reef managers and restoration specialists to acknowledge. For

example, there is significant interest in restoring reefs by outplanting coral colonies onto existing

reefs. While general oceanographic conditions are sometimes considered when defining sites for

these efforts, reef seawater microbial ecology is not typically factored into these site decisions.

As additional datasets like the one presented here emerge and we further link microbial dynamics

to reef health, microbial ecology may become a more prevalent and defining factor in reef

restoration efforts.

5.6 Methods

Reef surveys and sample collection

We conducted two separate research expeditions to JR (February 2015) and CAN/FK (April/May

2015) during the Caribbean dry season. Due to sampling limitations within JR, we were only

able to survey and collect samples from two reef sites in the JR offshore subregion and four reefs

within the JR gulf subregion. Scuba divers conducted reef surveys at all JR and five CAN reef

locations to assess the percent cover of different reef organisms and substrates (Supporting

Information Methods). Coverage of a wider diversity of biotypes was assessed at FK reefs using the same methods, but with a different research team (Supporting Information Methods).

At each reef location, hydrographic profiles of the water column were obtained and seawater from surface (<1 m depth) and reef-depth (~ 1 m above reef) was collected for a variety of different analyses (Table S1, Supporting Information Methods). Seawater (4 l for each depth) was collected using a submersible groundwater pump (Mini-monsoon sampling pump, Proactive Environmental Products) and replicate 2 l samples were each filtered onto 0.22 μm pore size, 25 mm Supor® filters (Pall Corporation). Filters were stored in cryovials, flash frozen in liquid nitrogen, transported in a dry-shipper in the field, and stored at -80 °C until DNA was extracted. Additionally, 20 l seawater from each site was filtered onto 0.1 μm pore size, 142 mm Supor® filters for shotgun metagenomic sequencing. Smaller-volume (1 ml) seawater samples were collected and preserved with 1% PFA (final concentration) for flow cytometry (Supporting Information Methods). Seawater (2 – 4 l) was also filtered onto 25 mm Whatman® GF/F glass microfiber filters (GE Healthcare Life Sciences) for phytoplankton pigment analyses. Seawater samples were collected in duplicate from both depths at each reef for analysis of organic and inorganic macronutrients (Supporting Information Methods).

Macronutrient analysis

Total non-purgeable organic carbon (TOC) and total nitrogen (TN; organic and inorganic) concentrations were measured using a Shimadzu TOC-V$_{CSH}$ total organic carbon analyzer (Hansell and Carlson 2001). Concentrations of inorganic macronutrients (PO_4^{3-}, $NO_2^- + NO_3^-$, NO_2^-, NH_4^+, silicate) were analyzed using a continuous segmented flow system (as described within Apprill and Rappé 2011). Nitrate (NO_3^-) concentrations for each sample were calculated by subtracting the NO_2^- concentration from the $NO_2^- + NO_3^-$ concentration. Total organic nitrogen

concentrations were calculated by subtracting the sum of the inorganic nitrogen species

concentrations ($NO_2^-+NO_3^-$ and NH_4^+) from the total nitrogen (TN) concentrations for each

sample.

Phytoplankton pigments

Pigment analysis was conducted using High Performance Liquid Chromatography (Van

Heukelem and Thomas 2001). The Chemtax addition to the R-package limSolve (Soetaert et al.

2009), based on the program CHEMTAX (Mackey et al. 1996), was used to estimate the algal

composition of chlorophytes, cyanobacteria, diatoms, cryptophytes, dinoflagellates, haptophytes

1- 4, and pasinophytes within each sample based upon the concentrations of 12 different

pigments (Pinckney et al. 2015). The initial pigment ratio matrix used to evaluate phytoplankton

composition was taken from Pickney et al. (2015). The converged initial pigment ratio matrix

was used because the phytoplankton assemblages in these samples were not determined with a

microscope.

Cell abundances

Samples collected for cell counts were analyzed with an Altra flow cytometer (Beckman

Coulter) and a laser excitation wavelength of 488. Unstained and stained (SybrGreen I,

Invitrogen™) subsamples were analyzed to estimate the abundance of picocyanobacteria

(*Prochlorococcus, Synechococcus*), picoeukaryotes, and unpigmented cells (proxy for

heterotrophic bacterial abundance) (Marie et al. 1997), respectively. Fluorescence spectra were

binned, analyzed, and transformed into abundances using FlowJo (v. 6.4.7) software. Total

abundances of cells per sample were calculated by adding the abundances for each picoplankton

class together. Estimates of the carbon biomass were calculated by multiplying the abundances

of *Prochlorococcus, Synechococcus*, and unpigmented cells by a carbon conversion factor and

then converting the concentration of carbon to micrograms per liter of seawater. The carbon

conversion factors for each cell type included: 30 fg of carbon per *Prochlorococcus* cell, 200 fg

of carbon per *Synechococcus* cell, and 10 fg of carbon per unpigmented cell (Fukuda et al. 1998;

Cermak et al. 2017).

Microbial community sequencing and analysis using 16S rRNA gene surveys

DNA was extracted from filters using two different DNA extraction methods in order to increase

DNA yield (Santoro et al. 2010; Urakawa et al. 2010). DNA was extracted from duplicate

samples taken at each site and depth to assess reproducibility between samples. Purified DNA

from the two different extraction methods was pooled per sample using the Genomic DNA Clean

and Concentrator kit (Zymo Research Corporation), quantified using the Qubit 2.0 HS dsDNA

assay (ThermoFisher Scientific), and screened for quality using gel electrophoresis (1% TBE

agarose gel) and the HyperLadder™ 1kb marker (Bioline) as a size reference. DNA extraction

and pooling controls (9) were also created to control for potential contamination from reagents.

Lastly, genomic DNA from a microbial mock community (HM-278D, BEI Resources) was

included in the final sample array to account for amplification and sequencing error.

Extracted DNA was amplified and sequenced at the W. M. Keck Center for Comparative

and Functional Genomics (University of Illinois, Urbana, IL). Briefly, V4 hypervariable regions

of the 16S rRNA gene were amplified using the Fluidigm® microfluidics quantitative PCR

platform and prepared for 2x250 bp paired-end Illumina MiSeq sequencing (Supporting

Information Methods). The Fluidgim V4 primer set 515F-Y: 5'-GTGYCAGCMGCCGCGGTAA

(Parada et al. 2016) and 806RB: 5'-GGACTACNVGGGTWTCTAAT (Apprill et al. 2015),

accompanied with Illumina adapters, index, pad, and linker sequences, were used for

amplification (Kozich et al. 2013). Primer-sorted and demultiplexed sequences were quality-

filtered using mothur v.1.36.1 (Schloss et al. 2009). Forward and reverse reads were united and

locus-specific forward and reverse primers were removed. Reads with ambiguous positions or

exceeding 275 bp in length were removed. Next, unknown, mitochondria, or eukaryotic

sequences were identified (method = 'knn') using the Silva database v119 (Quast et al. 2013) and

removed. UCHIME (Edgar et al. 2011) was used to identify and remove chimeric reads

(reference = self). Sequences detected in the DNA extraction and pooling controls are believed to

originate from amplicon contamination during sample processing or cross-talk between

multiplexed samples during sequencing (Wright and Vetsigian 2016) due to their classification

as marine bacteria (unclassified Rhodobacteraceae, Flavobacteria, and SAR11). To be

conservative, these sequences were removed from all samples (146,540 reads; 3% of dataset,

accounting for 107 MED nodes). This removal occurred prior to subsampling so that it had a

minimal impact on subsequent analyses. Mock community samples were removed from the

dataset prior to read clustering and analyzed separately. The sequencing error rate was 0.0027.

Sequences were then subsampled to 8,500 reads per sample to minimize the impacts of uneven

sequence coverage across samples, but retain as many samples within the dataset as possible.

Sequences were clustered into biologically meaningful groups (MED nodes) using Minimum

Entropy Decomposition (MED) (Eren et al. 2015). Sequences representing each MED node were

classified in mothur (Silva v119, method = 'knn') and this information, along with the read

counts and relative abundances, was used for microbial community analysis (Supporting

Information Methods). Raw sequences were deposited in the NCBI Sequence Read Archive

(SRA) under BioProject PRJNA517146.

Metagenomic sequencing

Total genomic DNA was extracted from nine reef-depth seawater samples from JR (n=4) and FK (n=5) using a modified cetyl-trimethylammonium bromide (CTAB) - phenol: chloroform: isoamyl alcohol extraction (Supporting Information Methods). No samples from CAN were chosen because the DNA extracted from CAN yielded 16S rDNA sequences that were highly variable between sites. Genomic libraries were prepared using the Hyper Library construction kit (Kapa Biosystems, Inc, Wilmington, MA, USA) and sequenced at the W. M. Keck Center using 2x150 bp paired-end Illumina HiSeq 4000 sequencing. Fastq files were demultiplexed, and library adaptors were trimmed from the 3' ends of the reads (Supporting Information Methods). BBTools (Bushnell 2016) was used to remove residual sequence adaptors (ktrim=r k=23 mink=11 hdist=1 tpe tbo) as well as trim reads using the Phred algorithm (qtrim=rl trimq=10). The program FMAP (Kim et al. 2016) was used to assign KEGG orthologs to the metagenomic reads using DIAMOND (Buchfink et al. 2015) as well as identify significantly different KEGG orthologs, KEGG pathways, and KEGG operons between JR and FK reef-depth seawater metagenomes (Kruskal – Wallis test; p-value<0.05, FDR adjusted to control for false positives). Raw files can also be found in SRA under bioProject PRJNA517146.

Community respiration measurements

Seawater samples (4 - 6 per site) were collected from reef-depth using a submersible groundwater pump and kept in the dark. Respiration rates incubations (~24 hours) were conducted with ~5X replication with water collected from 19 reefs in acid cleaned 60 ml glass Biological Oxygen Demand bottles with glass stoppers. Incubations using water from sites 2, 12, and 24 were conducted twice. Acid-washed bottles were equipped with oxygen optode 'dot' sensors (PreSens) affixed to the glass using food-grade silicone adhesive. In the laboratory, the perimeter of the glass stopper was filled with water using a squeeze bottle to prevent gas

exchange between the water in the bottle and the atmosphere. The concentration of oxygen in the bottles was measured over time using a handheld Fibox 4 (PreSens). Incubations were conducted in a static water bath in a darkened cooler located inside a darkened room at as close to *in situ* temperatures as possible in the remote location. The incubation temperatures were 26.6 ± 0.5 °C (JR), 25.0 ± 0.2 °C (CAN), and 26.5 ± 0.5 °C (FK). Initial oxygen measurements were taken every hour (h) for the first 4 h, and then approximately every 4 h after that. Ten oxygen measurement readings were taken for each incubation bottle at each timepoint. Prior to calculating respiration rates, oxygen data were quality controlled to remove any individual readings greater than one standard deviation from the mean value at a given timepoint. Linear fitting to the time course oxygen data was done in MATLAB (v. v7.13, MathWorks, Inc.) using the 'polyfit' function.

Statistical analyses

Due to the scope and breadth of this complex and nuanced dataset with sampling limitations, we implemented different statistical tests suitable for each dataset (e.g. inorganic nutrients, cell abundances, microbial community analysis) and tested for significance across different qualitative (e.g. subregion, reef-system) and quantitative (e.g. total organic carbon concentrations) parameters. To compare differences in reef cover, macronutrient concentrations, and cell abundances, data were inspected for normality using Shapiro-Wilk normality tests. Normally distributed data was tested using analysis of variance tests (ANOVA) followed by post-hoc Tukey multiple comparisons of means tests using a 95% family-wise confidence level (adjusted p value < 0.05). For data that was not normally distributed, Kruskal – Wallis rank sum tests, followed by either Dunn's or Conover-Iman tests using Bonferroni corrections were used to assess significant differences (adjusted p-value<0.05). We recognize that JR offshore and JR

233

gulf subregions have fewer data points compared to the other categories, but measurements from these locations were similar within each subregion and are likely representative of the environmental conditions. We performed linear regressions using ggplot2, geom_smooth and the method = "lm" (Wickham 2016) to investigate relationships between coral cover and picoplankton abundances, algal cover and total organic carbon concentrations, and unpigmented cell abundances and bacterial and archaeal observed richness. A principal components analysis (PCA) was conducted with biogeochemical, physicochemical, and microbial abundance data to assess collinearity between variables and to investigate which variables contributed to the most variation in both dimensions (Figure S6).

Amplicon-based microbial community statistical analyses were completed using R studio (R Core Development Team 2017). Reads identifying as chloroplasts (average 275 ± 198 reads per sample; 3% of all subsampled sequences) were removed from the dataset prior to beginning the analyses. Non-metric multidimensional scaling (NMDS) was conducted using the Bray-Curtis dissimilarity index and covariance matrices for each group were plotted as 95% confidence ellipses using 'vegan' (as in Eren et al. 2015; Oksanen et al. 2017). The 'vegan' package was also used to calculate the multivariate homogeneity of groups dispersions by subregion (function 'betadisper') and is defined as the average distance of group members to the group centroid (Oksanen et al. 2017). In addition, environmental vectors correlating maximally with each environmental variable were fit onto the NMDS ordination using the "envfit" function in 'vegan' (R^2 value indicates the scaled correlation coefficient). Reef-depth seawater collected from site FK 23 were omitted from the analysis because reef-depth TOC was not collected from this site. Nested PERMANOVA (Adonis) tests, also available within 'vegan', were conducted using the Bray-Curtis dissimilarity index (999 permutations) ($p<0.05$). To conduct this test, the

factors reef-depth, reef location, and subregion were nested within region (reef-system). The package 'phyloseq' (McMurdie and Holmes 2013) was used to calculate alpha richness. DESeq2 was used to identify significantly differently enriched MED nodes between JR and FK reef-depth seawater using default parameters with a "local" fit trend line (Love et al. 2014). This procedure is able to identify significantly differentially rich taxa even if they are at low relative abundances and is useful for investigating MED-specific differences in cryptic members of the community. Samples collected in CAN were not subjected to DESeq2 analysis due to the lower number of samples.

5.7 Acknowledgements

5.8 References

Apprill, A., and Rappé, M.S. (2011) Response of the microbial community to coral spawning in lagoon and reef flat environments of Hawaii, USA. *Aquat Microb Ecol* **62**: 251-266.

Apprill, A., Weber, L.G., and Santoro, A.E. (2016) Distinguishing between microbial habitats unravels ecological complexity in coral microbiomes. *mSystems* **1**: e00143-00116.

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.

Ault, J.S., Bohnsack, J.A., and Meester, G.A. (1998) A retrospective (1979 - 1996) multispecies assessment of coral reef fish stocks in the Florida Keys. *Fish Bull* **96**: 395-414.

Ault, J.S., Smith, S.G., Bohnsack, J.A., Luo, J., Harper, D.E., and McClellan, D.B. (2006) Building sustainable fisheries in Florida's coral reef ecosystem: positive signs in the Dry Tortugas. *B Mar Sci* **78**: 633-654.

Baisre, J. (2017) An overview of Cuban commercial marine fisheries: the last 80 years. *B Mar Sci*.

Barott, K.L., and Rohwer, F.L. (2012) Unseen players shape benthic competition on coral reefs. *Trends Microbiol* **20**: 621-628.

Bergmann, G.T., Bates, S.T., Eilers, K.G., Lauber, C.L., Caporaso, J.G., Walters, W.A. *et al.* (2011) The under-recognized dominance of Verrucomicrobia in soil bacterial communities. *Soil Biol Biochem* **43**: 1450-1455.

Bertilsson, S., Berglund, O., Pullin, M.J., and Chisholm, S.W. (2005) Release of dissolved organic matter by *Prochlorococcus*. *Vie Et Milieu* **55**: 225-231.

Briceno, H.O., and Boyer, J.N. (2015) 2014 annual report of the water quality monitoring project for the water quality protection program of the Florida Keys National Marine Sanctuary. In.

Bruce, T., Meirelles, P.M., Garcia, G., Paranhos, R., Rezende, C.E., de Moura, R.L. *et al.* (2012) Abrolhos bank reef health evaluated by means of water quality, microbial diversity, benthic cover, and fish biomass data. *PLoS One* **7**: e36687.

Buchfink, B., Xie, C., and Huson, D.H. (2015) Fast and sensitive protein alignment using DIAMOND. *Nat Methods* **12**: 59-60.

Bushnell, B. (2016) BBMap short read aligner.

Caballero Aragón, H., Armenteros, M., Perera Valderrama, S., Rey Villiers, N., Cobián Rojas, D., Campos Verdecia, K., and Alcolado Menéndez, P.M. (2019) Ecological condition of coral reef assemblages in the Cuban Archipelago. *Marine Biology Research*: 1-13.

Carpenter, R.C. (1988) Mass mortality of a Caribbean sea urchin: immediate effects on community metabolism and other herbivores. *P Natl Acad Sci USA* **85**: 511-514.

Cermak, N., Becker, J.W., Knudsen, S.M., Chisholm, S.W., Manalis, S.R., and Polz, M.F. (2017) Direct single-cell biomass estimates for marine bacteria via Archimedes' principle. *ISME J* **11**: 825-828.

Charpy, L., Casareto, B.E., Langlade, M.J., and Suzuki, Y. (2012) Cyanobacteria in Coral Reef Ecosystems: A Review. *Journal of Marine Biology* **2012**: 1-9.

Cox, E.F., Ribes, M., and Kinzie III, R.A. (2006) Temporal and spatial scaling of planktonic responses to nutrient inputs into a subtropical embayment. *Mar Ecol Prog Ser* **324**: 19-35.

Dinsdale, E.A., Pantos, O., Smriga, S., Edwards, R.A., Angly, F., Wegley, L. *et al.* (2008) Microbial ecology of four coral atolls in the Northern Line Islands. *PLoS One* **3**: e1584.

Dore, J.E., Letelier, R.M., Church, M.J., Lukas, R., and Karl, D.M. (2008) Summer phytoplankton blooms in the oligotrophic North Pacific Subtropical Gyre: Historical perspective and recent observations. *Prog Oceanogr* **76**: 2-38.

DuRand, M.D., Olson, R.J., and Chisholm, S.W. (2001) Phytoplankton population dynamics at the Bermuda Atlantic Time-series station in the Sargasso Sea. *Deep-Sea Res Pt II* **48**: 1983-2003.

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

Eren, A.M., Morrison, H.G., Lescault, P.J., Reveillaud, J., Vineis, J.H., and Sogin, M.L. (2015) Minimum entropy decomposition: unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences. *ISME J* **9**: 968-979.

Fernandez-Gomez, B., Richter, M., Schuler, M., Pinhassi, J., Acinas, S.G., Gonzalez, J.M., and Pedros-Alio, C. (2013) Ecology of marine *Bacteroidetes*: a comparative genomics approach. *ISME J* **7**: 1026-1037.

Ferrer, V.M., González-Díaz, S.P., Hernández-Fernández, L., Siciliano, D., Bretos, F., Apprill, A. *et al.* (2016) Salud de las comunidades de corales en arrecifes de Jardines de la Reina - Golfo de Ana Maria, region sur-central de Cuba. *Revista Investigaciones Marinas* **36**: 34-53.

Ferrier-Pages, C., and Gattuso, J.P. (1998) Biomass, production and grazing rates of pico- and nanoplankton in coral reef waters (Miyako Island, Japan). *Microb Ecol* **35**: 46-57.

Freitas, S., Hatosy, S., Fuhrman, J.A., Huse, S.M., Welch, D.B., Sogin, M.L., and Martiny, A.C. (2012) Global distribution and diversity of marine *Verrucomicrobia*. *ISME J* **6**: 1499-1505.

Fukuda, R., Ogawa, H., Nagata, T., and Koike, I. (1998) Direct determination of carbon and nitrogen contents of natural bacterial assemblages in marine environments. *Appl Environ Microbiol* **64**: 3352-3358.

237

Galford, G.L., Fernandez, M., Roman, J., Monasterolo, I., Ahamed, S., Fiske, G. *et al.* (2018) Cuban land use and conservation, from rainforests to coral reefs. *B Mar Sci* **94**: 000–000. 2018.

Gardner, T.A., Cote, I.M., Gill, J.A., Grant, A., and Watkinson, A.R. (2003) Long - term region wide declines in Caribbean corals. *Science* **301**: 958-960.

Glasl, B., Webster, N.S., and Bourne, D.G. (2017) Microbial indicators as a diagnostic tool for assessing water quality and climate stress in coral reef ecosystems. *Marine Biology* **164**.

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**: 94.

González-Díaz, P., González-Sansón, G., Aguilar Betancourt, C., Álvarez Fernández, S., Perera Pérez, O., Hernández Fernández, L. *et al.* (2018) Status of Cuban coral reefs. *B Mar Sci.*

Guannel, G., Arkema, K., Ruggiero, P., and Verutes, G. (2016) The Power of Three: Coral Reefs, Seagrasses and Mangroves Protect Coastal Regions and Increase Their Resilience. *PLoS One* **11**: e0158094.

Haas, A.F., Fairoz, M.F., Kelly, L.W., Nelson, C.E., Dinsdale, E.A., Edwards, R.A. *et al.* (2016) Global microbialization of coral reefs. *Nat Microbiol* **1**: 16042.

Haas, A.S., and Wild, C. (2010) Composition analysis of organic matter released by cosmopolitan coral reef-associated green algae. *Aquatic Biology* **10**: 131-138.

Hansell, D.A., and Carlson, C.A. (2001) Biogeochemistry of total organic carbon and nitrogen in the Sargasso Sea: control by convective overturn. *Deep-Sea Res II* **48**: 1649-1667.

Hughes, T.P., Rodrigues, M.J., Bellwood, D.R., Ceccarelli, D., Hoegh-Guldberg, O., McCook, L. *et al.* (2007) Phase shifts, herbivory, and the resilience of coral reefs to climate change. *Curr Biol* **17**: 360-365.

Jackson, J.B.C., Kirby, M.X., Berger, W.H., Bjorndal, K.A., Botsford, L.W., Bourque, B.J. *et al.* (2001) Historical Overfishing and the Recent Collapse of Coastal Ecosystems. *Science* **293**: 629-638.

Karl, D.M., Christian, J.R., Dore, J.E., Hebel, D.V., Letelier, R.M., Tupas, L.M., and Winn, C.D. (1996) Seasonal and interannual variability in primary production and particle flux at Station ALOHA. *Deep-Sea Res Pt II* **43**: 539-568.

Kearns, P.J., Angell, J.H., Howard, E.M., Deegan, L.A., Stanley, R.H.R., and Bowen, J.L. (2016) Nutrient enrichment induces dormancy and decreases diversity of active bacteria in salt marsh sediments. *Nat Commun* **7**.

Kelly, L.W., Barott, K.L., Dinsdale, E., Friedlander, A.M., Nosrat, B., Obura, D. *et al.* (2012) Black reefs: iron-induced phase shifts on coral reefs. *ISME J* **6**: 638-649.

Kelly, L.W., Williams, G.J., Barott, K.L., Carlson, C.A., Dinsdale, E.A., Edwards, R.A. *et al.* (2014) Local genomic adaptation of coral reef-associated microbiomes to gradients of natural variability and anthropogenic stressors. *Proc Natl Acad Sci U S A* **111**: 10227-10232.

Kim, J., Kim, M.S., Koh, A.Y., Xie, Y., and Zhan, X. (2016) FMAP: Functional Mapping and Analysis Pipeline for metagenomics and metatranscriptomics studies. *BMC Bioinformatics* **17**: 420.

Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K., and Schloss, P.D. (2013) Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol* **79**: 5112-5120.

Lapointe, B.E., and Clark, M.W. (1992) Nutrient inputs from the watershed and coastal eutrophication in the Florida Keys. *Estuaries* **15**: 465-476.

Lapointe, B.E., Barile, P.J., and Matzie, W.R. (2004) Anthropogenic nutrient enrichment of seagrass and coral reef communities in the Lower Florida Keys: discrimination of local versus regional nitrogen sources. *J Exp Mar Biol Ecol* **308**: 23-58.

Leeworthy, V.R., Loomis, D.K., and Paterson, S.K. (2010) Linking the economy and the environment of Florida Keys/Key West. In *Visitor profiles: Florida Keys/ Key West 2007-08*: National Oceanic and Atmospheric Administration.

Leichter, J.J., Stewart, H.L., and Miller, S.L. (2003) Episodic nutrient transport to Florida coral reefs. *Limnol Oceanogr* **48**: 1394-1407.

Lewis, J.B. (1977) Processes of organic production on coral reefs. *Biological Reviews* **52**: 305-347.

Louca, S., Polz, M.F., Mazel, F., Albright, M.B.N., Huber, J.A., O'Connor, M.I. *et al.* (2018) Function and functional redundancy in microbial systems. *Nat Ecol Evol* **2**: 936-943.

Love, M.I., Huber, W., and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**.

Mackey, M.D., Mackey, D.J., Higgins, H.W., and Wright, S.W. (1996) CHEMTAX - A program for estimating class abundances from chemical markers: Application to HPLC measurements of phytoplankton. *Mar Ecol Prog Ser* **144**: 265-283.

Marie, D., Partensky, F., Jacquet, S., and Vaulot, D. (1997) Enumeration and cell cycle analysis of natural populations of marine picoplankton by flow cytometry using the nucleic acid stain SYBR Green I. *Appl Environ Microb* **63**: 186-193.

Martiny, A.C., Tai, A.P., Veneziano, D., Primeau, F., and Chisholm, S.W. (2009) Taxonomic resolution, ecotypes and the biogeography of Prochlorococcus. *Environ Microbiol* **11**: 823-832.

McDole Somera, T., Bailey, B., Barott, K., Grasis, J., Hatay, M., Hilton, B.J. *et al.* (2016) Energetic differences between bacterioplankton trophic groups and coral reef resistance. *Proc Biol Sci* **283**.

McDole, T., Nulton, J., Barott, K.L., Felts, B., Hand, C., Hatay, M. *et al.* (2012) Assessing coral reefs on a Pacific-wide scale using the microbialization score. *PLoS One* **7**: e43233.

McMurdie, P.J., and Holmes, S. (2013) phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* **8**: e61217.

McNally, S.P., Parsons, R.J., Santoro, A.E., and Apprill, A. (2017) Multifaceted impacts of the stony coral *Porites astreoides* on picoplankton abundance and community composition. *Limnol Oceanogr* **62**: 217-234.

Miller, J., Muller, E., Rogers, C., Waara, R., Atkinson, A., Whelan, K.R.T. *et al.* (2009) Coral disease following massive bleaching in 2005 causes 60% decline in coral cover on reefs in the US Virgin Islands. *Coral Reefs* **28**: 925-937.

Moreira, A.P., Meirelles, P.M., Santos Ede, O., Amado-Filho, G.M., Francini-Filho, R.B., Thompson, C.C., and Thompson, F.L. (2015) Turbulence-driven shifts in holobionts and planktonic microbial assemblages in St. Peter and St. Paul Archipelago, Mid-Atlantic Ridge, Brazil. *Front Microbiol* **6**: 1038.

Mumby, P.J., Edwards, A.J., Arias - Gonzalez, J.E., Lindenman, K.C., Blackwell, P.G., Gall, A. *et al.* (2004) Mangroves enhance the biomass of coral reef fish communities in the Caribbean. *Nature* **427**: 533-536.

Nelson, C.E., Alldredge, A.L., McCliment, E.A., Amaral-Zettler, L.A., and Carlson, C.A. (2011) Depleted dissolved organic carbon and distinct bacterial communities in the water column of a rapid-flushing coral reef ecosystem. *ISME J* **5**: 1374-1387.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D. *et al.* (2017) vegan: community ecology package. R package version 2.4-5.

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.

Patten, N.L., Wyatt, A.S.J., Lowe, R.J., and Waite, A.M. (2011) Uptake of picophytoplankton, bacterioplankton and virioplankton by a fringing coral reef community (Ningaloo Reef, Australia). *Coral Reefs* **30**: 555-567.

Pina-Amargos, F., Gonzalez-Sanson, G., Martin-Blanco, F., and Valdivia, A. (2014) Evidence for protection of targeted reef fish on the largest marine reserve in the Caribbean. *PeerJ* **2**: e274.

Pinhassi, J., Sala, M.M., Havskum, H., Peters, F., Guadayol, O., Malits, A., and Marrase, C. (2004) Changes in bacterioplankton composition under different phytoplankton regimens. *Appl Environ Microbiol* **70**: 6753-6766.

Pizarro, V., Rodriguez, S.C., Lopez-Victoria, M., Zapata, F.A., Zea, S., Galindo-Martinez, C.T. et al. (2017) Unraveling the structure and composition of Varadero Reef, an improbable and imperiled coral reef in the Colombian Caribbean. *PeerJ* **5**: e4119.

Precht, W.F., Gintert, B.E., Robbart, M.L., Fura, R., and van Woesik, R. (2016) Unprecedented Disease-Related Coral Mortality in Southeastern Florida. *Sci Rep* **6**: 31374.

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

R Core Development Team (2017) R: A language and environment for statistical computing. In. Vienna, Austria: R Foundation for Statistical Computing.

Reese, A.T., and Dunn, R.R. (2018) Drivers of Microbiome Biodiversity: A Review of General Rules, Feces, and Ignorance. *MBio* **9**.

Rocap, G., Larimer, F.W., Lamerdin, J., Malfatti, S., Chain, P., Ahlgren, N.A. et al. (2003) Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* **424**: 1042-1047.

Santoro, A.E., Casciotti, K.L., and Francis, C.A. (2010) Activity, abundance and diversity of nitrifying archaea and bacteria in the central California Current. *Environ Microbiol* **12**: 1989-2006.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B. et al. (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75**: 7537-7541.

Smith, J.E., Brainard, R., Carter, A., Grillo, S., Edwards, C., Harris, J. et al. (2016) Re-evaluating the health of coral reef communities: baselines and evidence for human impacts across the central Pacific. *Proc Biol Sci* **283**.

Soetaert, K., Van den Meersche, K., and van Oevelen, D. (2009) limSolve: Solving Linear Inverse Models. R-package version 1.5.5.3. In.

Sorokin, Y.I. (1995) Role of plankton in the turnover of organic matter on the Great Barrier Reef, Australia. *Hydrobiologia* **308**: 35-44.

Sorokin, Y.I. (1995) Coral Reef Ecology: Springer-Verlag.

Stuhldreier, I., Sanchez-Noguera, C., Rixen, T., Cortes, J., Morales, A., and Wild, C. (2015) Effects of Seasonal Upwelling on Inorganic and Organic Matter Dynamics in the Water Column of Eastern Pacific Coral Reefs. *PLoS One* **10**: e0142681.

Szmant, A.M., and Forrester, A. (1996) Water column and sediment nitrogen and phosphorus distribution patterns in the Florida Keys, USA. *Coral Reefs* **15**: 21-41.

Teeling, H., Fuchs, B.M., Becher, D., Klockow, C., Gardebrecht, A., Bennke, C.M. *et al.* (2012) Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. *Science* **336**: 608-611.

Urakawa, H., Martens-Habbena, W., and Stahl, D.A. (2010) High abundance of ammonia-oxidizing Archaea in coastal waters, determined using a modified DNA extraction method. *Appl Environ Microbiol* **76**: 2129-2135.

Valdivia, A., Cox, C.E., and Bruno, J.F. (2017) Predatory fish depletion and recovery potential on Caribbean reefs. *Science Advances* **3**: e1601303.

Van Heukelem, L., and Thomas, C.S. (2001) Computer-assisted high-performance liquid chromatography method development with applications to the isolation and analysis of phytoplankton pigments. *J Chromatogr A* **910**: 31-49.

Westrum, B.L., and Meyers, P.A. (1978) Organic carbon content of seawater from over three Caribbean reefs. *B Mar Sci* **28**: 153-158.

Wickham, H. (2016) ggplot2: elegant graphics for data analysis. New York: Springer-Verlag.

Wright, E.S., and Vetsigian, K.H. (2016) Quality filtering of Illumina index reads mitigates sample cross-talk. *BMC Genomics* **17**: 876.

Yeo, S.K., Huggett, M.J., Eiler, A., and Rappe, M.S. (2013) Coastal bacterioplankton community dynamics in response to a natural disturbance. *PLoS One* **8**: e56207.

Zinser, E.R., Coe, A., Johnson, Z.I., Martiny, A.C., Fuller, N.J., Scanlan, D.J., and Chisholm, S.W. (2006) *Prochlorococcus* ecotype abundances in the North Atlantic Ocean as revealed by an improved quantitative PCR method. *Appl Environ Microbiol* **72**: 723-732.

5.9 Tables

Table 1. PERMANOVA (Analysis of variance using distance matrices, ADONIS) comparisons based on Bray – Curtis dissimilarities of seawater bacterial and archaeal SSU rRNA amplicons grouped into MED nodes across collected across reef-systems.

Factors^	Df*	Sums of squares	Mean squares	Pseudo F statistic	$R^{2\&}$	p-value (perm.)
Region	2	2.3619	1.18093	12.5082	0.18028	0.001
Water-type (Region)	2	2.3047	1.15236	12.2057	0.17592	0.001
Reef (Water-type (Region))	18	4.7583	0.26435	2.7999	0.36320	0.001
Depth (Reef (Water-type (Region)))	9	1.3157	0.14619	1.5485	0.10043	0.039
Residuals	25	2.3603	0.09441		0.18016	
Total	56	13.1009			1.000	

^Nested comparisons are denoted by parentheses; e.g. Reef (Water-type (Region)) indicates that the factor 'Reef' is nested within the factor 'Water-type' that is nested within the factor 'Region'.
*Df = degrees of freedom
$^{\&}R^2$ = percentage of variation explained by each factor
perm. = 999 permutations

Table 2. Enriched KEGG metabolic modules and pathways of seawater microbial communities in Jardines de la Reina, Cuba (JR, shaded in grey) and Florida Keys, USA (FK).

KEGG module definition	Orthology count	Reef-system of enrichment	Coverage*	p-value
Capsular polysaccharide transport system (M00249)	2	JR	0.6667	0.0051
NAD(P)H: quinone oxidoreductase, chloroplasts and cyanobacteria (M00145)	3	JR	0.2143	0.0189
Assimilatory nitrate reduction, nitrate => ammonia (M00531)	2	JR	0.3333	0.0236
Nitrate assimilation (M00615)	1	JR	1.0000	0.0421
Fumarate reductase, prokaryotes (M00150)	2	FK	0.5000	0.0100
Benzene degradation, benzene => catechol (M00548)	2	FK	0.3333	0.0236

KEGG pathway definition				
Photosynthesis (map00195)	5	JR	0.0794	0.0019
Nitrogen metabolism (map00910)	4	JR	0.0667	0.0098
Oxidative phosphorylation (map00190)	7	JR (5) FK (2)	0.0326	0.0316
Pentose and glucuronate interconversions (map00040)	4	FK	0.0571	0.0166
Lipopolysaccharide biosynthesis (map00540)	3	FK	0.0750	0.0182
Toluene degradation (map00623)	3	FK	0.0652	0.0264
Valine, leucine and isoleucine biosynthesis (map00290)	2	FK	0.1053	0.0285
Fructose and mannose metabolism (map00051)	7	FK (6) JR (1)	0.0654	0.0007
Drug metabolism - other enzymes (map00983)	2	FK (1) JR (1)	0.0909	0.0375
Pyruvate metabolism (map00620)	4	FK (3) JR (1)	0.0417	0.0457
Microbial metabolism in diverse environments (map01120)	22	FK (16) JR (6)	0.0204	0.0474

*coverage indicates the normalized coverage of genes in either KEGG pathways or modules
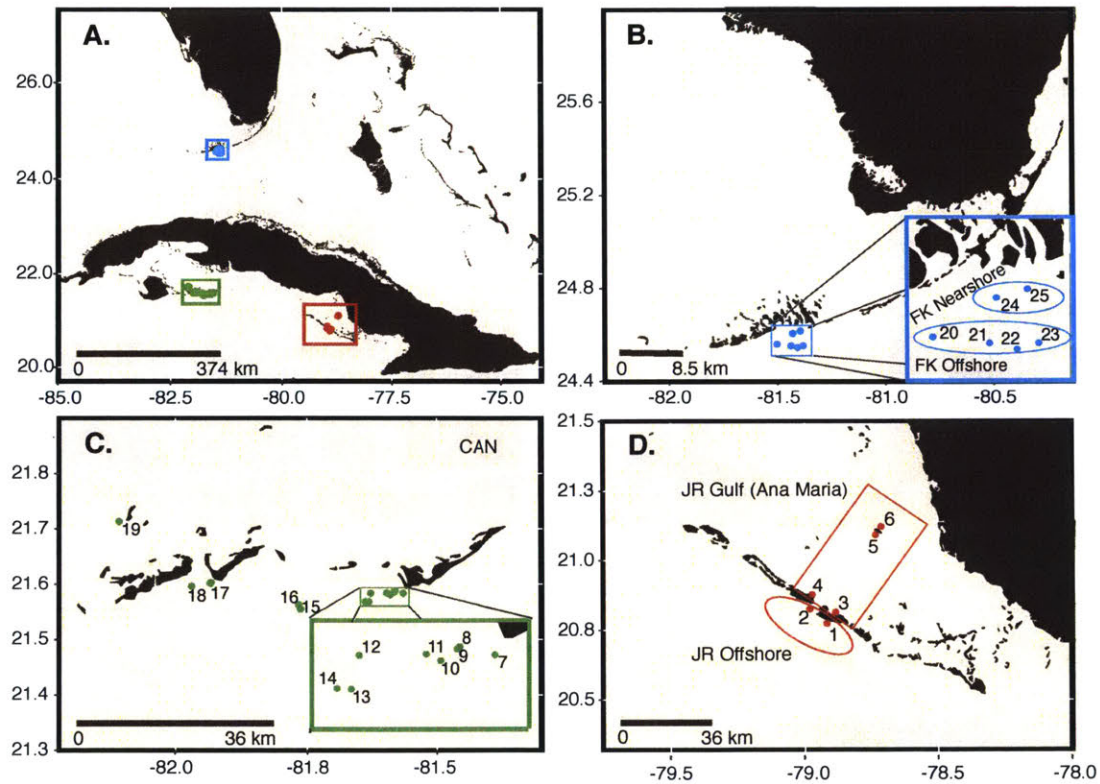
Figure 1. Overview map of the reef-systems (A) and reef locations studied: B) Florida Keys (FK), Florida, USA, C) Los Canarreos (CAN), Cuba, and D) Jardines de la Reina (JR), Cuba. Outlined shapes in B) and D) delineate the two subregion reef groupings within the Florida Keys (nearshore and offshore) and Jardines de la Reina (offshore and gulf).
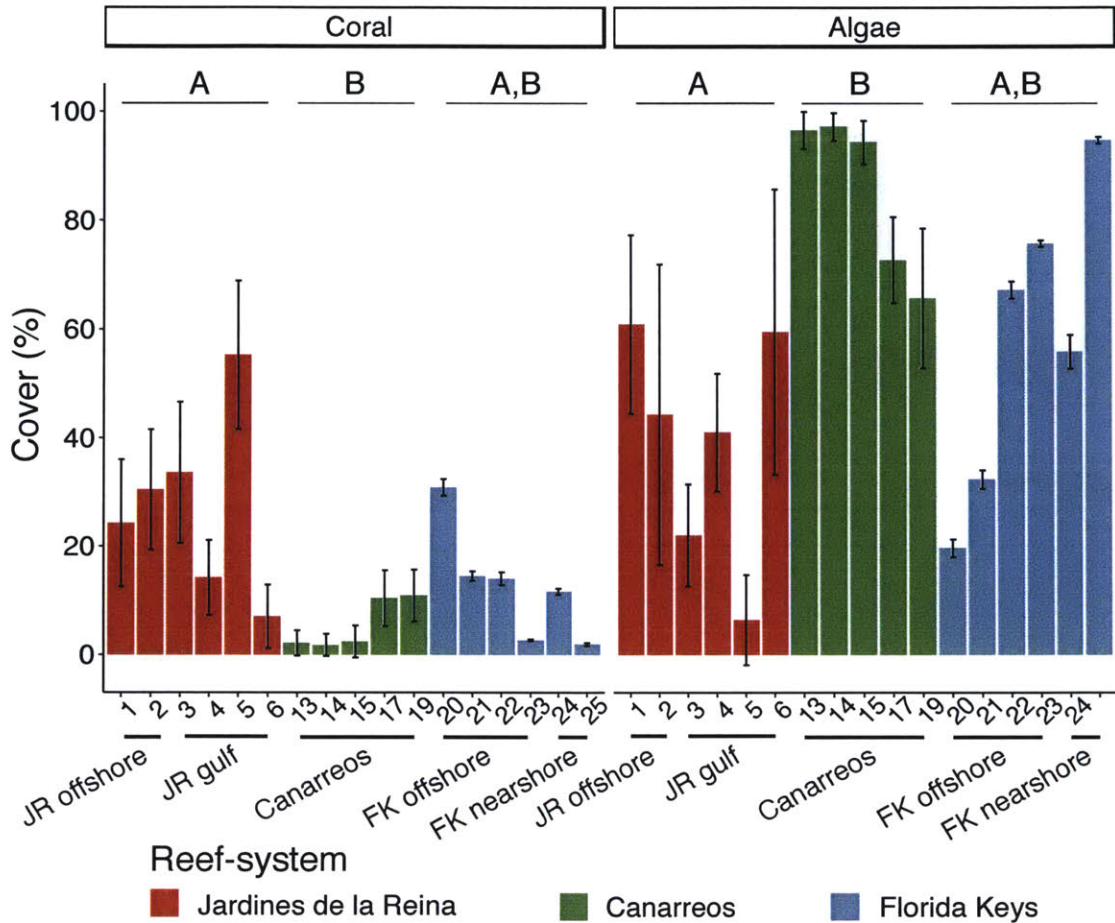
Figure 2. Comparison of mean coral and algal cover of the reefs indicates that Jardines de la Reina (JR) and the Florida Keys (FK) have similar coral and algal cover, whereas the Los Canarreos reef-system has significantly lower coral cover and higher algal cover. Reef-systems with different letters are significantly different from each other (Kruskal-Wallis rank sum test, Dunn's test with Bonferroni corrections; ANOVA, Tukey's multiple comparisons of means test; $p < 0.05$) and are indicated by the letters A and B. Error bars reflect standard deviation in percent cover.
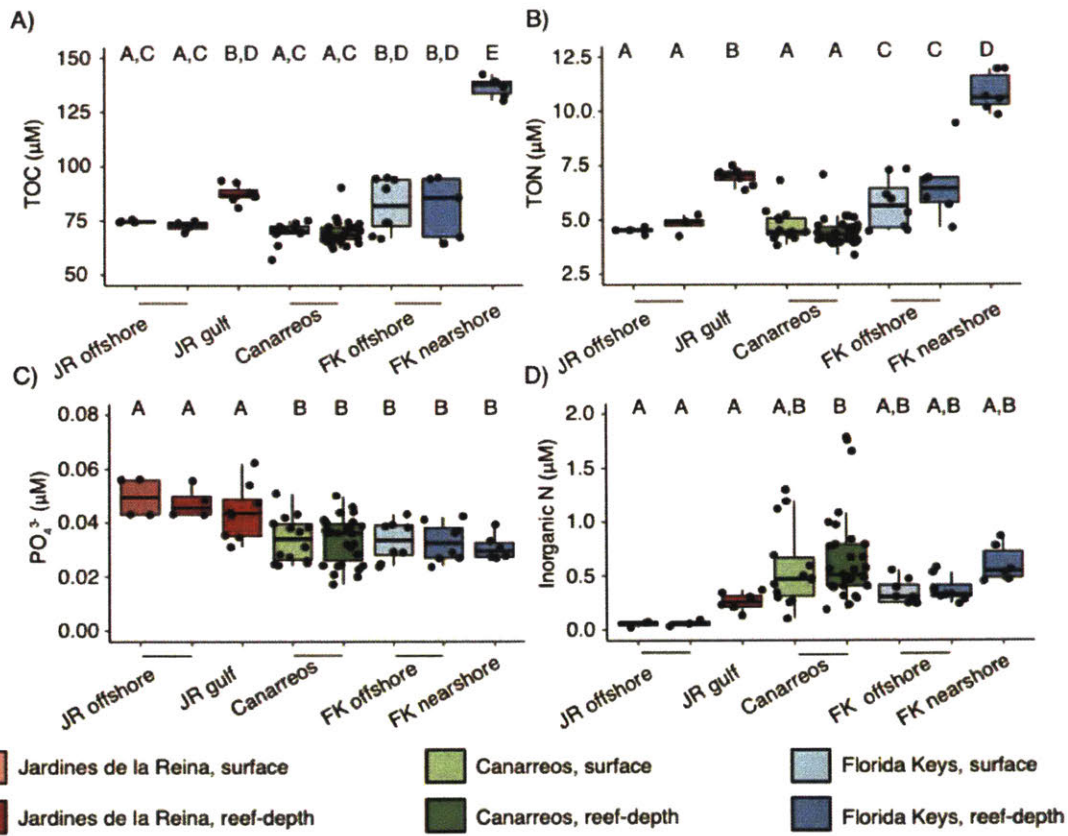
Figure 3. Comparison of seawater organic matter and nutrient concentrations across the reefs shows differential concentrations between the reef-systems. Measurements include A) total organic carbon (TOC), B) total organic nitrogen (TON), C), phosphate (PO_4^{3-}) and D) inorganic nitrogen ($NO_2^- + NO_3^- + NH_4^+$). Lower and upper edges of the boxplot correspond to the first and third quartiles, the whiskers extend to the largest or smallest value at 1.5 times the interquartile, and the black bar across the box represents the median. Boxplots with different letters denote concentrations that are significantly different from each other (ANOVA, Tukey's HSD test, $p < 0.05$) within each plot.

247
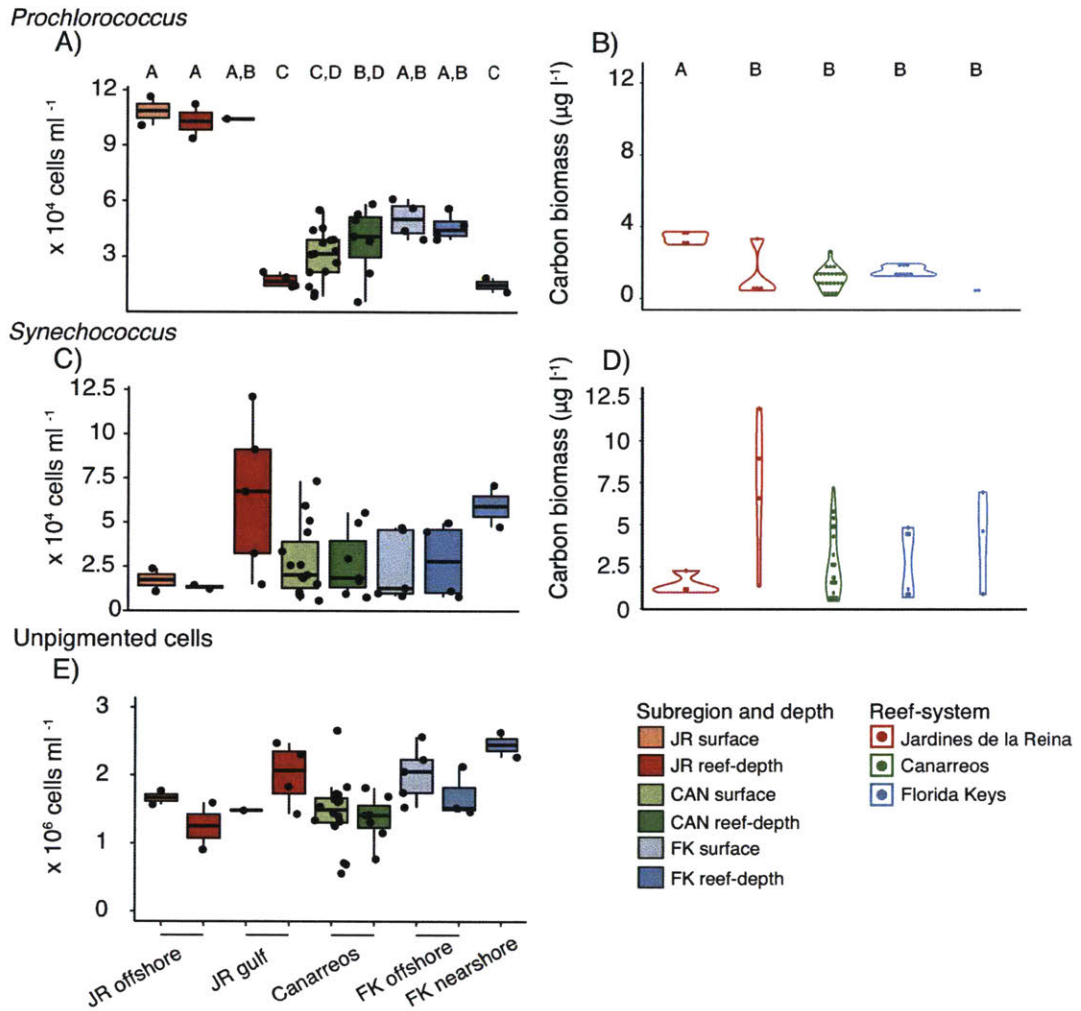
Figure 4. Differential microbial abundances and their carbon contributions in seawater across different reef subregions for A, B) *Prochlorococcus* and C, D) *Synechococcus*. Unpigmented cell abundance (largely heterotrophic bacteria and archaea) are shown in E). Lower and upper edges of the boxplot correspond to the first and third quartiles, the whiskers extend to the largest or smallest value at 1.5 times the interquartile, and the black bar across the box represents the median. Boxplots and violin plots with different letters are significantly different from each other (Kruskal-Wallis Rank Sum test and Conover-Iman or Dunn's tests, $p < 0.05$).
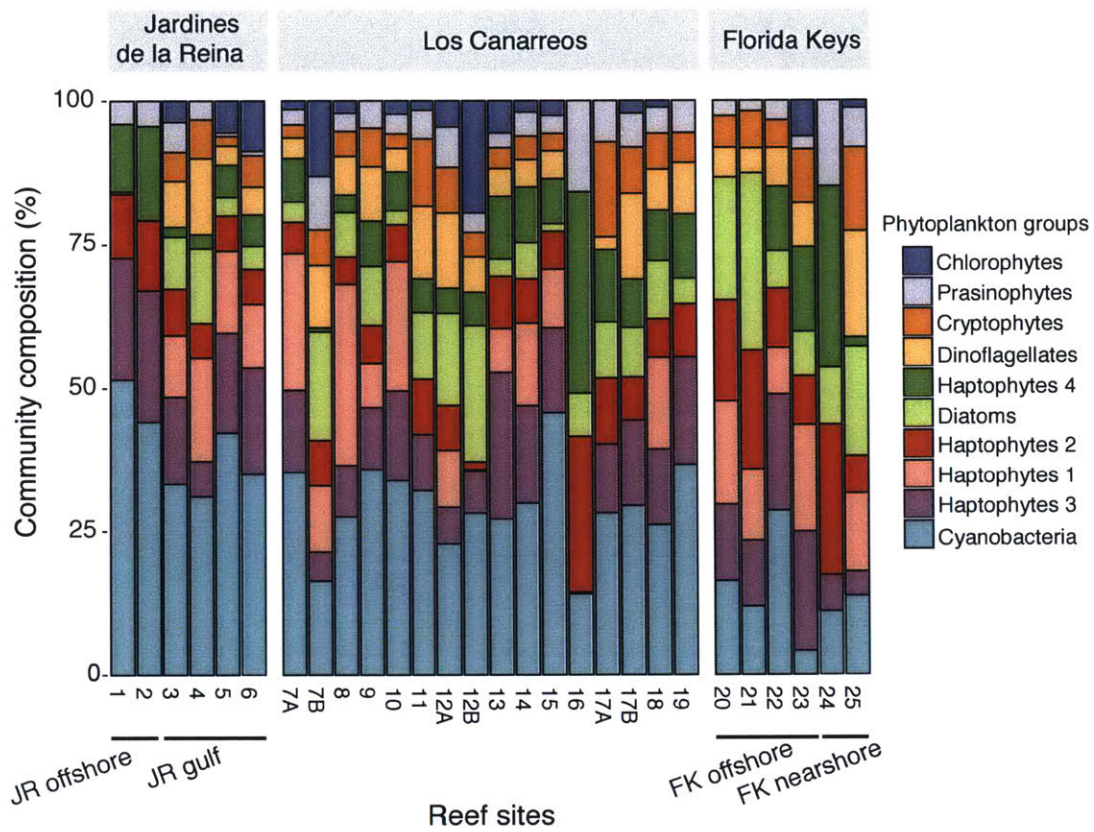
Figure 5. Relative abundance of phytoplankton groups across the reefs as determined by pigment analysis. Site numbers with either an A or B were sampled twice, but on different days.
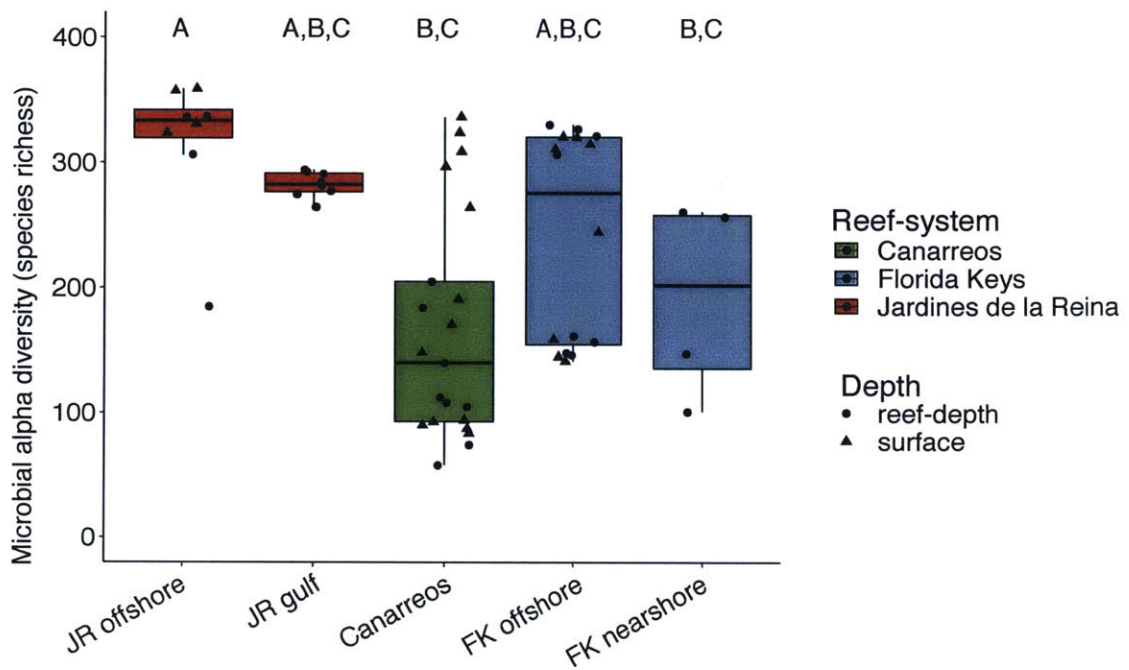
Figure 6. Microbial species richness (alpha diversity), as indicated by the number of bacterial and archaeal SSU rRNA gene amplicons grouped into minimum entropy decomposition (MED) nodes, is significantly highest in the Jardines de la Reina offshore reefs compared to the other reefs. Lower and upper edges of the boxplot correspond to the first and third quartiles, the whiskers extend to the largest or smallest value at 1.5 times the interquartile, and the black bar across the box represents the median. Boxplots with different letters are significantly different from each other (Kruskal-Wallis rank sum test, Dunn's test using Bonferroni corrections $p < 0.05$).
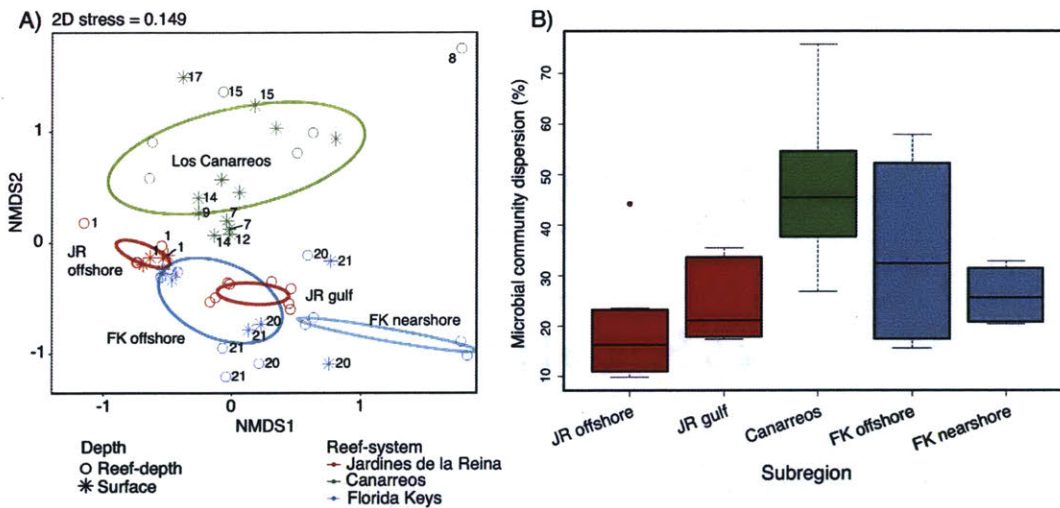
Figure 7. Bacterial and archaeal community beta diversity and Bray-Curtis microbial community dispersion between the reefs. A) Non-metric multidimensional scaling analysis (NMDS) of reef seawater bacterial and archaeal SSU rRNA gene amplicons grouped into MED nodes and compared using Bray-Curtis dissimilarity. Confidence ellipses are drawn using the subregion category (using the covariance matrix), the region is indicated by color, and depth and region information for each sample is indicated by the symbol. Samples that are located outside of the covariance ellipse are labeled with the reef number where they were collected. B) Boxplots of the microbial community dispersion within each reef grouping. Lower and upper edges of the boxplot correspond to the first and third quartiles, the whiskers extend to the largest or smallest value at 1.5 times the interquartile, and the black bar across the box represents the median.

251
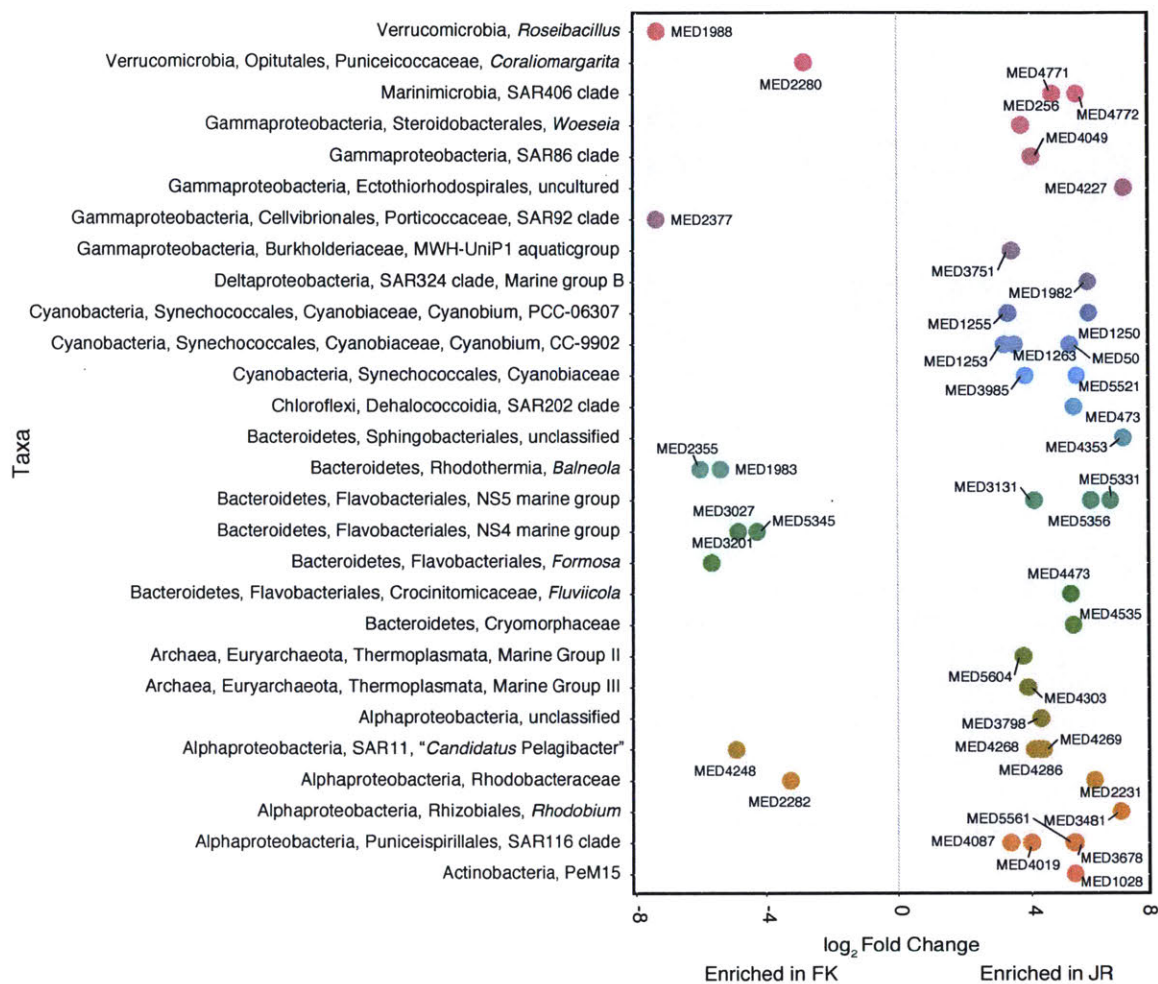
Figure 8. Log$_2$ fold change of the 44 significantly enriched MED nodes detected within Jardines de la Reina (JR) compared to the Florida Keys (FK) reef-depth seawater. Taxa are labeled using the highest resolved level of taxonomic annotation and colors indicate phyla or subphyla membership. Taxa identified to genus are plotted.

252

253

Figure 9. Comparison of the reef-depth seawater metagenomes showed 163 significantly different KEGG orthologs (KOs) between Jardines de la Reina and the Florida Keys. KO abundances were scaled using the 10th and 90th quantiles of the data for visualization. The dendrogram reflects hierarchical clustering of the samples using the 'hclust' function in R.
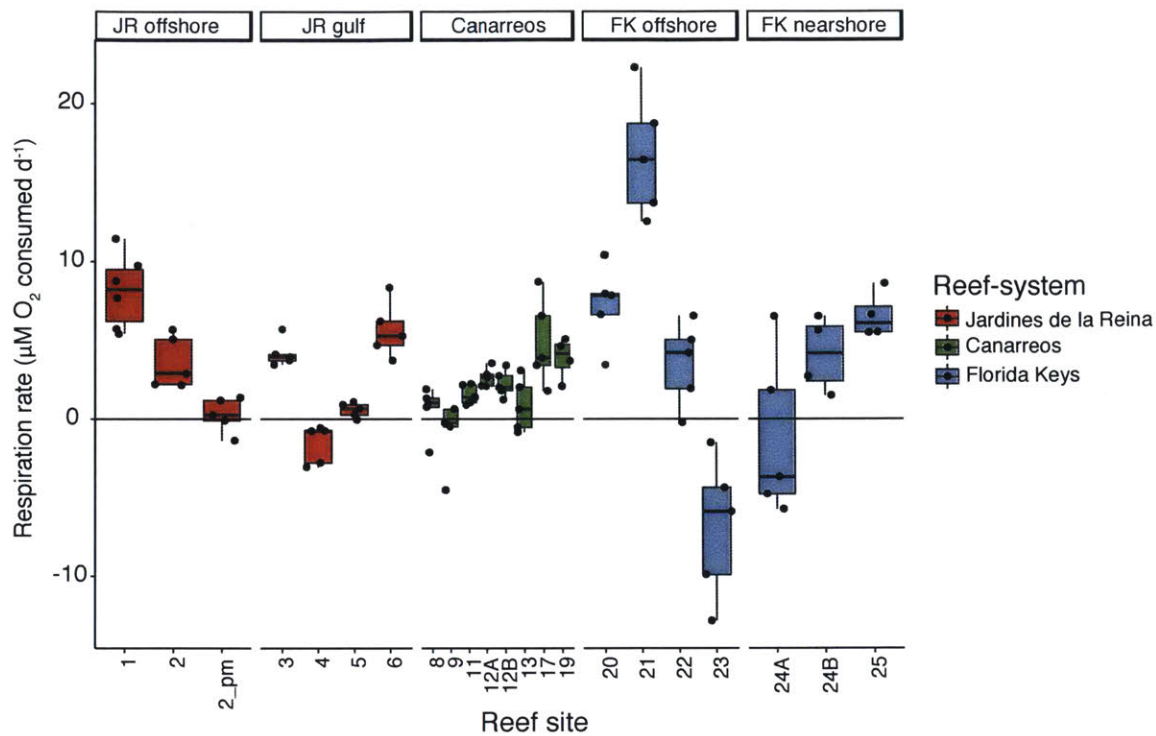
Figure 10. Comparison of net community respiration rates for reef-depth seawater across reef-systems. Lower and upper edges of the boxplot correspond to the first and third quartiles, the whiskers extend to the largest or smallest value at 1.5 times the interquartile, and the black bar across the box represents the median. Reef sites with either an A or B after the number were sampled twice, but on different days. All incubations were completed with reef seawater collected in the morning with the exception of one incubation at site 2 that was collected in the afternoon (labeled as '2_pm').
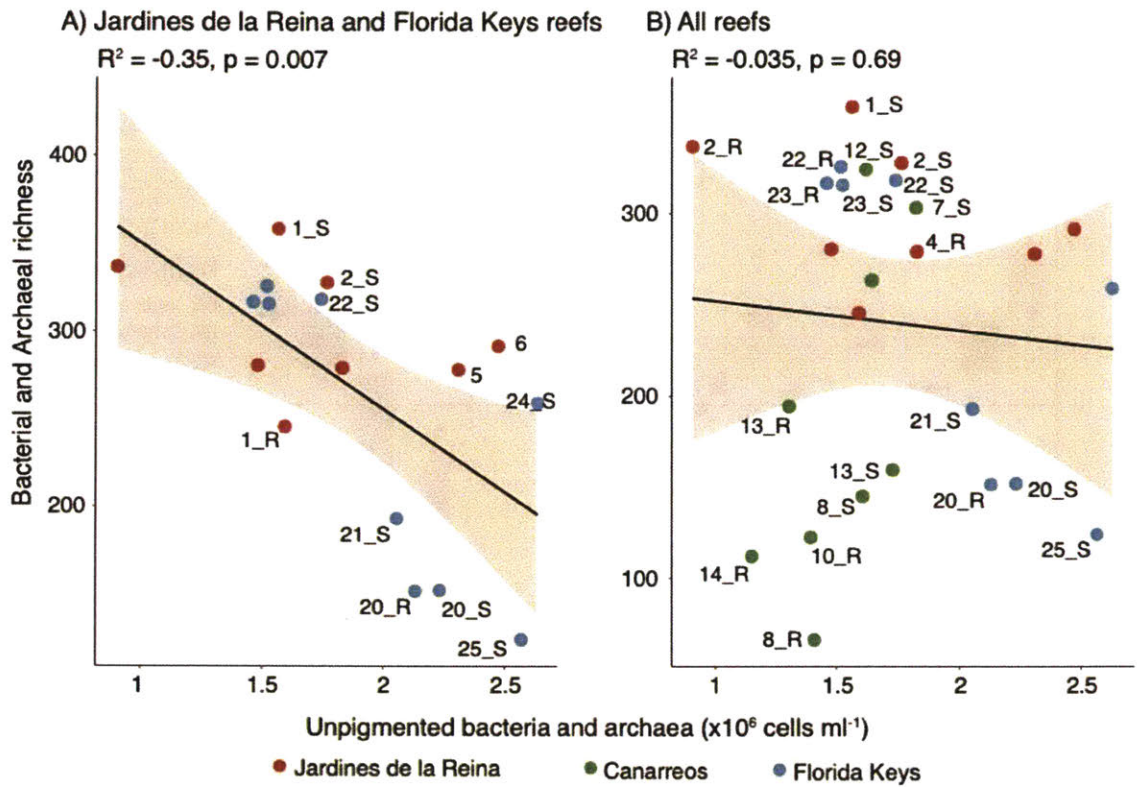
Figure 11. Negative regressions between bacterial and archaeal richness and abundance of unpigmented cells across the three reef-systems. Symbols that lie outside the line are labeled with the site number and depth. S denotes surface and R denotes reef-depth samples. A) includes only Jardines de la Reina and Florida Keys samples. B) includes samples from all reef-systems.

5.11 Supporting Information

Supporting Methods

Reef surveys
Scuba divers conducted reef surveys at reefs within Jardines de la Reina, Los Canarreos, and the Florida Keys (Supporting Information Table S1). At all Jardines de la Reina and five Los Canarreos reefs, divers estimated the percent cover of dominant reef biotypes (macroalgae, coral, sponge, and sand) by recording the distance (cm) that each biotype directly intersected with the transect tape at each meter over a total distance of 10 m. This distance was then recorded as a percent cover of each biotype at each meter. This was done for 12-20 transects at each site. Coverage of a wider diversity of biotypes including bare rock (covered in sand, turf algae, or crustose coralline algae), clionid sponge, dead coral, fire coral, gorgonian, green zoanthid, live coral, macroalgae, palythoa, rubble/sand, sand, and sponge, was assessed at all FK reefs using the same methods, but by a different research team. In order to compare reef survey data collected in the FK with surveys completed in Cuba, the bare rock (covered with sand, turf algae, or CCA) category was added to the percent cover of macroalgae on each reef to represent the total algal cover. This decision was made because turf algae or CCA usually covers most surfaces on the reef that are not covered with reef organisms and this estimation was used to complete the surveys in Cuba.

Hydrography and sample collection
At each reef location, a YSI EXO Sonde (YSI Inc./Xylem Inc.) was lowered next to the boat and used to collect temperature, salinity, dissolved oxygen, and pH profiles of the water column (Supporting Information Table S1). A custom Matlab (Mathworks®) script was used to extract values from surface (1.5 m) and reef depths (Supporting Information Table S1).

To evaluate planktonic microbial biomass, 1 ml seawater samples from each site and reef depth were collected, transported back to the field laboratory on ice, preserved using 1% PFA (final concentration) for 30 minutes at 4°C, and flash frozen with liquid nitrogen. Unfiltered seawater samples (40 ml) were collected for the measurement of total non-purgeable organic carbon (TOC) and total nitrogen (TN) and these samples were acidified with concentrated phosphoric acid (70 µl) to remove inorganic carbon. Smaller volume seawater samples (30 ml) were collected and filtered using 0.22 µm, Sterivex ™ filter units for analysis of phosphate (PO_4^{3-}), nitrite and nitrate ($NO_2^- + NO_3^-$), silicate (SiO_4^{4-}), nitrite (NO_2^-), and ammonium (NH_4^+) concentrations. Macronutrient samples were transported back to the field laboratory in a cooler on ice and then frozen at -20° C for long- term storage until they could be analyzed.

Fluidigm amplification
DNA extracts were amplified using Fluidigm microfluidic amplification. Before amplification, 2 ng of each DNA extract was combined with 4 µl of PCR mastermix (Roche High Fidelity Fast Start Kit) in a PCR plate. PCR primers were added to a second plate (50 µM each) and diluted with the Fluidigm loading reagent and water. The primers and extracts suspended within the mastermix were loaded into a primed Fluidigm 48.48 Access Array Integrated Fluidic Circuit (IFC) and the IFC was placed within an AX controller. The Fluidigm Biomark HD PCR machine

257

was then used to amplify the DNA extracts (with no imaging). The following amplification steps and cycle numbers were used: 50 °C for 2 minutes (1 cycle); 70 °C for 20 minutes (1 cycle); 95 °C for 10 minutes (1 cycle); 95 °C for 15 seconds, 55 °C for 30 seconds, and 72 °C for 1 minute (10 cycles); 95 °C for 15 seconds, 80 °C for 30 seconds, 60 °C for 30 seconds, and 72 °C for 1 minute (2 cycles); 95 °C for 15 seconds, 55 °C for 30 seconds, and 72 °C for 1 minute (8 cycles), 95 °C for 15 seconds, 80 °C for 30 seconds, 60 °C for 30 seconds, and 72 °C for 1 minute (2 cycles); 95 °C for 15 seconds, 55 °C for 30 seconds, and 72 °C for 1 minute (8 cycles); and 95 °C for 15 seconds, 80 °C for 30 seconds, 60 °C for 30 seconds, and 72 °C for 1 minute (5 cycles). The total number of cycles in the first amplification process was 38.

After the first amplification, PCR products from each sample were collected and then diluted (1:100) in water. Diluted product (1 μl) from each sample was amplified using Illumina linkers and barcodes in 20 μl volume reactions. The PCR reaction conditions included 95 °C for 10 minutes (1 cycle); 95 °C for 15 seconds, 60 °C for 30 seconds, and 72 °C for 1 minute (15 cycles); and an extension step at 72 °C for 3 minutes. The total number of cycles for the second amplification process was 16.

PCR products were harvested from the second amplification and quantified. Amplicon regions and expected sizes were confirmed using a Fragment Analyzer (Advanced Analytics, Ames, IA). After size confirmation, PCR products were pooled into equal ratios. PCR product pools were run on a gel for size selection and the product was gel purified (Qiagen gel extraction kit). A Bioanalyzer (Agilent) was used to inspect the size and profiles of the pooled and purified PCR products.

Metagenomic sequencing
Four samples were chosen from Jardines de la Reina (sites 2, 4, 5, and 6), and five samples were chosen from the Florida Keys (20, 21, 22, 23, 24). Additionally, a DNA extraction control sample was sequenced to account for potential reagent contamination, but was not analyzed.

A modified cetyl-trimethylammonium bromide (CTAB) - phenol: chloroform: isoamyl alcohol extraction was used to extract DNA from half of each 142 mm filter. Cells on the filters were exposed to a series of physical, enzymatic, and chemical disruptions to enhance cellular lysis by using 3 freeze-thaw cycles, incubating the filters with proteinase-k (20 mg/ml) and lysozyme (20 mg/mL), and vortexing the filters. CTAB, an effective surfactant used for purifying DNA in the presence of polysaccharides (Clarke 2009), was added to the sample, followed by a phenol: chloroform (24:1), phenol: chloroform: isoamyl alcohol (25:24:1), phenol: chloroform (24:1) rinsing series. The aqueous phase was precipitated using molecular grade isopropanol overnight at -20 °C and the DNA pellet was rinsed with 70% ethanol twice before it was eluted into 50 μl of TE buffer (10 mM Tris-Cl, pH 7.5; 1 mM Ethylenediaminetetraacetic acid).

After sequencing, 274,418,737 paired reads were generated with an average read number of 27,441,874 (+/- 9,096,570) paired reads per sample. DNA fragment sizes for the seawater samples ranged from 280-700 bp while the DNA control sample had fragments ranging between 80-600 bp.

Supporting Table S1: Summary of reef descriptions and surface (S) and reef-depth (R) water column properties.

Region *	Site #	Depth (m)	Subregion	Reef type	Latitude and Longitude	Temp. (°C)	Sal. (psu)	DO+ (mg/l)	pH
JR	1	10	JR offshore	back reef	20.77453 N, -78.91517 W	S: 26.86	37.3	6.68	8.16
						R: 26.82	37.3	7.03	8.19
JR	2	17	JR offshore	Fore-reef	20.82598 N, -78.97931 W	S: 26.76	37.4	6.45	8.14
						R: 26.74	37.4	6.40	8.14
JR	3	2	JR gulf	lagoon	20.81478 N, -78.88320 W	S: 25.75	38.9	6.78	8.13
						R: 25.73	39.0	7.19	8.14
JR	4	1.5	JR gulf	back reef	20.87765 N, -78.97028 W	S: 24.68	38.9	6.34	8.11
						R: 24.68	38.9	6.34	8.11
JR	5	1.3	JR gulf	back reef	21.09232 N, -78.73354 W	S:24.91	39.7	6.62	8.17
						R: 24.91	39.7	6.62	8.17
JR	6	0.75	JR gulf	back reef	21.10845 N, -78.72080 W	S: 24.12	39.9	7.02	8.19
						R: 24.12	39.9	7.02	8.19
CAN	7	7	CAN	deep fore-reef with wall drop-off	21.58422 N, -81.56530 W	S: 29.45	37.43	6.29	8.16
						R: 29.35	37.4	6.31	8.16
CAN	8	5	CAN	reef crest	21.58693 N, -81.58308 W	S: 29.73	37.5	5.32	8.10
						R: 29.41	37.48	4.61	8.07
CAN	9	5	CAN	reef crest	21.58802 N, -81.58180 W	S: 28.68	37.37	5.56	8.06
						R: 28.69	37.38	5.50	8.07
CAN	10	15	CAN	deep fore-reef	21.58158 N, -81.59057 W	S: 27.94	37.36	6.41	8.12
						R: 27.93	37.36	6.39	8.15
CAN	11	4	CAN	reef crest	21.58462 N, -81.59720 W	S: 28.82	37.40	5.54	8.14
						R: 28.83	37.41	5.45	8.14
CAN	12	3	CAN	reef crest	21.58408 N, -81.62805 W	S: 28.82	37.41	5.75	8.14
						R: 28.82	37.41	5.72	8.14
CAN	13	~7	CAN	deep fore-reef	21.56855 N, -81.63165 W	S: 28.18	37.31	6.04	8.12
						R: 27.63	37.40	5.98	8.14
CAN	14	9	CAN	deep fore-reef	21.56893 N, -81.63820 W	S: 28.18	37.37	6.08	8.11
						R: 27.85	37.37	6.14	8.15
CAN	15	10	CAN	Fore-reef, 500m off reef crest	21.55521 N, -81.76323 W	S: 28.07	37.34	6.36	8.14
						R: 28.08	37.35	6.47	8.19
CAN	16	~1	CAN	back reef	21.56272 N, -81.76676 W	S: 28.03	37.39	6.22	8.12
						R: 27.91	37.36	6.16	8.15
CAN	17	~1	CAN	back reef	21.60300 N,	S: 27.23	37.40	5.9	8.10

					-81.93300 W	R: 27.20	37.41	5.96	8.12
CAN	18	~1	CAN	back reef	21.59684 N,	S: 26.81	37.38	5.43	5.40
					-81.96867 W	R: 26.82	37.37	5.40	8.09
CAN	19	10	CAN	mid-depth fore-reef	21.71333 N, -82.10417 W	S: 27.62	37.39	6.10	8.12
						R: 27.63	37.41	6.03	8.13
FK	20	6	FK offshore	mid-channel patch reef	24.55945 N, -81.50098 W	S: 27.59	37.36	6.33	8.20
						R: 27.59	37.36	6.34	8.21
FK	21	7	FK offshore	offshore patch reef	24.55228 N, -81.43700 W	S: 27.12	37.32	6.58	8.19
						R: 27.12	37.32	6.55	8.21
FK	22	6	FK offshore	Spur and groove reef	24.54500 N, -81.40600 W	S: 27.35	37.26	6.14	8.16
						R: 27.35	37.26	6.11	8.17
FK	23	6	FK offshore	reef flat	24.55228 N, -81.38130 W	S: 27.26	37.22	6.28	8.19
						R: 27.26	37.22	6.28	8.19
FK	24	1	FK nearshore	nearshore reef	24.60548 N, -81.42930 W	S: 27.98	37.25	6.65	8.23
						R: 27.98	37.25	6.64	8.23
FK	25	1	FK nearshore	nearshore patch reef	24.61565 N, -81.39390 W	S: 28.42	37.42	5.11	8.13
						R: 28.42	37.43	5.04	8.14

Sal. = salinity
*JR = Jardines de la Reina, Cuba; CAN = Los Canarreos, Cuba; FK = Florida Keys, USA.
+ DO = dissolved oxygen

Table S2. Average percent cover of dominant reef organisms and substrates at reef sites across Jardines de la Reina, Los Canarreos, and the Florida Keys.

Biotype Category	Average % Cover (S.D.)					
	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6
Jardines de la Reina						
coral	24.3 (11.8)	30.4 (11.1)	33.6 (13.1)	14.2 (6.9)	55.3 (13.7)	7.0 (5.8)
algae	60.9 (16.4)	44.2 (27.7)	22.0 (9.4)	41 (10.9)	6.4 (8.3)	59.5 (26.2)
sand	2.9 (3.9)	0.6 (1.3)	26.3 (21.0)	39.8 (15.9)	24.0 (12.8)	18.3 (21.4)
sponge	9.7 (6.2)	6.4 (4.9)	13.7 (5.9)	4.1 (2.3)	8.2 (5.3)	1.9 (1.6)
Los Canarreos	Site 13	Site 14	Site 15	Site 17	Site 19	
coral	2.2 (2.3)	1.8 (2.0)	2.4 (3.0)	10.4 (5.1)	10.9 (4.8)	
algae	96.5 (3.4)	97.2 (2.5)	94.3 (4.0)	72.7 (7.9)	65.7 (12.9)	
sand	0.1 (0.02)	0	0.3 (0.7)	13.2 (8.9)	11.8 (6.5)	
sponge	0.9 (1.5)	0.7 (1.6)	1.4 (2.1)	0.9 (1.4)	1.3 (1.2)	
Florida Keys	Site 20	Site 21	Site 22	Site 23	Site 24	Site 25
bare rock (w/ sand, turf, or CCA)	19.7 (1.6)	33.7 (1.3)	63.6 (1.4)	74.1 (0.6)	41.4 (0.0)	68.7 (1.1)
clionid sponge	0	0.73 (0.16)	0	0	0	0
dead coral	13.3 (1.1)	5.6 (0.8)	2.1 (0.2)	0.3 (0.1)	6.5 (0.9)	0
fire coral	0.1 (0.0)	0.4 (0.1)	0.6 (0.1)	1.0 (0.1)	0.1 (0.0)	0.4 (0.1)
gorgonian	3.8 (0.3)	3.4 (0.1)	1.8 (0.2)	2.4 (0.1)	1.4 (0.1)	0.5 (0.1)
green zoanthid	0	0	0.2 (0.05)	0	0	0
live coral	30.8 (1.5)	14.5 (0.9)	14.0 (1.2)	2.7 (0.2)	11.6 (0.6)	1.9 (0.3)
macroalgae	0	1.4 (0.1)	3.6 (0.3)	1.7 (0.1)	14.5 (0.7)	26.0 (1.0)
palythoa	0	1.5 (0.2)	11.9 (0.7)	0	0	0
rubble/sand	6.2 (0.5)	8.7 (1.7)	0.5 (0.1)	2.8 (0.4)	7.9 (1.8)	1.2 (0.4)
sand	21.3 (2.0)	19.9 (1.3)	0	2.4 (0.3)	12.0 (2.1)	1.1 (0.2)
sponge	4.9 (0.3)	10.2 (0.3)	1.8 (0.1)	5.3 (0.2)	4.6 (0.5)	0.2 (0.0)
total algae*	19.6 (1.6)	32.3 (1.7)	67.2 (1.5)	75.7 (0.6)	55.9 (3.2)	94.7 (0.6)

S.D. = standard deviation
*total algae = average of the sum of macroalgae and bare rock (w/ sand, turf, or CCA) categories

Table S3. Relative abundances (%) and standard deviations of significantly enriched (grey shading) or depleted MED nodes in Jardines de la Reina or Florida Keys reef seawater as revealed by DESeq2.

MED node	Taxonomy	Mean (SD) JR	Mean (SD) FK
MED1988	Verrucomicrobia, *Roseibacillus*	0	0.30 (0.60)
MED2280	Verrucomicrobia, Opitutales, Puniceicoccaceae, *Coraliomargarita*	0.10 (0.12)	1.03 (0.96)
MED4771	Marinimicrobia, SAR406 clade	0.11 (0.14)	0.01 (0.02)
MED4772	Marinimicrobia, SAR406 clade	0.090 (0.14)	0.001 (0.006)
MED256	Gammaproteobacteria, Steroidobacterales, *Woeseia*	0.06 (0.06)	0.0070 (0.016)
MED4049	Gammaproteobacteria, SAR86 clade	0.12 (0.14)	0.013 (0.035)
MED4227	Gammaproteobacteria, Ectothiorhodospirales, uncultured	0.313 (0.440)	0.0036 (0.0091)
MED2377	Gammaproteobacteria, Cellvibrionales, Porticoccaceae, SAR92 clade	0	0.40 (0.53)
MED3751	Gammaproteobacteria, Burkholderiaceae, MWH-UniP1 aquatic group	0.10 (0.12)	0.030 (0.08)
MED1982	Deltaproteobacteria, SAR324 clade, Marine group B	1.09 (0.76)	0.04 (0.05)
MED1255	Cyanobacteria, Synechococcales, Cyanobiaceae, Cyanobium, PCC-06307	0.27 (0.24)	0.02 (0.05)
MED1250	Cyanobacteria, Synechococcales, Cyanobiaceae, Cyanobium, PCC-06307	0.027 (0.33)	0.005 (0.012)
MED1253	Cyanobacteria, Synechococcales, Cyanobiaceae, Cyanobium, CC-9902	0.26 (0.27)	0.04 (0.05)
MED1263	Cyanobacteria, Synechococcales, Cyanobiaceae, Cyanobium, CC-9902	0.07 (0.08)	0.0080 (0.015)
MED50	Cyanobacteria, Synechococcales, Cyanobiaceae, Cyanobium, CC-9902	0.074 (0.14)	0
MED3985	Cyanobacteria, Synechococcales, Cyanobiaceae	0.077 (0.075)	0.01 (0.01)
MED5521	Cyanobacteria, Synechococcales, Cyanobiaceae	0.13 (0.21)	0.0050 (0.013)
MED473	Chloroflexi, Dehalococcoidia, SAR202 clade	0.16 (0.16)	0.0040 (0.11)
MED4353	Bacteroidetes, Sphingobacteriales, unclassified	0.18 (0.34)	0
MED2355	Bacteroidetes, Rhodothermia, *Balneola*	0.0090 (0.16)	0.72 (0.95)
MED1983	Bacteroidetes, Rhodothermia, *Balneola*	0	0.29 (0.70)
MED3131	Bacteroidetes, Flavobacteriales, NS5 marine group	0.06 (0.08)	0.004 (0.008)
MED5356	Bacteroidetes, Flavobacteriales, NS5 marine group	0.11 (0.16)	0.0030 (0.012)
MED5331	Bacteroidetes, Flavobacteriales, NS5 marine group	0.19 (0.23)	0.004 (0.0130)
MED3027	Bacteroidetes, Flavobacteriales, NS4 marine group	0.014 (0.021)	0.53 (0.81)
MED5345	Bacteroidetes, Flavobacteriales, NS4 marine group	0.016 (0.019)	0.24 (0.29)
MED3201	Bacteroidetes, Flavobacteriales, *Formosa*	0.0018 (0.0072)	0.14 (0.23)
MED4473	Bacteroidetes, Flavobacteriales, Crocinitomicaceae, *Fluviicola*	0.089 (0.14)	0.00071 (0.0032)
MED4535	Bacteroidetes, Cryomorphaceae	0.79 (0.83)	0.018 (0.043)
MED5604	Archaea, Euryarchaeota, Thermoplasmata, Marine Group II	0.11 (0.16)	0.0064 (0.0098)

| MED4303	Archaea, Euryarchaeota, Thermoplasmata, Marine Group III	0.11 (0.12)	0.0010 (0.020)
MED3798	Alphaproteobacteria, unclassified	0.045 (0.058)	0.00066 (0.0029)
MED4248	Alphaproteobacteria, SAR11, "*Candidatus* Pelagibacter"	0.0018 (0.0049)	0.070 (0.13)
MED4268	Alphaproteobacteria, SAR11, "*Candidatus* Pelagibacter"	0.038 (0.056)	0.0014 (0.0044)
MED4286	Alphaproteobacteria, SAR11, "*Candidatus* Pelagibacter"	0.14 (0.22)	0.0064 (0.019)
MED4269	Alphaproteobacteria, SAR11, "*Candidatus* Pelagibacter"	0.14 (0.17)	0.013 (0.023)
MED2282	Alphaproteobacteria, Rhodobacteraceae	0.19 (0.30)	1.90 (2.12)
MED2231	Alphaproteobacteria, Rhodobacteraceae	0.18 (0.23)	0.0078 (0.019)
MED3481	Alphaproteobacteria, Rhizobiales, *Rhodobium*	0.94 (1.36)	0.0099 (0.024)
MED4087	Alphaproteobacteria, Puniceispirillales, SAR116 clade	0.24 (0.19)	0.29 (0.045)
MED4019	Alphaproteobacteria, Puniceispirillales, SAR116 clade	0.28 (0.34)	0.0082 (0.021)
MED5561	Alphaproteobacteria, Puniceispirillales, SAR116 clade	0.12 (0.13)	0.0014 (0.0042)
MED3678	Alphaproteobacteria, Puniceispirillales, SAR116 clade	0.073 (0.12)	0
MED1028	Actinobacteria, PeM15	0.13 (0.11)	0.0056 (0.0096)
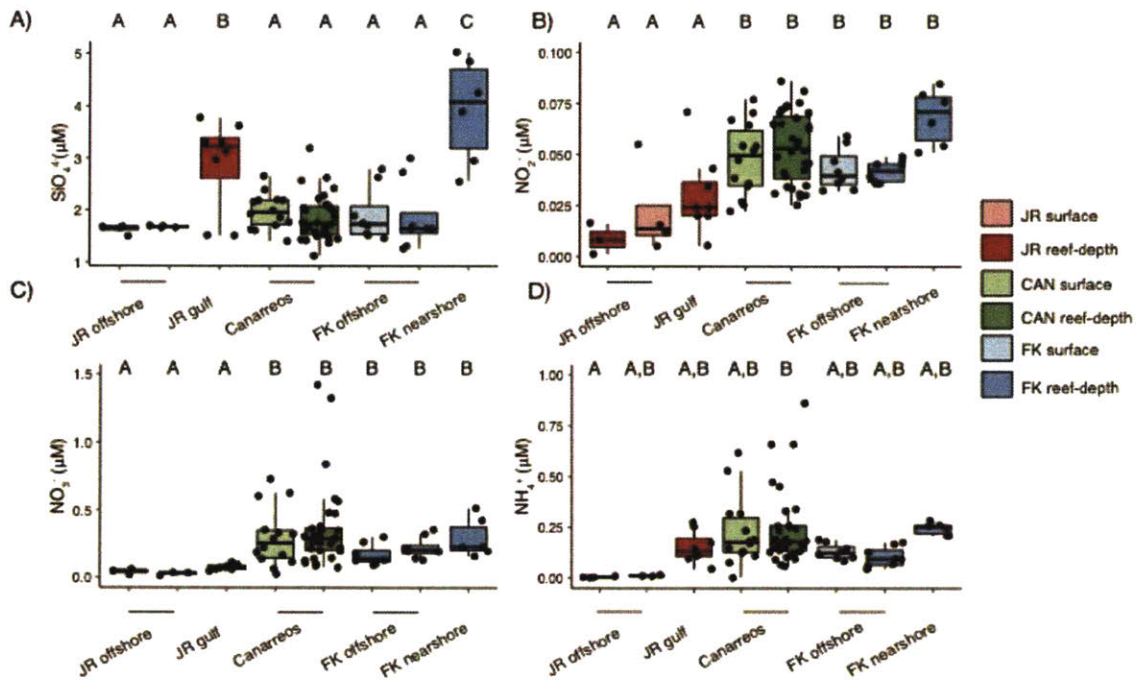
Figure S1. Concentrations of organic and inorganic macronutrients measured by subregion and reef-system. A) SiO_4^{4-} (Silicate) B) NO_2^- (nitrite), C), NO_3^- (nitrate), D) NH_4^+ (ammonium). Boxplots are drawn as follows: the lower and upper edges of the boxplot correspond to the first and third quartiles, the whiskers extend to the largest or smallest value at 1.5 times the interquartile, and the black bar across the box represents the median. Boxplots with different letters are significantly different from each other (ANOVA, Tukey's HSD test, $p < 0.050$). JR = Jardines de la Reina, CAN = Canarreos, FK = Florida Keys. Surface refers to surface seawater.
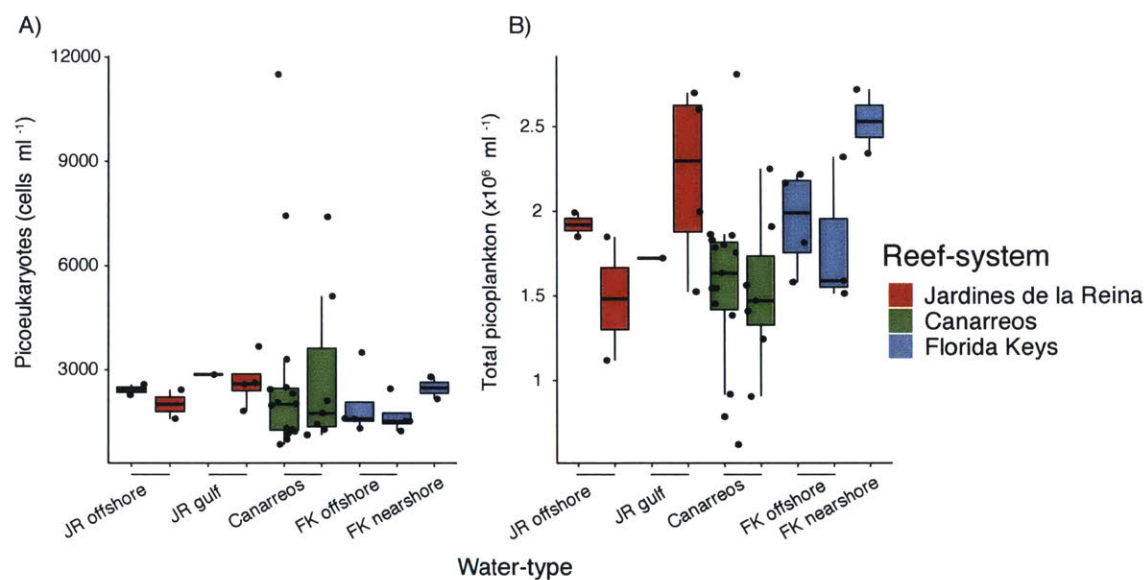
Figure S2. Cell abundances of picoplankton functional groups, including picoeukaryotes (A) and total cells (summation of *Prochlorococcus*, *Synechococcus*, Picoeukaryotes, and unpigmented cells at each depth and subregion) (B). Lower and upper edges of the boxplot correspond to the first and third quartiles, the whiskers extend to the largest or smallest value at 1.5 times the interquartile, and the black bar across the box represents the median. No significant differences were detected (Kruskal-Wallis Rank Sum test and Dunn's test, $p < 0.050$).
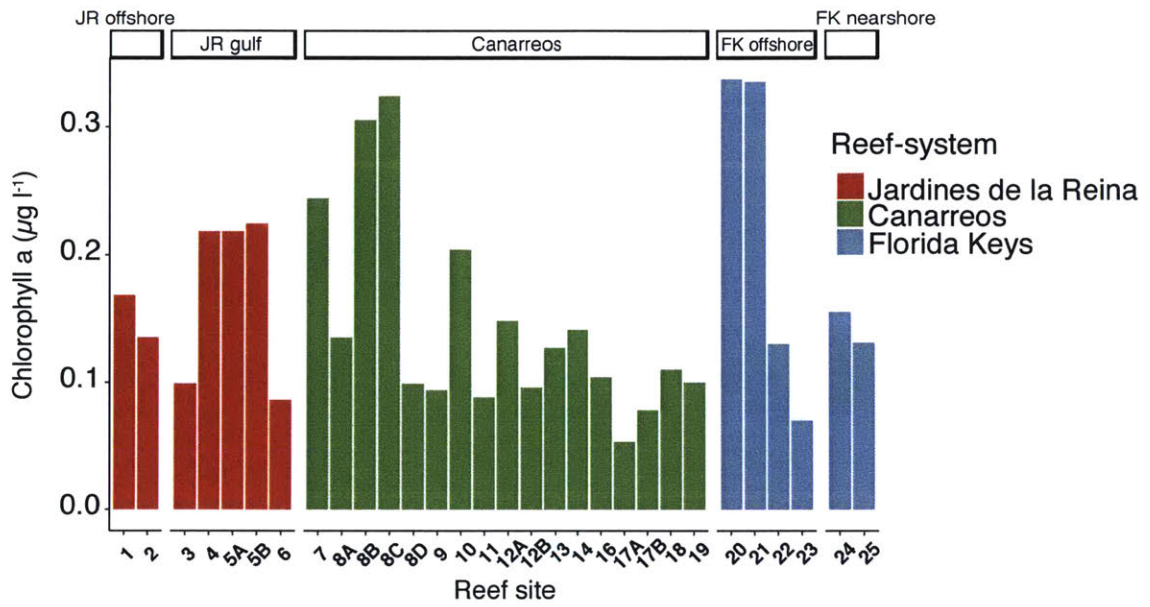
Figure S3. Concentrations of total chlorophyll *a* by subregion and reef-system.
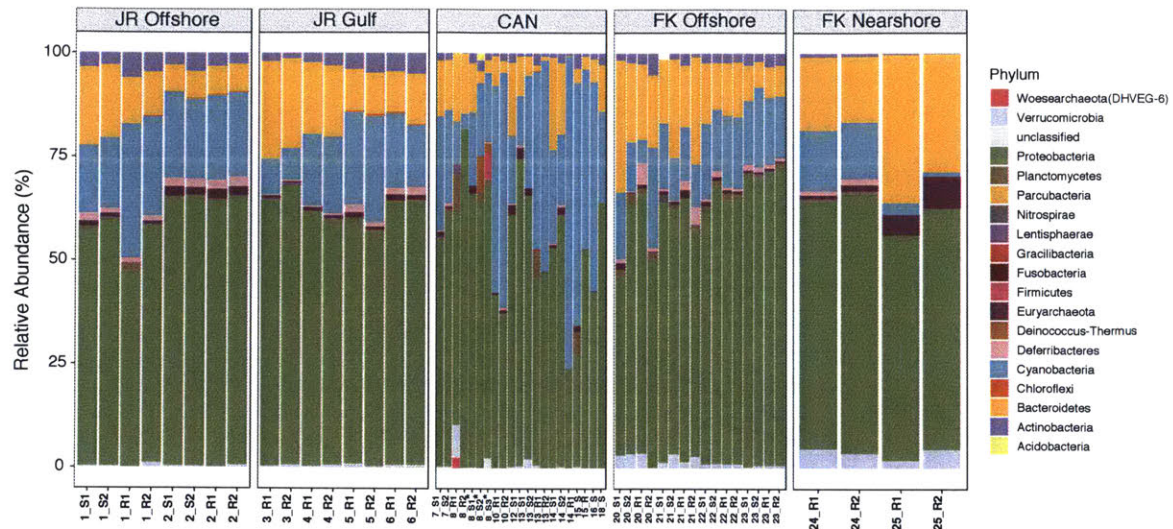
Figure S4. Relative abundance (%) of bacterial and archaeal phyla determined from SSU rRNA gene amplicons. S indicates surface and R indicates reef-depth. Replicate samples are numbered. Samples with * were collected from the same location but on a different day.

267
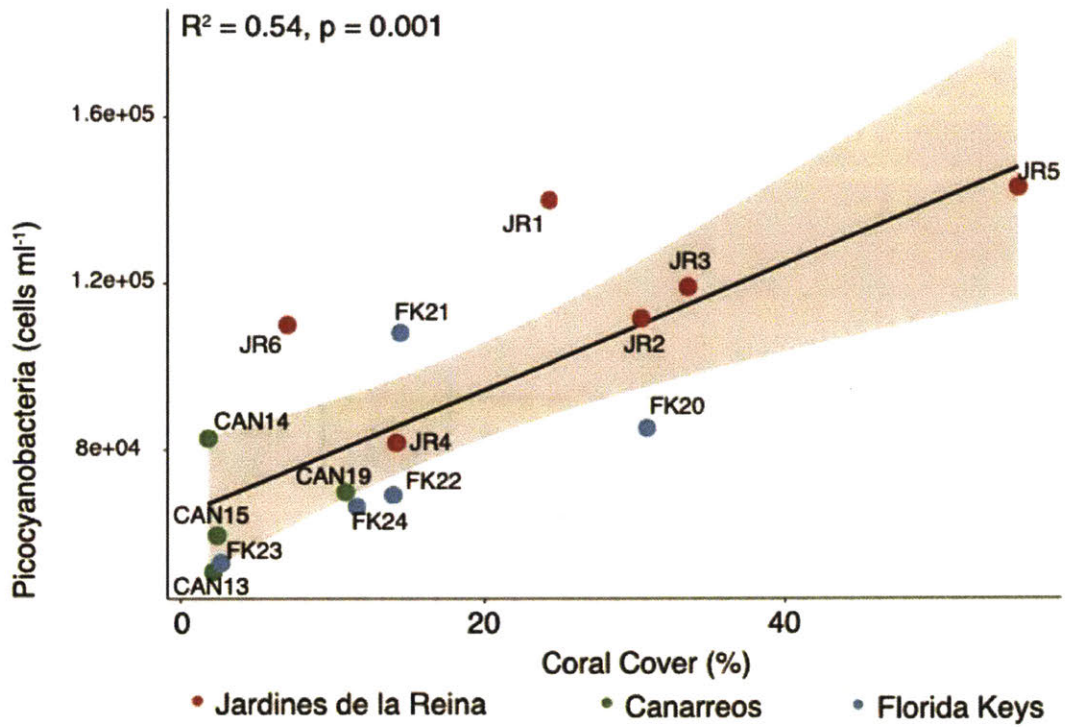
Figure S5. Regression between the abundance of picocyanobacteria (the summation of the abundance of *Prochlorococcus* and *Synechococcus* cells) detected in reef depth seawater and coral cover across the three reef-systems. Each point is labeled with the site name.
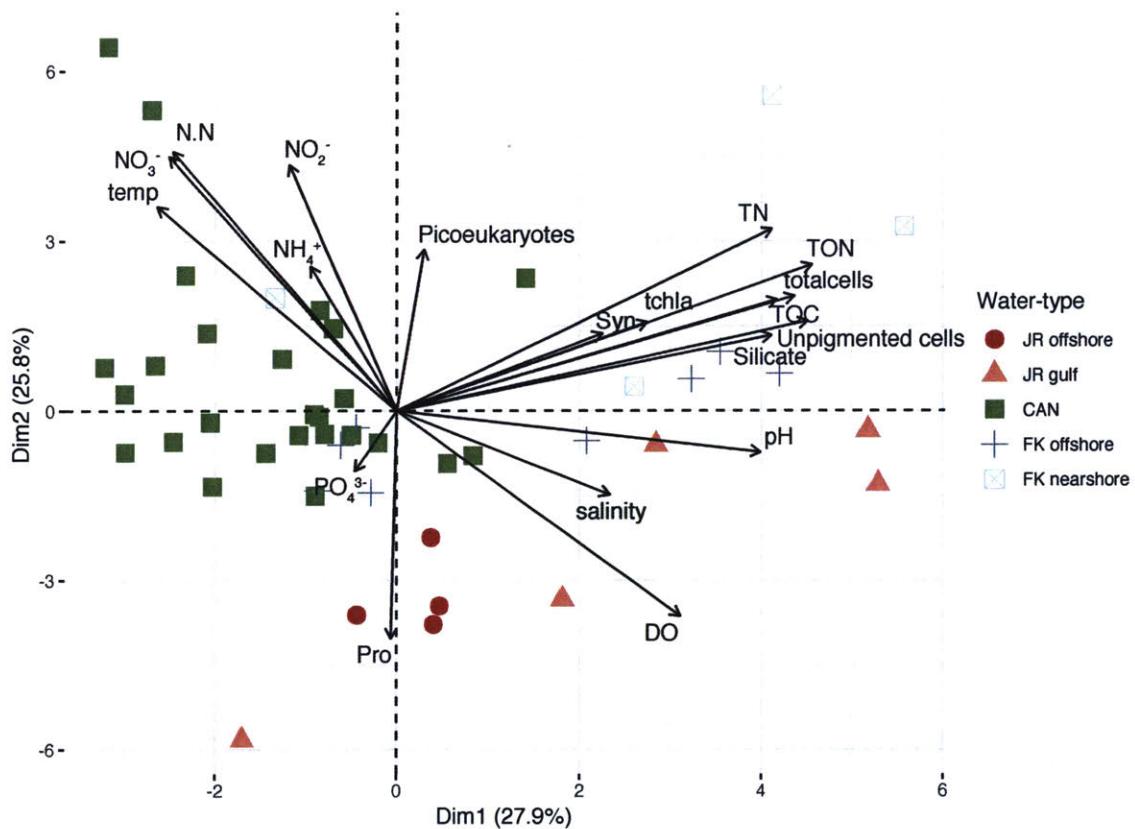
Figure S6. Principal components analysis (PCA) biplot of the physicochemical, biogeochemical, and cell abundance measurements collected across reef-systems. Symbol color and shape reflect subregion. N.N = nitrate + nitrite, Pro = *Prochlorococcus*, Syn = *Synechococcus*, DO = dissolved oxygen, TOC = total organic carbon, TON = total organic nitrogen, TN = total nitrogen, tchla = total chlorophyll *a*, temp. = temperature.
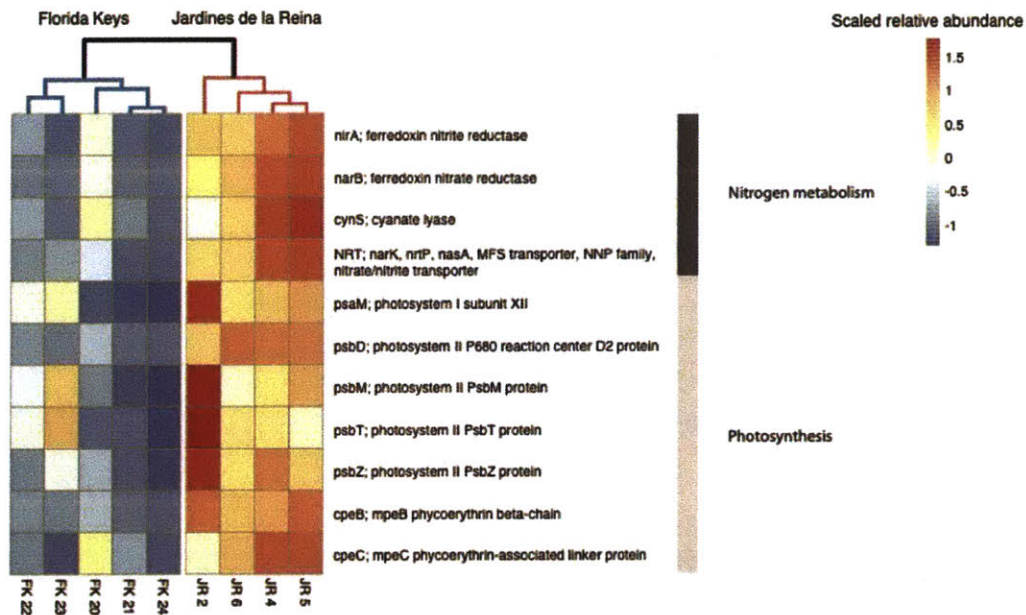
Figure S7. Photosynthesis and nitrogen metabolism genes are enriched in Jardines de la Reina compared to the Florida Keys. The abundances of KEGG orthologs (KOs) were scaled using the 10th and 90th quantiles of the data for visualization. The dendrogram reflects hierarchical clustering of the samples using the 'hclust' function in R.

Chapter 6

Extracellular reef metabolites across the protected Jardines de la Reina, Cuba reef-system

6.1 Abstract

Coral reef ecosystems are incredibly diverse marine biomes that rely on nutrient cycling by microorganisms to sustain high productivity and biomass in oligotrophic regions. Understanding extracellular reef metabolites in seawater, the small organic molecules that serve as the currency for microorganisms, may provide insight into benthic-pelagic coupling as well as the complexity of nutrient cycling in coral reef ecosystems. In this research, we extracted, identified, and quantified specific known metabolites of interest (targeted approach) as well as surveyed trends in metabolite feature composition (untargeted approach) from surface and reef-depth (6 – 14 m) seawater as well as off-reef seawater collected across the protected Caribbean reef-system of Jardines de la Reina, Cuba. We found that metabolite composition was fairly similar between reef-depth, surface, and bluewater seawater across the archipelago, corresponding with other biogeochemical and physicochemical measurements that suggest that environmental conditions are homogenous across fore reefs within Jardines de la Reina. We identified 33 metabolites and quantified the environmental concentrations of 22 of these metabolites including amino acids, nucleosides, vitamins, and metabolic intermediates. We observed that riboflavin concentrations were higher in reef-depth seawater compared to surface seawater, suggesting that riboflavin may be produced by reef organisms and degraded in surface seawater through photochemical lysis. Additionally, methylthioadenosine (MTA) concentrations increased significantly within the central region of the archipelago, displaying biogeographic patterns that warrant future investigation. Here we provide novel baseline knowledge about the extracellular metabolite composition of seawater from relatively healthy forereefs within the protected Jardines de la Reina reef-system for comparative analysis with less healthy reef-systems. We also lay the groundwork for future investigations into the variation in metabolite composition across a reef, sources and sinks of different metabolites, and changes in metabolites over stronger environmental, temporal, and reef quality gradients.

6.2 Introduction

Coral reefs are productive and regenerative ecosystems that generally exist in oligotrophic waters, but tight nutrient recycling by microorganisms and benthic productivity by corals, algae, sponges and their symbionts sustain an incredible biomass and diversity of vertebrates and invertebrates (Hatcher 1988; Atkinson 2010). At depth, reef organisms, like corals, macroalgae, and sponges, release dissolved and particulate organic matter (DOM and POM, respectively) into the surrounding seawater (Wild et al. 2004; Tanaka et al. 2009; Haas et al. 2011; de Goeij et al. 2013). Microorganisms living within the seawater degrade benthic-derived OM and recycle limiting nutrients like nitrogen and phosphorous back into the dissolved phase, making these

nutrients available for primary producers like eukaryotic phytoplankton and picoplanktonic cyanobacteria (reviewed within Azam and Malfatti 2007). Furthermore, bulk exudates from reef organisms impact seawater microbial community composition and function (Haas et al. 2011; Haas et al. 2013; Nelson et al. 2013) and reef composition influences microbial community structure and function at the level of the ecosystem (Haas et al. 2016).

Metabolites, compounds classifying as small molecular weight organic matter, serve as the chemical currency for life and are used as growth substrates, for communication, and for chemical defense (reviewed by Kujawinski 2011). Study of metabolites can reveal interactions occurring between cells (through investigation of extracellular metabolites) as well as the metabolic responses of cells to different conditions (intracellular metabolites). In the context of coral reef ecology, understanding the flux of extracellular metabolites through this intricate microbial loop may help tease apart the importance of benthic-pelagic coupling on reefs as well as reveal essential yet cryptic chemical transactions between reef organisms and the microorganisms surrounding them within the seawater. While bulk approaches for measuring the flux of carbon, nitrogen, and phosphorous through coral reefs have provided fundamental information about nutrient cycling and reef ecosystem metabolism (reviewed within Atkinson 2010), these studies are lacking a highly resolved understanding of the extracellular metabolite landscape on reefs and information about the minutia and intricacy of chemical transactions across the reef water column. Research approaches that combine the analysis of bulk measurements with the study of environmental metabolites, made possible by recent methodological advances in the field of metabolomics, will provide insight into the complexity and diversity of molecules across coral reefs and permit scientists to link dynamics of individual metabolites to larger-scale organic matter cycling.

The study of metabolites is usually divided into two complementary analytical strategies, each with their own trade-offs. The targeted strategy identifies and quantifies environmental concentrations of specific, known metabolites whereas the untargeted strategy provides semi-quantitative information about all the observable metabolic features (a unique combination of a mass to charge ratio and retention time) within a sample. When the identities and concentrations of specific metabolites are known and need to be monitored throughout the course of an investigation, the targeted approach is most suitable, but it does not support discovery of new metabolites. In contrast, the untargeted method can be used to investigate multi-variate patterns across a dataset and for putative metabolite discovery, but is only semi-quantitative and metabolite identifications require follow-up analyses. Metabolomic investigations usually use a combination of both methods in order to comprehensively query the processes of interest.

Identification and quantification of coral-associated intracellular metabolites using methodologies have recently been performed to resolve the response of corals to stressors like bleaching and ocean acidification (Sogin et al. 2016) and competitive interactions with other coral species, a fungal mat, and different types of macroalgae (Quinn et al. 2016). Coral tissue intracellular metabolomes are also correlated with coral microbiomes and their dinoflagellate symbionts and are specific to coral species (Sogin et al. 2017; Vohsen et al. 2019), implying an intricate connection between the coral microbiome and coral nutrition and metabolism.

The study of dissolved extracellular metabolites in reef seawater has not been widely investigated and is an emerging area of research for coral reef microbial ecologists (Kelly et al. 2018). A study of both targeted and untargeted extracellular metabolites in sponge inhalant and exhalant seawater compared to off-reef seawater found that the sponge exhalant had a higher diversity of untargeted metabolite features as well as higher concentrations of nucleosides (Fiore

276

et al. 2017). An investigation of untargeted extracellular metabolites derived from seawater sampled along a gradient (0 to 5 cm away from the coral surface) surrounding two different coral species in the Arabian Gulf found differences in the elemental composition across the gradient and putatively detected infochemicals involved in quorum-sensing and chemical defense among others (Ochsenkuhn et al. 2018). Despite the knowledge contributed by these studies, there is little baseline information about the metabolite landscape across reefs and how the metabolite composition changes along various temporal, spatial, or environmental gradients. The primary questions that need to be addressed include: 1) what are the dominant extracellular metabolites in reef seawater (and their environmental concentrations)? and 2) are there spatial or biogeographical patterns in the distribution of specific metabolites across the water column and different reefs within the same reef-system?

To obtain this baseline understanding of the metabolite composition in coral reef seawater, we surveyed the reef composition and collected seawater from surface and reef-depths across nine shallow forereefs within the Cuban reef-system of Jardines de la Reina and subjected this seawater to targeted and untargeted Ultra Performance Liquid Chromatography (UPLC) tandem mass spectrometry. Additionally, our analysis is complemented by quantification of inorganic and organic macronutrient concentrations, chlorophyll a and phaeophytin concentrations, and picoplankton cell abundances in order to relate coral reef exometabolomes to physicochemical and bulk biogeochemical measurements.

6.3 Methods

Coral Surveys and sample collection

277

Reef composition and seawater microbial biogeochemistry were surveyed at nine shallow (6 – 14 m) forereefs during a cruise to Jardines de la Reina (JR), Cuba in November of 2017 (Figure 1). Reef surveys were conducted by scuba divers at all forereefs using the line-intercept survey method (as described in Obura 2014). The percent cover of corals, macroalgae, sponges, rock, and sand were recorded on each reef by divers who conducted 10 - 20, 10 m transects across the reef substrate (surveys were completed between 6 - 14 m depth). In addition to surveying forereefs, we also surveyed two off-reef surface seawater 'bluewater' locations (800 – 1600 m depth) to make comparisons between reef and off-reef samples.

At each reef, CTD casts were completed with a YSI Exo Sonde (Xylem Inc., Yellow Springs, OH, USA) to describe the physicochemical properties (e.g. temperature, salinity, dissolved oxygen, and pH) of the water-column (File S1). Surface (1 m) and reef-depth (1 m off reef; 5 – 13 m depth) seawater was sampled from nine JR forereefs and surface seawater was sampled from two bluewater locations adjacent to the forereefs (Figure 1). Seawater collected for cell counts (1 mL) and macronutrient analyses (30 and 40 mLs) was collected from surface and reef-depth using a submersible groundwater pump. Samples collected for chlorophyll *a* and phaeophytin analysis (4 L) were obtained from bluewater and reef-depth seawater.

To collect reef-depth seawater at each site for metabolomics analyses, a scuba diver collected seawater in an acid-washed, Teflon-lined 8 L Niskin Bottle (General Oceanics, Miami, Florida, USA). To collect reef-depth reef seawater, the diver descended with the Niskin bottle in the cocked position and left the Niskin on a barren part of the reef (i.e. sand patch) for 30 – 50 minutes while completing reef surveys. At the end of each dive, the diver retrieved the Niskin bottle, thoroughly rinsed the Niskin bottle with reef-depth seawater, located an area of the reef that was topographically complex (covered with hard and soft corals as well as sponges), and

triggered the Niskin bottle to close, capturing reef-depth seawater within the Niskin chamber. Upon ascent, the Niskin bottle was immediately placed within an iced cooler. Before leaving the site, surface seawater was collected by hand off of the back of the dive boat into acid-washed and autoclaved 2 L polycarbonate bottles (Nalgene™, Thermo Scientific™, Waltham, MA, USA). Bottles were triple-rinsed with surface seawater, samples were collected in duplicate, and the bottles were filled to the brim and capped. At the two bluewater sites, surface seawater was collected in duplicate using the same process. Surface samples were stored in the iced cooler until processing.

Sample processing

Aboard the research vessel, surface and reef-depth seawater collected for analysis of total organic carbon and nitrogen (including particulate and dissolved) concentrations was acidified with 75 μl of concentrated phosphoric acid, capped, and stored at room temperature. Surface and reef-depth seawater samples collected for analyses of inorganic macronutrient concentrations (30 mL) were frozen after subsamples (1 mL) of unfiltered seawater were collected for quantifying cell abundances. Seawater, collected for enumeration of *Prochlorococcus*, *Synechococcus*, picoeukaryotic cells, and unpigmented cells (heterotrophs) using flow cytometry, was fixed with paraformaldehyde (1% final volume), incubated at 4 °C in the dark for 30 minutes, frozen at -50 °C on the research vessel, and then stored at -80 °C prior to analysis.

For metabolomics analyses and within an hour of collection, 4 L of reef-depth seawater collected within the Niskin bottle was transferred into two acid-washed and autoclaved 2 L polycarbonate bottles using acid-washed PharMedBPT tubing (L/S #24, Masterflex™, Cole – Parmer, Vernon Hills, IL, USA). Reef-depth and surface seawater were filtered separately through PTFE 0.2 μm pore size, 47 mm filters (Omnipore, EMD Millipore Corporation,

279

Billerica, MA, USA) using peristalsis (MasterFlex L/S pump and pump heads, Cole-Parmer, Vernon Hills, IL, USA). PharMed® BPT (Masterflex® B/T®, Cole-Parmer) tubing and acid-washed Fluorinated Ethylene Propylene (FEP) tubing (890 Tubing, Nalgene™, Thermo Scientific™, Waltham, MA, USA) were used to transfer pumped seawater through the filter membrane into acid-washed and autoclaved polycarbonate collection bottles. Filters were encased in acid-washed and autoclaved 47 mm, PFA in-line filter holders (Adventec, Cole-Parmer).

The filtrate was collected into cleaned polycarbonate bottles, subsamples were collected for analysis of dissolved organic carbon (DOC) post-filtration, and the filtrate was acidified to a pH of ~3 using 12 M trace-metal grade hydrochloric acid (HCl, Optima™, Fisher Chemical, Fisher Scientific, Hampton, NH, USA). During filtering, the flow rate was kept low to minimize the lysis of cells on the surface of the filter, but we did note an approximate 5 μM increase in DOC concentrations in the filtrate compared to unfiltered total organic carbon (TOC), indicating potential contamination from cellular lysis.

We performed solid phase extraction (SPE) to concentrate and extract metabolites (e.g. low molecular weight dissolved organic matter) from the filtered seawater using a vacuum manifold (Waters Corporation, Milford, MA, USA). Prior to and after SPE, the bottles were weighed in order to calculate the volume of seawater that passed through the SPE cartridges. Filtrate was then passed through acid-washed FEP tubing and 1g/6 cc PPL cartridges (BondElut; Agilent, Santa Clara, CA, USA) using gentle vacuum pressure (Kido Soule et al. 2015; Fiore et al. 2017). After filtrate passed through the cartridge, cartridges were removed from the manifold, wrapped with combusted aluminum foil, placed within sterile Whirl-Pak™ bags (Nasco, Fisher Scientific, Hampton, NH, USA), and stored in the field at -50° C. Wrapped cartridges were

shipped back to Woods Hole Oceanographic Institution (WHOI) at -20 °C and then stored at -80

°C until the dissolved organic matter (DOM) could be eluted off of the cartridges. To finish the

extraction process, cartridges were rinsed with 4 bed-volumes of 0.01 M HCl, dried using gentle

vacuum pressure for 5 minutes, and then eluted into pre-combusted glass vials using 100%

methanol (volume of eluents was 6 mL). Methanol extracts were transferred into pre-combusted

glass amber vials using pre-combusted glass pipettes, and stored at -20 °C.

To prepare the samples for mass spectrometry, methanol extracts were dried down using

a vacuum centrifuge. Dried extracts were re-suspended in 200 μl of a 95: 5 (v/v) MilliQ water:

acetonitrile solution with deuterated biotin (final concentration 0.05 mg ml^{-1}) and vortexed. The

extracts were then divided into three different preparations: 100 μl was transferred into a

separate analysis vial containing a combusted glass insert (targeted metabolite analysis), 50 μl

was transferred into a combusted 4 ml vial to be used for a 'pooled' sample, and 25 μl was

transferred into a 4 ml vial to be diluted and prepared for the untargeted metabolite analysis. The

'pooled' sample is a combination of extracts from all samples and is used to as a quality control

during the instrument runs. Untargeted extracts were then diluted with 600 μl of the deuterated

standard and water: ACN solution and vortexed. A 100 μl subsample of this diluted extract was

then aliquoted into an analysis vial and 75 μl of the dilution was diverted into a 'pooled' sample.

Prepared extracts were stored at -20 °C until analysis.

Macronutrient and chlorophyll analysis and cell abundances quantification

Non-purgeable total organic carbon (TOC, unfiltered), dissolved organic carbon (DOC, filtered),

total nitrogen (TN, unfiltered), and total dissolved nitrogen (TDN, filtered) concentrations were

analyzed using a Shimadzu TOC-V$_{CSH}$ total organic carbon analyzer (Hansell and Carlson 2001).

Inorganic macronutrient (PO_4^{3-}, $NO_2^-+NO_3^-$, NO_2^-, NH_4^+, silicate) concentrations were measured

with a continuous segmented flow system (as used in Apprill and Rappé 2011). Nitrite concentrations were subtracted from the nitrite + nitrate concentrations to obtain the nitrate concentrations. The concentration of total organic nitrogen was obtained by subtracting the sum of the inorganic nitrogen species ($NO_2^-+NO_3^-$ and NH_4^+) from the total nitrogen concentration for each sample.

Chlorophyll *a* and phaeophytin were extracted from GFF filters using 90% acetone in water and the optical density (OD) values were measured on a calibrated spectrophotometer using standard optics (Lambda 18, Perkin Elmer, Waltham, MA, USA). Enumeration of *Prochlorococcus*, *Synechococcus*, picoeukaryotes, and unpigmented cells was conducted via flow cytometry using a collinear analyses method and a UV wavelength of 488 nm. Unpigmented cells were stained with Hoeschst stain at a final concentration of 1 μg ml^{-1}.

Untargeted analysis of metabolites

Untargeted metabolite analysis was performed using UPLC coupled with an ultrahigh resolution quadrupole/linear ion trap/ Orbitrap tribid mass spectrometer (Orbitrap Fusion Lumos, Thermo Scientific™). A Waters Acquity HSS T3 column (2.1 × 100 mm, 1.8 μm), equipped with a Vanguard pre-column, was used for chromatographic separation at 40 °C. The column was eluted first with A) 0.1% formic acid in water followed by B) 0.1% formic acid in acetonitrile with a flow rate of 0.5 mL min^{-1}. The chromatographic gradient was as follows: 1% B (1 min), 15% B (1-3 min), 50% B (3-6 min), 95% B (6-9 min), 95% B (10 min). The column was washed and re-equilibrated with 1% B (2 min) between injections. Individual autosampler injections (5 μL each) were taken for positive and negative mode analyses. The electrospray voltage was set to 3600 V for positive mode and 2600 V for negative mode. Settings for source gases were 55 (sheath) and 20 (auxiliary) and these settings are presented in arbitrary units. The heated

capillary temperature was 375 °C and the vaporizer temperature was 400 °C. Full scans of the

mass spectrometry (MS) data were collected in the Orbitrap analyzer with a mass resolution of

120,000 FWHM at m/z 200. The automatic gain control (AGC) target was 4e5, the maximum

injection time was 50 seconds, and the scan range was 100 – 1000 m/z. Data-dependent MS/MS

data were acquired in the Orbitrap analyzer using higher energy collisional dissociation (HCD)

with a normalized collision energy of 35%. The AGC target value for fragmentation spectra was

5e4 and the intensity threshold was 2e4. Isolation within the quadropole was set at 1 m/z. All

data were collected in profile mode. Pooled samples were run after every 6 samples to assess

instrument performance and drift over the sample run.

After the instrument run, files were obtained from the instrument and converted into

MZml files using msConvert and then processed using XCMS (Smith et al. 2006). Peak-picking

was performed using the CentWave algorithm and a Guassian fit with the following parameters:

noise = 10000, peak-width = 3-15, ppm = 15, prefilter = 4 -10000, integrate = 2, mzdiff = -0.005,

snthresh = 100. Retention times were then adjusted using the Obiwarp and correspondence

between the peaks was conducted. There was good agreement among the untargeted pooled

samples and these were removed from further analyses. Untargeted peak intensities were

normalized by dividing the peak intensities by the total sample volume as well as the DOC

concentration for each sample. In the untargeted analyses, metabolite features are defined as a

unique combination of mass to charge ratio and retention time combination.

Targeted analysis of metabolites

Extracts prepared for the targeted method were run using ultra performance liquid

chromatography (UPLC) (Accela Open Autosampler and Accela 1250 Pump, Thermo

Scientific™) coupled to a heated electrospray ionization source (H-ESI) and a triple stage

quadrupole mass spectrometer (TSQ Vantage, Thermo Fisher Scientific™, Waltham, MA, USA),

operated in selective reaction monitoring (SRM) mode. The same UPLC column, column

conditions, gradient, and flow rates were used for both targeted and untargeted analysis.

Autosampler injections of 5 μl each were made separately for positive and negative modes.

Samples were analyzed in a random order and pooled samples were run between every 6 samples

to assess instrument drift. SRM parameters were optimized for each target compound using an

authentic standard as described in Kido Soule et al. (2015) and two transitions were monitored

for quantification and confirmation. Target metabolites included a range of environmentally

relevant vitamins, amino acids, and other metabolites (Kido Soule et al. 2015). Calibration

curves based on peak area were generated for each compound (8 points). Data were then

converted into mzML files using msConvert (Chambers et al. 2012) and processed with MAVEN

(Melamud et al. 2010). Environmental concentrations of metabolites were determined by volume

correcting the concentrations for metabolites detected in the extracts that met the threshold limits

of detection and quantification for the PPL column (see Johnson et al. 2017). Finally,

concentrations were corrected using extraction efficiencies that have been reported for each

metabolite in seawater (Johnson et al. 2017).

Statistical analyses

We used principal components analysis (PCA) to visualize dominant patterns in the

environmental data (macronutrient, chlorophyll *a*, and phaeophytin concentrations, picoplankton

cell abundances, and physicochemical measurements) across reef locations in JR and the

bluewater sites using the R package 'FactoMineR' (Le et al. 2008). Significant differences in

macronutrient, chlorophyll *a*, and phaeophytin concentrations, and cell abundances were

investigated using several statistical tests. The Shapiro-Wilk test was used to investigate data

normality. If data were normally distributed, an ANOVA test followed by a Tukey's honestly significant difference (HSD) test was used. If data was not normally distributed, a Kruskal-Wallis test followed by Dunn's test was used. Non-metric multidimensional scaling analysis (NMDS) was performed on the Bray – Curtis dissimilarity matrix yielded from square – root transformed untargeted metabolite feature peak intensities using the function 'metaMDS' from the Vegan R package (Oksanen et al. 2017). We used the function 'indval' from the R package labdsv (Roberts 2016) to perform indicator metabolite analysis on normalized untargeted feature peak intensities. We conducted this test to identify features significantly ($p < 0.05$) indicative of either surface or reef-depth reef seawater.

Linear regressions were used to test for relationships between the averaged untargeted feature peak intensities by site and depth and microbial cell abundances as well as inorganic and organic macronutrient concentrations. An ANOVA was completed on the linear model fit for each test and then p-values were adjusted using the false discovery rate (FDR).

Overall patterns in targeted metabolite concentrations were investigated and a subset of these metabolites were chosen for further analyses. Pearson correlations were conducted between targeted metabolites that could be quantified (adjusted p-value < 0.05, Benjamini-Hochberg test) to assess co-occurrence between targeted metabolites. Linear regressions (as described above) were completed between targeted metabolites and cell count and nutrient data. Targeted metabolite concentration distributions were tested for normality using the Shapiro-Wilks test. ANOVA followed by Tukey's HSD tests were used to identify significant differences in metabolite concentrations by depth or reef location in data that was normally distributed ($p < 0.05$). For non-parametric testing, Wilcoxon rank sum tests were used to identify significant differences between two factors ($p < 0.05$). Samples were also categorized into bluewater (BW1

and BW2), eastern (JR 2, 4, 5), central (JR 6, 10, 11), and western (JR 11, 12, 13, 14) forereef

groupings in order to examine broader patterns in metabolite concentrations across the

archipelago. Lastly, we conducted linear regressions (as explained above) between specific

targeted metabolites of interest with untargeted features in order to detect if a suite of untargeted

features followed the same trends.

6.4 Results

Reef composition, biogeochemistry, and cell abundances

Reef composition was similar across forereefs (Figure 2). Coral cover ranged from 11% to 35%

with an average cover of 26% and JR2 and JR12 had the lowest coral cover (Figure 2). The

average cover of macroalgae was 31% and ranged from 24 to 41% (Figure 2). Sponge cover was

similarly low (4% average cover) across all surveyed reefs (Figure 2). Forereef bare substrate

was mostly rock (average cover of 30%) and sand cover varied between forereefs (Figure 2).

Physicochemical conditions in surface and reef-depth seawater were similar across all the

forereefs (average temperature: 28.49 ± 0.29 (S.D.) °C, average pH: 8.25 ± 0.05, average

salinity: 37.07 ± 0.17 psu, and average dissolved oxygen: 6.45 ± 0.52 mgL^{-1}) (Table S1). Total

organic carbon (TOC), dissolved organic carbon (DOC), phosphate, and ammonium were similar

across all reefs and reef-depths (Table S1). Total nitrogen concentrations were significantly

higher in reef-depth seawater (3.4 ± 1.6 µM) compared to bluewater seawater (0.68 ± 0.90 µM)

(Kruskal-Wallis test, Chi^2 = 5.5974, df = 2, p = 0.06; Dunn's test, adjusted p = 0.03).

Additionally, nitrite + nitrate concentrations were significantly higher in reef-depth seawater

(0.23 ± 0.11 µM) compared to bluewater seawater (0.025 ± 0.05 µM) (Kruskal-Wallis test, Chi^2=

9.2507, df = 2, p = 0.01; Dunn's test, adjusted p = 0.004). When tested separately, nitrite

concentrations were higher in bluewater (0.055 ± 0.017 µM) compared to reef surface and reef-depth seawater (0.036 ± 0.01 µM), but the trends were not significant (Kruskal-Wallis test, $Chi^2 = 3.9435$, df = 2, p = 0.1392). Additionally, nitrate did not display significant differences by sample-type (i.e. bluewater, reef-depth, reef surface) (ANOVA, $F_{(2,19)} = 1.3066$, p > 0.05).

While the abundance of unpigmented cells (generally heterotrophic bacteria and archaea) did not vary widely across locations (average of $413,925 \pm 77,745$ cells mL^{-1}), *Prochlorococcus* cell abundances were significantly higher in bluewater (average of $54,397 \pm 5,665$ cells mL^{-1}) compared to reef-depth seawater (average of $36,177 \pm 10,631$ cells mL^{-1}) (ANOVA, $F_{(2,19)} = 6.1045$, p = 0.009; Tukey's HSD, adjusted p = 0.009). In contrast, *Synechococcus* cell abundances were on average higher in reef seawater (average of $34,766 \pm 16,100$ cells mL^{-1}) compared to bluewater seawater (average of $19,111 \pm 3,378$ cells mL^{-1}), but not significantly higher due to local variability across reefs (ANOVA, $F_{(2,19)} = 2.061$, p = 0.1549). Picoeukaryotic cell abundances followed the same pattern as *Synechococcus* and were higher in reef seawater (average of 890 ± 548 cells mL^{-1}) compared to bluewater seawater (average of 471 $\pm$ 83 cells mL^{-1}). Chlorophyll *a* and phaeophytin concentrations were generally similar across reefs (average concentrations of 0.14 ± 0.05 and 0.087 ± 0.04 µgL^{-1}, respectively), excluding JR10 and JR12 which had much higher concentrations of chlorophyll *a* (0.89 ± 0.07 µgL^{-1}) and phaeophytin (1.05 ± 0.14 µgL^{-1}).

Overall patterns in environmental data were investigated using PCA (Figure 3). Bluewater locations grouped together, indicating their similarity, and the vector representing *Prochlorococcus* cell abundance was oriented in their direction (Figure 3). Surface and reef-depth seawater were distributed across the PCA and did not group together by depth. Reef-depth seawater from JR11, 10, and 12 emerged as outliers (Figure 3). Total chlorophyll *a* and

phaeophytin were strongly correlated with each other, oriented in the direction of JR12, and contributed the most variation to principal component one (total chlorophyll a correlation = 0.95, p = $1.08e^{-10}$, phaeophytin correlation = 0.94, p = $6.18e^{-10}$; Figure 3). Silicate concentrations contributed the most variation to principal component two (correlation = 0.85, p = $1.19e^{-5}$; Figure 3).

Untargeted metabolites

Overall, 1794 metabolite features were detected using the untargeted method in negative mode and the average number of features detected across all samples was 1791 ± 6. Untargeted metabolite feature composition was highly similar across reefs and samples did not separate clearly by location, depth, or biome (bluewater vs. forereef) (Figure 4). That being said, reef-depth and surface seawater from JR12 and surface seawater from JR10 were outliers in the NMDS, indicating larger differences in metabolite composition in these samples compared to all the other samples (Figure 4).

Dissolved nitrogen (158 features), nitrite + nitrate (107 features), nitrate (127 features), and ammonium (369 features) concentrations significantly regressed with untargeted metabolite features (adjusted p-value < 0.05). There were no significant pair-wise relationships between untargeted metabolite features and DOC, TOC, TN, phosphate, silicate, or nitrite concentrations. A few metabolite features significantly regressed with *Prochlorococcus* (3 features) or *Synechococcus* (1 feature) cell abundances, but not with picoeukaryotic or unpigmented (generally bacteria and archaea) cell abundances.

Indicator feature analysis was conducted to reveal features that were significantly indicative of reef seawater (surface and reef-depth), bluewater, surface reef seawater, and reef-depth seawater based on their frequency of detection and peak intensity. In a comparison

between reef seawater and bluewater, there were 76 features indicative of reef seawater (surface

and reef-depth) and 171 features indicative of bluewater seawater. Across reef seawater samples,

41 features were indicative of reef-depth seawater and 39 features were indicative of surface reef

seawater (Figure 5).

Targeted metabolites

Using targeted metabolomics, we detected 33 metabolites and quantified the environmental

concentrations of 22 metabolites, using determined extraction efficiencies reported by Johnson et

al. (2017) (Table S2). Among these metabolites, we found 6 significant (adjusted p-value < 0.05,

Benjamini-Hochberg corrections) co-correlations between detected and quantified metabolites.

The strongest positive correlation was found between adenosine and guanosine (Pearson

correlation = 0.831 adjusted p-value = 0.002). Tyrosine and phenylalanine were also strongly

correlated (Pearson correlation = 0.794, adjusted p-value = 0.008), followed by

methylthioadenosine (MTA) and guanosine (Pearson correlation = 0.756 adjusted p-value =

0.016), riboflavin and taurochlorate (Pearson correlation = 0.754, adjusted p-value = 0.016),

phenylalanine and tryptophan (Pearson correlation = 0.734, adjusted p-value = 0.025), and MTA

and taurochlorate (Pearson correlation = 0.707, adjusted p-value = 0.048). Most (85%) of the

identified metabolites were detected in both bluewater and reef seawater (reef-depth and surface)

except for adenosine 5' – monophosphate, glyphosate, guanosine, pantothenic acid, and

xanthine, which were only detected in reef seawater (Table S2). All quantified targeted

metabolites had nanomolar or picomolar concentrations in reef and bluewater seawater (Table

S2). Malic acid was present at the highest concentration, ranging from 9,534.79 – 27, 839.39 pM

(Table S2).

Linear regressions of targeted metabolites with nutrient and cell abundances revealed several significant relationships. Riboflavin concentrations regressed significantly with nitrate concentrations (adjusted p-value = 0.028) and MTA concentrations regressed significantly with nitrite concentrations (adjusted p-value = 0.021). Seven targeted compounds, including MTA, adenosine, desthiobiotin, guanosine, pantothenic acid, riboflavin, and taurochlorate, regressed significantly with *Prochlorococcus* abundances. Only one compound, S-adenosyl-L-homocysteine, significantly regressed with *Synechococcus* abundances.

Riboflavin and MTA concentrations exhibited strong biogeographical patterns across Jardines de la Reina (Figure 6A, C). Riboflavin concentrations were significantly higher (Wilcoxon rank sum test, W = 370, p-value = 2.046e-07) in reef-depth seawater (median concentration = 4.9 pM) compared to surface reef seawater (median concentration = 1.9 pM) (Figure 6B). Additionally, median riboflavin concentrations were slightly elevated in reef-depth seawater collected from the central forereefs (median concentration = 4.4 pM) compared to the eastern (median concentration = 3.2 pM) and western (median concentration = 2.9 pM) forereefs, but this difference was not statistically significant. However, the concentration of riboflavin in reef-depth seawater was significantly higher compared to bluewater surface seawater (Kruskal-Wallis, Chi^2 = 27.768, df = 6, p-value = 0.0001; Dunn's test, p < 0.05). The normalized peak intensities of 129 untargeted features significantly regressed with riboflavin concentrations. MTA concentrations were significantly elevated (ANOVA, F(3,36) = 37.019, p < 0.05; Tukey's HSD, adjusted p-value <0.05) in central forereef seawater compared to bluewater, eastern and western forereef seawater (6D). MTA concentrations did not differ significantly between surface and reef-depth seawater (ANOVA, F(1,38) = 2.2109, p = 0.1453). Lastly, 63 untargeted features significantly regressed with MTA concentrations.

6.5 Discussion

This study surveyed dissolved organic metabolites present in forereef seawater from the protected Jardines de la Reina coral reef archipelago in an effort to provide baseline information about the extracellular metabolite composition of coral reef seawater. Using both targeted and untargeted approaches, we have quantified specific metabolites including nucleosides, amino acids, vitamins, and metabolic intermediates as well as observed that untargeted metabolite feature composition across the archipelago was quite similar between reef-depth and surface seawater as well as bluewater seawater. We also investigated if the metabolite feature landscape changed in concert with reef composition and bulk biogeochemical and physical measurements and found similar patterns with a few subtle differences.

Similar metabolite compositions across the Jardines de la Reina Archipelago

Untargeted metabolite composition and diversity was generally similar across all sampled biomes and depths, likely reflecting homogenous reef compositions and environmental conditions across the Jardines forereefs. The other measurements generally followed suit, with the exception of total nitrogen and nitrate + nitrite concentrations (higher on reefs) and significantly higher abundances of *Prochlorococcus* in the off-reef sites (bluewater seawater). The metabolite feature composition of seawater from sites JR10 and JR12 emerged as outliers and these observations, together with the elevated total chlorophyll *a* and phaeophytin concentrations in this seawater, may suggest a phytoplankton bloom. The biogeochemical and physicochemical similarities across the forereefs are likely due to their location, depth, and hydrography. The forereefs of Jardines de la Reina are adjacent to deep, pelagic environments

and are flushed with oligotrophic water via the Caribbean current that meanders through the Caribbean basin.

Presence of amino acids and nucleosides in reef and bluewater seawater

Targeted metabolite analysis revealed the presence of 33 known extracellular metabolites in reef seawater, some of which could be quantified at their nM or pM environmental concentrations. Most (85%, presence/absence) of these metabolites could be detected in both bluewater and reef seawater, consistent with the observation that the forereefs in Jardines are flushed with oligotrophic seawater from offshore. We were able to quantify five nucleosides (adenosine, guanosine, inosine, thymidine, and xanthosine) and three out of four amino acids (tyrosine, phenylalanine, and tryptophan). Fiore et al. (2017) reported higher concentrations of nucleosides and the amino acid tryptophan in sponge exhalent seawater compared to inhalant seawater, demonstrating that sponges can be sources of amino acids and nucelosides to the surrounding seawater. Additionally, corals (both stony and soft) are capable of releasing dissolved free amino acids into the water column, either via their mucus (Ducklow and Mitchell 1979) or dissolved exudation (Schlichter and Liebezeit 1991). There were several co-correlations between the concentrations of various nucleosides (e. g. adenosine and guanosine) as well as essential (phenylalanine) and non-essential (tyrosine) amino acids.

Depth-related and geographical patterns in Riboflavin and MTA distributions

We found interesting depth-related trends in the riboflavin concentration distributions, with riboflavin being higher in reef-depth seawater compared to surface reef and bluewater seawater. Incubation- and field-based experiments have demonstrated that corals (Dunlap and Susic 1985) and sponges (Fiore et al. 2017) can be sources of riboflavin to the surrounding reef seawater. Within the cell, riboflavin (vitamin B2) is a required precursor for cofactors used in oxidative

metabolism (Abbas and Sibirny 2011; Gutierrez-Preciado et al. 2015). Extracellularly, riboflavin

can also be used by bacteria as a quorum-sensing (Rajamani et al. 2008), electron-transfer

(Marsili et al. 2008), and signal molecule in plant-microbe systems (Dakora et al. 2015).

Riboflavin degrades via ultraviolet oxidation into lumichrome or lumiflavin (Dunlap and Susic

1986), representing a potential abiotic loss of the molecule from the system in high light and

shallow environments like the forereefs in Jardines de la Reina. Our targeted data confirms that

extracellular riboflavin is present on reefs and demonstrates that concentrations of riboflavin are

higher closer to the reef, indicating that the reef is a source of riboflavin. Additionally, riboflavin

concentrations were lowest in bluewater surface seawater, providing more evidence that reefs

may be a source of extracellular riboflavin to the surrounding seawater. We hypothesize that the

decrease in riboflavin concentrations in surface seawater is due to both biological uptake by

planktonic organisms and ultraviolet degradation, but this hypothesis needs to be investigated

further.

MTA exhibited interesting geographic changes in concentration, displaying humplike

patterns with maximum concentrations of these metabolites in the center of the archipelago.

MTA is an intermediate metabolite that is formed when the S-adenosyl-l-methionine (SAM)

cosubstrate undergoes various reactions within the cell (e.g. to form polyamines, ethylene

precursors for plants, quorum sensing compounds, phytosiderophores, and betaine lipids) and

accumulation of MTA within the cell can cause cell death (reviewed within Miller et al. 2018).

In freshwater systems or environments with low supply of bioavailable sulfur, MTA can be

scavenged for use of the sulfur atom (Miller et al. 2018). The role of extracellular MTA in

marine environments is unknown, but there is evidence that intra- and extracellular MTA

accumulation was linked with the synthesis of an acyl homoserine lactone (AHL) when the

marine bacterial isolate *Ruegeria pomeroyi* DSS-3 was grown with dimethylsulfoniopropionate

(DMSP) as its sole carbon source (Johnson et al. 2016). The finding that MTA exhibits this

geographic distribution within extracellular reef seawater provides preliminary evidence that

MTA is an important metabolite on these reefs and that its production and loss need to be

investigated in order to understand the processes behind its interesting geographic distribution.

Future directions

This work has provided fundamental knowledge about the extracellular metabolite feature

inventory across the protected Jardines de la Reina reef-system and has opened up many

questions pertaining to metabolic cycling and benthic-pelagic coupling on reefs. For one, it

would be useful to identify the untargeted metabolite features that were indicative of surface or

reef-depth seawater. This could be done if corresponding MS2 scans were available for these

features, but is beyond the scope of this dissertation chapter. Regardless of these indicative

features, this study found that the overall untargeted metabolic feature composition was similar

across surface and reef-depths as well as between bluewater and reef seawater environments, but

did not survey potential biological variability of metabolite composition across individual reefs

by collected biologically independent samples in different locations adjacent to the reef

substrate. Future work should address this by collecting >3 samples at each reef location and

subjecting these samples to both targeted and untargeted analyses in order to investigate if

specific benthic organisms like corals have distinct extracellular metabolite signatures. At a

larger comparative scale, this approach can also be used to understand if reef metabolomes

reflect the overall quality or health of the reef. We were not able to assess this question in our

current study due to very similar reef and environmental conditions across the archipelago, but a

study that compared reef metabolites collected from different reefs across reef-systems and/or

along environmental and anthropogenic gradients may be able to detect associations between metabolite compositions and reef health.

Additionally, samples should be collected at other marine biomes, like seagrass or mangrove habitats, in order to tease apart if, why, and how the extracellular metabolite composition differs and if there are signature metabolite profiles that define each biome. Metabolite samples could also be collected over diel or daily time-series in order to track how metabolite composition shifts in the presence/absence of light or through time concomitant with shifts in environmental conditions. Secondly, this work should be integrated with genomic analyses of the microbial community so that the potential functional roles of these communities can be cross-checked with metabolic evidence as well as to generate concrete hypotheses about planktonic metabolic cycling in coral reef environments (e. g. understanding sources and sinks of MTA). Lastly, continuing efforts to increase the retention of more molecules from seawater during the extraction step will help improve our understanding of the extracellular metabolite composition of reef seawater. The PPL cartridge used for SPE has been reported to extract 43 – 62% of the DOC in seawater (Dittmar et al. 2008) and performs better than other extraction columns, but still fails to recover the smallest and most polar metabolites (Johnson et al. 2017).

Here we have provided the first, comprehensive survey of specific extracellular metabolites, their concentrations, and the overall extracellular metabolite feature composition within seawater from the protected Caribbean fore reefs of Jardines de la Reina, Cuba. This exploratory work investigates trends between reef metabolites and other bulk biogeochemical parameters as well as reef composition and quality and lays the groundwork and the methodological framework for future investigations of extracellular reef metabolites.

6.6 Acknowledgements

6.7 References

Abbas, C.A., and Sibirny, A.A. (2011) Genetic control of biosynthesis and transport of riboflavin and flavin nucleotides and construction of robust biotechnological producers. *Microbiol Mol Biol Rev* **75**: 321-360.

Apprill, A., and Rappé, M.S. (2011) Response of the microbial community to coral spawning in lagoon and reef flat environments of Hawaii, USA. *Aquat Microb Ecol* **62**: 251-266.

Atkinson, M.J. (2010) Biogeochemistry of nutrients. In *Coral reefs an ecosystem in transition.* Dubinsky, Z., and Stambler, N. (eds): Springer Science & Business Media, pp. 199-206.

Azam, F., and Malfatti, F. (2007) Microbial structuring of marine ecosystems. *Nat Rev Microbiol* **5**: 782-791.

Chambers, M.C., Maclean, B., Burke, R., Amodei, D., Ruderman, D.L., Neumann, S. *et al.* (2012) A cross-platform toolkit for mass spectrometry and proteomics. *Nat Biotechnol* **30**: 918-920.

Dakora, F.D., Matiru, V.N., and Kanu, A.S. (2015) Rhizosphere ecology of lumichrome and riboflavin, two bacterial signal molecules eliciting developmental changes in plants. *Front Plant Sci* **6**: 700.

de Goeij, J.M., van Oevelen, D., Vermeij, M.J.A., Osinga, R., Middelburg, J.J., de Goeij, A.F.P.M., and Admiraal, W. (2013) Suriving in a marine desert: the sponge loop retains resources within coral reefs. *Science* **342**: 108 - 110.

Dittmar, T., Koch, B., Hertkorn, N., and Kattner, G. (2008) A simple and efficient method for the solid-phase extraction of dissolved organic matter (SPE-DOM) from seawater. *Limnology and Oceanography: Methods* **6**: 230-235.

Ducklow, H.W., and Mitchell, R. (1979) Composition of mucus released by coral reef coelenterates. *Limnol Oceanogr* **24**: 706-714.

Dunlap, W.C., and Susic, M. (1985) Determination of pteridines and flavins in seawater by reverse phase, high-performance liquid chromatography with fluorometric detection. *Marine Chemistry* **17**: 185-198.

Dunlap, W.C., and Susic, M. (1986) Photochemical decomposition rates of pteridines and flavins in seawater exposed to surface solar radiation. *Marine Chemistry* **19**: 99-107.

Fiore, C.L., Freeman, C.J., and Kujawinski, E.B. (2017) Sponge exhalent seawater contains a unique chemical profile of dissolved organic matter. *PeerJ* **5**: e2870.

Gutierrez-Preciado, A., Torres, A.G., Merino, E., Bonomi, H.R., Goldbaum, F.A., and Garcia-Angulo, V.A. (2015) Extensive Identification of Bacterial Riboflavin Transporters and Their Distribution across Bacterial Species. *PLoS One* **10**: e0126124.

Haas, A.F., Nelson, C.E., Wegley Kelly, L., Carlson, C.A., Rohwer, F., Leichter, J.J. *et al.* (2011) Effects of coral reef benthic primary producers on dissolved organic carbon and microbial activity. *PLoS One* **6**: e27973.

Haas, A.F., Nelson, C.E., Rohwer, F., Wegley-Kelly, L., Quistad, S.D., Carlson, C.A. *et al.* (2013) Influence of coral and algal exudates on microbially mediated reef metabolism. *PeerJ* **1**: e108.

Haas, A.F., Fairoz, M.F., Kelly, L.W., Nelson, C.E., Dinsdale, E.A., Edwards, R.A. *et al.* (2016) Global microbialization of coral reefs. *Nat Microbiol* **1**: 16042.

Hansell, D.A., and Carlson, C.A. (2001) Biogeochemistry of total organic carbon and nitrogen in the Sargasso Sea: control by convective overturn. *Deep-Sea Res II* **48**: 1649-1667.

Hatcher, B.G. (1988) Coral reef primary productivity: a beggar's banquet. *Trends Ecol Evol* **3**: 106-111.

Johnson, W.M., Kido Soule, M.C., and Kujawinski, E.B. (2016) Evidence for quorum sensing and differential metabolite production by a marine bacterium in response to DMSP. *ISME J* **10**: 2304-2316.

Johnson, W.M., Kido Soule, M.C., and Kujawinski, E.B. (2017) Extraction efficiency and quantification of dissolved metabolites in targeted marine metabolomics. *Limnology and Oceanography: Methods* **15**: 417-428.

Kelly, L.W., Haas, A., and Nelson, C.E. (2018) Ecosystem microbiology of coral reefs: linking genomic, metabolomic, and biogeochemical dynamics from animal symbioses to reefscape processes. *mSystems* **3**: e00162-00117.

Kido Soule, M.C., Longnecker, K., Johnson, W.M., and Kujawinski, E.B. (2015) Environmental metabolomics: Analytical strategies. *Marine Chemistry* **177**: 374-387.

Kujawinski, E.B. (2011) The impact of microbial metabolism on marine dissolved organic matter. *Ann Rev Mar Sci* **3**: 567-599.

Le, S., Josse, J., and Husson, F. (2008) FactoMineR: an R package for multivariate analysis. *Journal of Statistical Software* **25**: 1-18.

Marsili, E., Baron, D.B., Shikhare, I.D., Coursolle, D., Gralnick, J.A., and Bond, D.R. (2008) Shewanella secretes flavins that mediate extracellular electron transfer. *Proc Natl Acad Sci U S A* **105**: 3968-3973.

Melamud, E., Vastag, L., and Rabinowitz, J.D. (2010) Metabolomic analysis and visualization engine for LC-MS data. *Anal Chem* **82**: 9818-9826.

Miller, A.R., North, J.A., Wildenthal, J.A., and Tabita, F.R. (2018) Two distinct aerobic methionine salvage pathways generate volatile methanethiol in *Rhodopseudomonas palustris*. *Microbiome* **9**: e00407-00418.

298

Nelson, C.E., Goldberg, S.J., Wegley Kelly, L., Haas, A.F., Smith, J.E., Rohwer, F., and Carlson, C.A. (2013) Coral and macroalgal exudates vary in neutral sugar composition and differentially enrich reef bacterioplankton lineages. *ISME J* 7: 962-979.

Obura, D. (2014) Coral reef monitoring manual. *Ébène, Mauritius: Indian Ocean Commission*: 67 pp.

Ochsenkuhn, M.A., Schmitt-Kopplin, P., Harir, M., and Amin, S.A. (2018) Coral metabolite gradients affect microbial community structures and act as a disease cue. *Commun Biol* 1: 184.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D. *et al.* (2017) vegan: community ecology package. R package version 2.4-5.

Quinn, R.A., Vermeij, M.J., Hartmann, A.C., Galtier d'Auriac, I., Benler, S., Haas, A. *et al.* (2016) Metabolomics of reef benthic interactions reveals a bioactive lipid involved in coral defence. *Proc Biol Sci* 283.

Rajamani, S., Bauer, W.D., Robinson, J.B., Farrow, J.M., 3rd, Pesci, E.C., Teplitski, M. *et al.* (2008) The vitamin riboflavin and its derivative lumichrome activate the LasR bacterial quorum-sensing receptor. *Mol Plant Microbe Interact* 21: 1184-1192.

Roberts, D.W. (2016) labdsv: ordination and multivariate analysis for ecology. *R package version 18-0.*

Schlichter, D., and Liebezeit, G. (1991) The natural release of amino acids from the symbiotic coral *Heteroxenia fuscescens* (Ehrb.) as a function of photosynthesis. *J Exp Mar Bio Ecol* 150: 83-90.

Smith, C.A., Want, E.J., O'Maille, G., Abagyan, R., and Siuzdak, G. (2006) XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem* 78: 779-787.

Sogin, E.M., Putnam, H.M., Anderson, P.E., and Gates, R.D. (2016) Metabolomic signatures of increases in temperature and ocean acidification from the reef-building coral, Pocillopora damicornis. *Metabolomics* 12.

Sogin, E.M., Putnam, H.M., Nelson, C.E., Anderson, P., and Gates, R.D. (2017) Correspondence of coral holobiont metabolome with symbiotic bacteria, archaea and Symbiodinium communities. *Environ Microbiol Rep* 9: 310-315.

Tanaka, Y., Miyajima, T., Umezawa, Y., Hayashibara, T., Ogawa, H., and Koike, I. (2009) Net release of dissolved organic matter by the scleractinian coral *Acropora pulchra*. *J Exp Mar Biol Ecol* 377: 101-106.

Vohsen, S.A., Fisher, C.R., and Baums, I.B. (2019) Metabolomic richness and fingerprints of deep-sea coral species and populations. *Metabolomics* 15: 34.

Wild, C., Rasheed, M., Werner, U., Franke, U., Johnstone, R., and Huettel, M. (2004) Degradation and mineralization of coral mucus in reef environments. *Mar Ecol Prog Ser* **267**: 159-171.

6.8 Figures
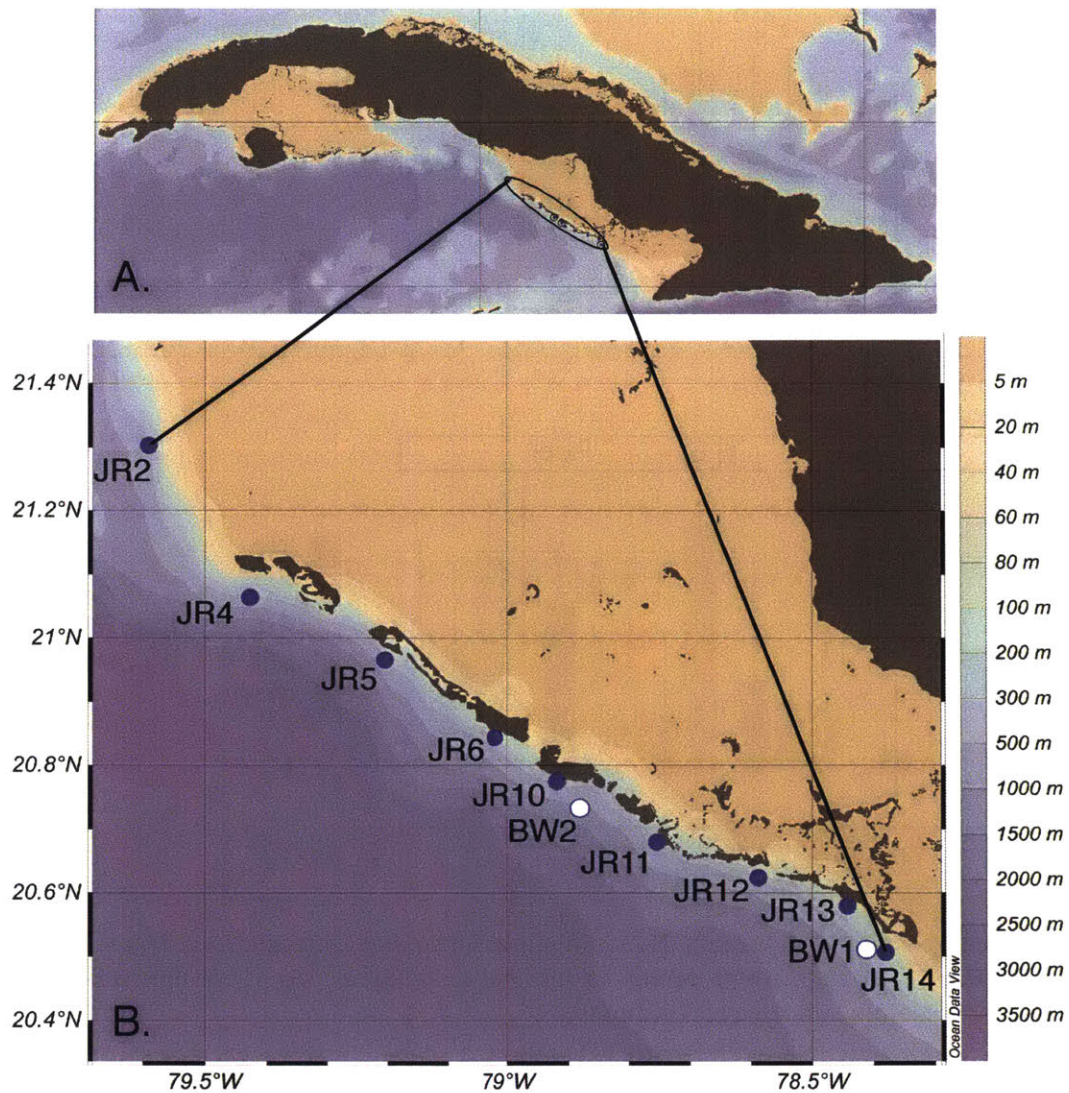
Figure 1. A) Overview of the encircled Jardines de la Reina (JR) archipelago adjacent to Cuba. B) Coral reef forereef (blue circles) and bluewater (white circles) locations surveyed across the archipelago. Symbols represent sites and are labeled with the site name. BW = bluewater.
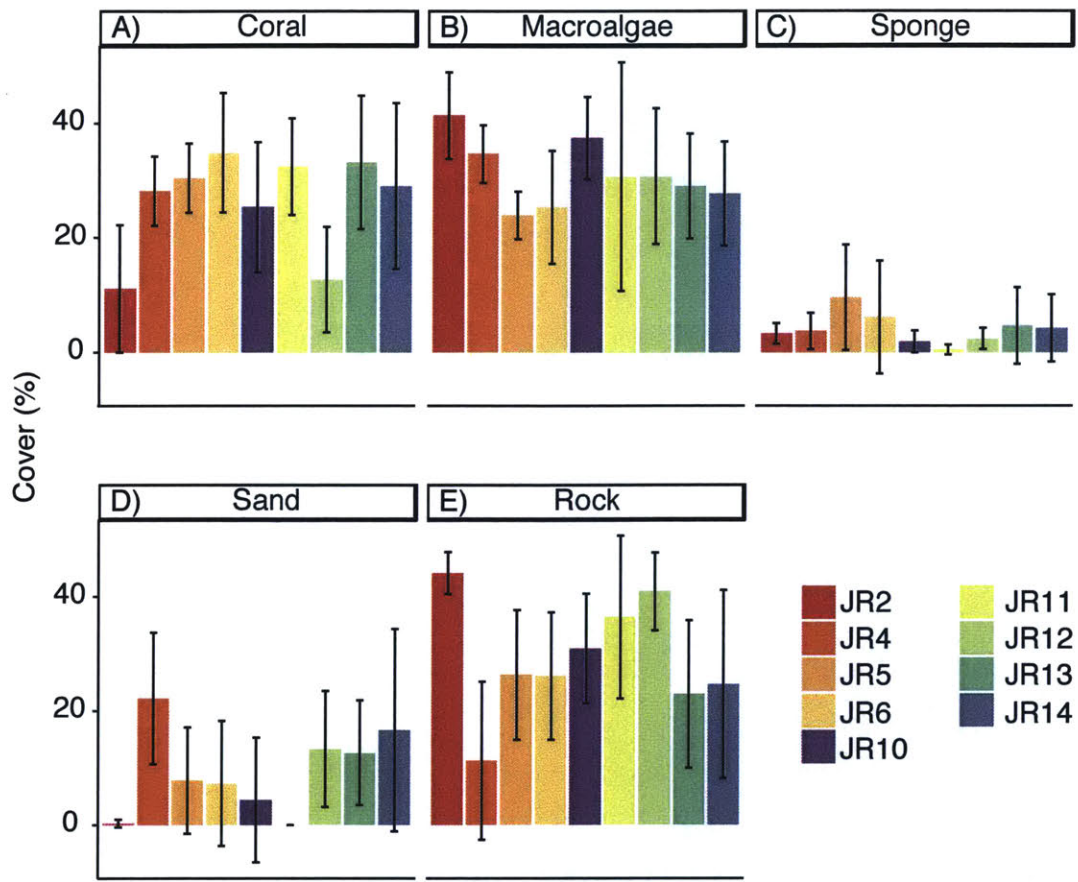
Figure 2. Reef composition of A) corals, B) macroalgae, C) sponges, D) sand, and E) rock measured in average % cover across forereefs in Jardines de la Reina (JR). Error bars reflect standard deviation. Colors indicate reef locations.
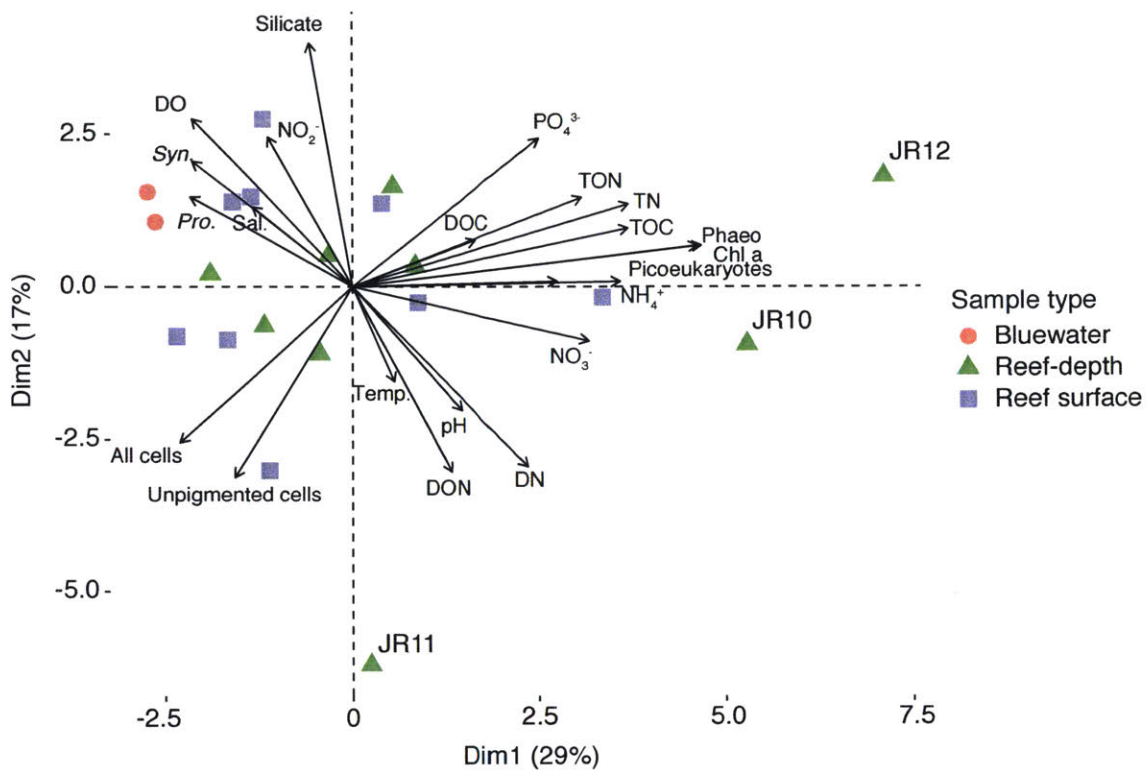
Figure 3. Principal components analysis (PCA) performed using physicochemical seawater properties, inorganic and organic nutrient concentrations, chlorophyll and phaeophytin concentrations, and microbial cell abundances. Symbols represent each location and the color represents the depth of collection. Outlier reef locations are labeled. Phaeo = phaeophytin, Sal. = salinity, Chl a = Chlorophyll a, Temp. = temperature, *Syn.* = *Synechococcus* cell abundance, *Pro.* = *Prochlorococcus* cell abundance, DO = dissolved oxygen, TOC = total organic carbon, DOC = dissolved organic carbon, PO_4^{3-} = phosphate, TN = total nitrogen, TON = total organic nitrogen, DON = dissolved organic nitrogen, DN = dissolved nitrogen, NO_2^- = nitrite, NO_3^- = nitrate, NH_4^+ = ammonium.

303
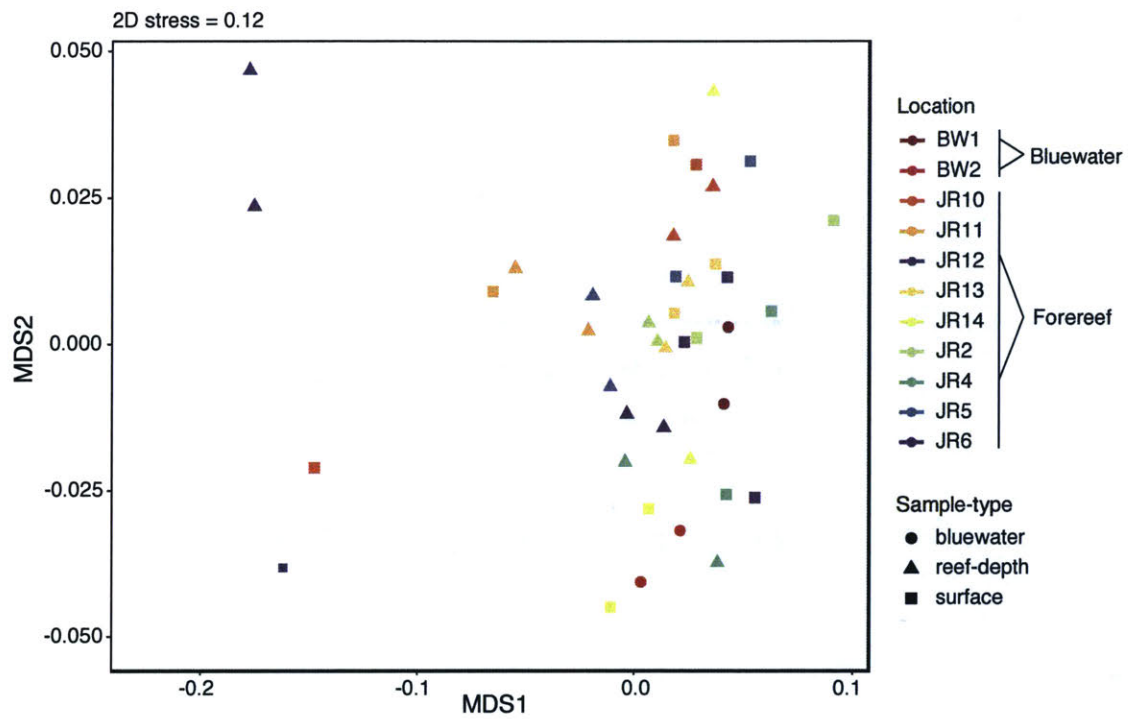
Figure 4. Non-metric multidimensional scaling (NMDS) analysis performed on the Bray-Curtis dissimilarity index of untargeted metabolite features. Symbol colors represent sampling locations and symbol shape indicates sample-type.
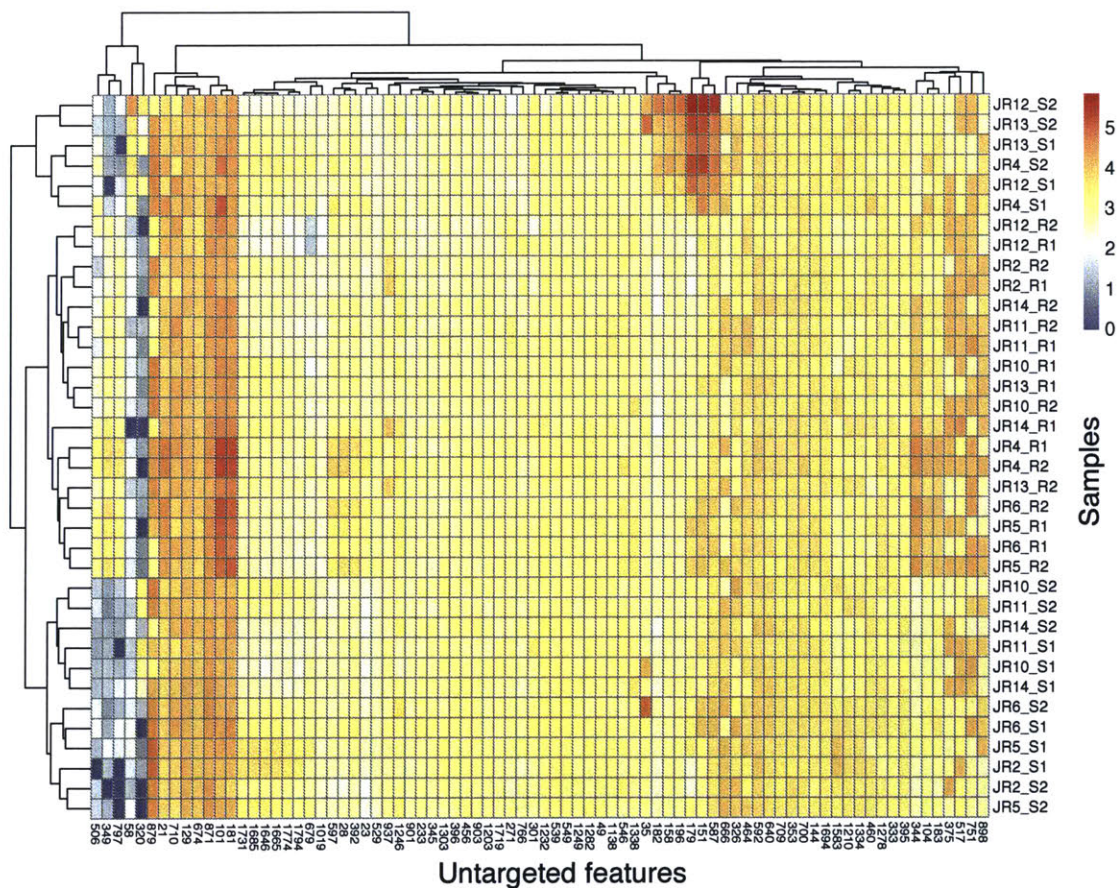
Figure 5. Heatmap of log-transformed untargeted metabolite peak intensities indicative of either reef surface or reef-depth seawater. Warm colors reflect higher peak intensities whereas cooler colors represent lower peak intensities. Surface samples are indicated by S1 or S2 and reef-depth samples are indicated by R1 or R2, with 1 and 2 from each depth representing technical replicates.
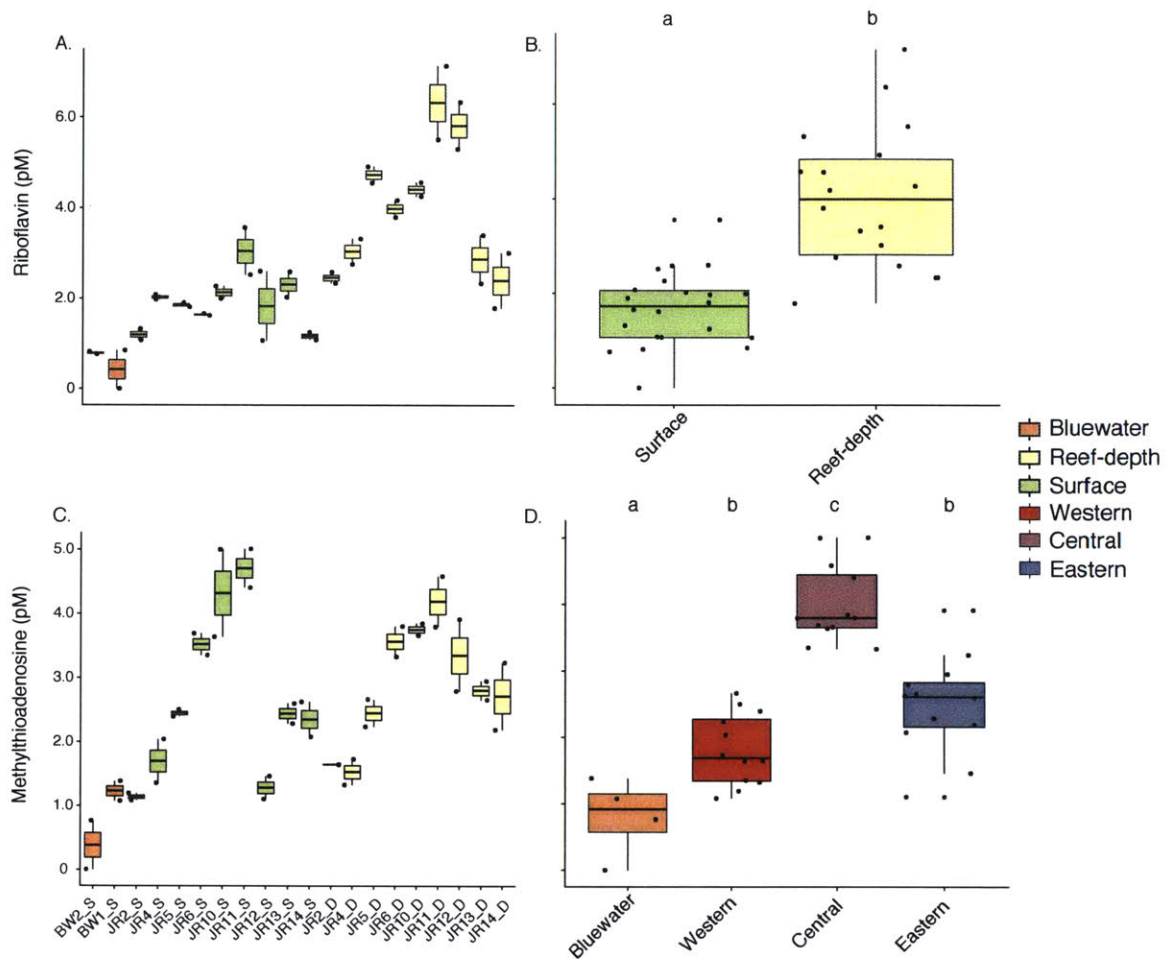
Figure 6. Concentrations of riboflavin (A, B) and methylthioadenosine (MTA) (C, D) in reef seawater across reefs in Jardines de la Reina, presented by reef site and depth (A, C), depth irrespective of site (B), and depth within geographical region (D). Boxplots with different letters are significantly different from each other. In panel B, concentrations of riboflavin were significantly higher in reef-depth seawater (Wilcoxon rank sum test, $p < 0.05$) compared to surface seawater. In D, MTA concentrations by geographic location were significantly higher in reef seawater compared to bluewater and highest in seawater collected from central forereefs (ANOVA, Tukey multiple comparison of means, $p < 0.05$). Colors indicate reef seawater type (e.g. bluewater, reef-depth, or surface seawater) and the symbols represent individual concentrations. Boxplots with different letters are significantly different from each other.

6.9 Supporting Information

Table S1. Locations, depths, and physicochemical measurements of the water column across Jardines de la Reina (JR) coral reefs and the blue water (BW) sites.

Date	Site	Type	Latitude	Longitude	Depth types (m)	CTD depth (m)	Temp. (°C)	pH	Salinity (PSU)	DO (mg/L)
11/5/17	JR4b	Reef	21° 03.825	-79° 25.635	Surface	0.38	28.62	8.16	36.91	8.13
					14	10.84	28.65	8.25	36.94	6.48
11/6/17	JR5b	Reef	20° 57.924	-79° 12.273	Surface	0.40	27.75	8.17	37.54	6.76
					12	14.71	28.63	8.26	36.96	6.38
11/7/17	JR6b	Reef	20° 50.634	-79° 01.299	Surface	0.38	28.29	8.15	37.39	6.53
					12	7.46	28.73	8.27	37.05	6.51
11/8/17	JR13b	Reef	20° 34.747	-78° 26.563	Surface	0.35	28.69	8.27	37.04	6.35
					6	10.04	28.64	8.28	37.05	6.23
11/9/17	JR14b	Reef	20° 30.390	-78° 22.864	Surface	0.53	28.78	8.25	37.06	6.45
					8	11.12	28.11	8.29	37.19	6.37
11/9/17	BW1	Blue water	20° 30.680	-78° 24.798	Surface	-	-	-	-	-
11/10/17	JR12b	Reef	20° 37.446	-78° 35.326	Surface	0.39	28.84	8.20	37.07	6.54
					12	12.05	28.65	8.23	37.09	5.93
11/11/17	JR11b	Reef	20° 40.802	-78° 45.275	Surface	0.50	28.68	8.22	37.04	6.34
					12	13.84	28.58	8.26	37.10	5.44
11/11/17	BW2	Blue water	20° 44.017	-78° 52.835	Surface	-	-	-	-	-
11/12/17	JR10b	Reef	20° 46.500	-78° 55.134	Surface	0.35	28.49	8.29	36.98	6.45
					9	9.59	28.54	8.29	36.98	5.93
11/20/17	JR2b	Reef	21° 18.200	-79° 35.464	Surface	0.38	28.13	8.29	36.87	6.72
					11	10.98	28.16	8.31	36.94	6.53

307

Table S2. Detection and concentration ranges (when applicable) of targeted metabolites as well as the sample types and the percentage of sample detection in each category.

Metabolites	Sample-type	% of samples	Concentration range (pM)
2,3-Dihydroxypropane-1-sulfonate*	Bluewater	100	9395.53 – 10554.81
	Surface	100	3449.45 – 15262.06
	Reef-depth	100	4337.62 – 17165.13
4- Aminobenzoic acid	Bluewater	100	26.29 – 50.55
	Surface	100	18.87 – 63.52
	Reef-depth	100	20.07 – 67.68
4- Hydroxybenzoic acid	Bluewater	100	70.73 – 91.23
	Surface	100	47.72 – 139.49
	Reef-depth	100	55.69 – 122.20
5-Methylthioadenosine	Bluewater	75	0 – 1.38
	Surface	100	1.08 – 5.01
	Reef-depth	100	1.33 – 4.58
Adenosine 5'-monophosphate*	Bluewater	0	-
	Surface	0	0 – 686.48
	Reef-depth	50	0 – 532.85
NAD§*	Bluewater	50	0 – 5.80
	Surface	67	0 – 5.66
	Reef-depth	57	0 – 3.16
Adenosine	Bluewater	100	22.30 – 30.86
	Surface	100	23.90 – 58.36
	Reef-depth	100	28.64 – 70.32
Caffeine	Bluewater	100	18.81 – 80.57
	Surface	100	13.94 – 193.85
	Reef-depth	100	14.47 – 30.65
Chitobiose*	Bluewater	75	0 – 1734.30
	Surface	83.3	0 - 3175
	Reef-depth	77.8	0 – 3373.79
Chitotriose*	Bluewater	75	0 – 165.83
	Surface	23	0 – 179.82
	Reef-depth	23	0 – 175.32
Citric acid*	Bluewater	100	1760.74 – 3217.70
	Surface	100	0 – 9401.37
	Reef-depth	100	1547.95 – 6471.09
Desthiobiotin	Bluewater	100	13.15 – 16.22
	Surface	88.9	0 – 16.80
	Reef-depth	72.2	0 – 18.65
Glyphosate*	Bluewater	0	-

	Surface	11.1	0 – 7.00
	Reef-depth	11.1	0 – 10.53
Guanosine	Bluewater	0	-
	Surface	83.3	0 – 31.91
	Reef-depth	83.3	0 – 42.43
Inosine	Bluewater	100	4.95 – 5.94
	Surface	100	4.61 – 14.22
	Reef-depth	100	3.44 – 13.04
Glutathione*	Bluewater	100	183.76 – 416.47
	Surface	100	104.35 – 364.80
	Reef-depth	100	112.76 – 724.34
Kynurenine	Bluewater	100	3.22 – 6.94
	Surface	83.3	0 – 12.75
	Reef-depth	77.8	0 – 6.96
(Iso)Leucine*	Bluewater	75	0 – 712.50
	Surface	94.4	0 – 2388.18
	Reef-depth	94.4	0 – 1447.29
Tyrosine	Bluewater	100	480.66 – 659.26
	Surface	100	478.82 – 1762.52
	Reef-depth	100	444.90 – 1158.48
Malic acid*	Bluewater	100	17657.23 – 24727.60
	Surface	100	11461.47 – 27839.39
	Reef-depth	100	9534.79 – 27322.32
N-acetylglutamic acid	Bluewater	50	0 – 61.72
	Surface	55.6	0 – 69.90
	Reef-depth	55.6	0 – 136.09
N-acetylmuramic acid	Bluewater	100	301.57 – 424.37
	Surface	88.9	0 – 531.80
	Reef-depth	88.9	0 – 629.91
Pantothenic acid	Bluewater	0	-
	Surface	77.8	0 – 6.55
	Reef-depth	88.9	0 – 4.72
Phenylalanine	Bluewater	100	49.87 – 69.64
	Surface	100	30.70 – 394.58
	Reef-depth	100	38.54 – 133.35
Pyridoxine	Bluewater	100	17.60 – 23.30
	Surface	88.3	0 – 24.39
	Reef-depth	66.7	0 – 33.85
Riboflavin	Bluewater	75	0 – 0.85
	Surface	100	1.06 – 3.55

	Reef-depth	100	1.78 – 7.13
s-5'adenosyl -L-homocysteine	Bluewater	100	0.58 – 1.93
	Surface	94.4	0 – 2.79
	Reef-depth	100	0 – 2.45
Taurocholic acid	Bluewater	100	1.61 – 5.12
	Surface	100	5.64 – 32.51
	Reef-depth	100	9.87 – 81.02
Thymidine	Bluewater	25	0 – 11.69
	Surface	72.2	0 – 18.16
	Reef-depth	55.6	0 – 16.71
Tryptamine	Bluewater	50	0 – 21.43
	Surface	50	0 – 20.18
	Reef-depth	50	0 – 19.91
Typtophan	Bluewater	75	0 – 5.76
	Surface	66.7	0 – 16.90
	Reef-depth	88.9	0 – 5.75
Xanthine*	Bluewater	0	-
	Surface	72.2	0 – 655.06
	Reef-depth	66.7	0 – 1054.99
Xanthosine	Bluewater	50	0 – 7.28
	Surface	61.1	0 – 11.12
	Reef-depth	27.8	0 – 6.33

§NAD = Nicotinamide adenine dinucleotide

*Extraction efficiency for these metabolites was determined to be under 2% in Johnson et al. 2017 so these metabolites have not been corrected for extraction efficiency.

Chapter 7

Conclusion

7.1 Conclusions and future directions

Coral reefs are essential ecosystems for a healthy global ocean due to their biological productivity, organismal diversity, and aesthetic beauty. Coral reefs are estimated to sustain up to 25% of all known marine fish species (reviewed within Lavides et al. 2016), leading to the maintenance and replenishment of fish biomass within the ocean (Barbier et al. 2011). In addition to fish, coral reefs host a wide diversity of other benthic and planktonic organisms and microorganisms that are critical components of the food web, responsible for passing energy up the food chain and recycling limiting nutrients in these oligotrophic waters (Odum and Odum 1955). Coral reefs are also inextricably linked to humans due to their generally coastal locations and high productivity, providing food, natural resources, and storm protection, among other ecosystem services, to humans (reviewed within Barbier et al. 2011). In fact, a partial estimate of the worth of U.S. coral reef ecosystems amounts to 3.4 billion dollars per year (Brander and Van Beukering 2013). A separate study found that the annual value of U.S. coral reef flood risk reduction is 1.8 billion dollars (the 2010 dollar) and 18,000 lives annually (Storlazzi et al. 2019).

Coral reefs are also some of the most sensitive ecosystems to global climate change and other direct stressors induced by human activity like overfishing, nutrient pollution, and coastal development. Coral reefs benefit humans, but their very existence is threatened by humans. As a result, coral reef ecosystems have been selected as a target for conservation and restoration. Much of the focus in coral reef restoration has been geared towards microfragmenting and outplanting individual coral colonies on existing reefs, creating marine protected areas that limit human access to reefs, and educating the public about the benefits of, and threats to, coral reefs. However, due to the complexity of these systems, as well as the lack of baseline monitoring of

coral reefs prior to the onset of the aforementioned stressors, we have limited understanding of how reef ecosystems function, especially from a microbial perspective.

Microorganisms are present on coral reefs in multitudes, living within benthic reef organisms, reef sediment, and the surrounding seawater, and are critical components of these ecosystems for nutrient recycling. The functions and compositions of these complex communities likely play important roles in successful coral restoration efforts. However, coral reef microbial ecology is rarely factored into reef restoration management plans because there is not enough information about how reef microbial communities change in response to various environmental and anthropogenic gradients, their variation over time and space, or how they interact with benthic marine organisms like corals, especially for reef ecosystems within the Caribbean. The overarching goal of this dissertation was to characterize reef seawater microbial community diversity, composition, and function over different spatial and temporal scales and to integrate this information with biogeochemical and physicochemical measurements to provide critical, baseline information about microbial community dynamics that can inform future restoration efforts.

Along with meeting this goal, the completion of this dissertation has resulted in several broad contributions to the field of microbial ecology. For one, this dissertation helped improve and apply new methods and analyses for studying coral reefs over a variety of scales and environmental gradients. Additionally, this dissertation has integrated genome-based analyses (amplicon sequencing and metagenomics) with biogeochemical and physicochemical measurements, as well as reef composition surveys, for a comprehensive analyses of the environmental gradients and differences in reef conditions that impact these microbial communities. Lastly, three chapters of this dissertation provided the first information about the

315

microbial ecology of the remote, protected, and understudied reef-system of Jardines de le Reina, Cuba , placing it within the context of other reef-systems within the Caribbean.

Chapter two focused on improving the methodology for extracting microbial DNA (specifically bacterial and archaeal 16S rRNA) from coral tissue and investigating if different DNA extraction methods biased microbial community composition. The motivation for this work was to satisfy a need within the scientific community to standardize and address the methodological issues involved with extracting and amplifying DNA from diverse coral species. Overall, this work demonstrated that coral species, not DNA extraction method, significantly influenced the community composition of the coral microbiome. However, one DNA extraction method was able to isolate DNA from more cryptic bacterial members of the coral microbiome and led to higher estimates of microbial community diversity, suggesting that researchers need to consider DNA extraction methods when designing their studies if they intend to survey rare or specific bacteria living within the coral microbiome. This chapter also highlighted questions about the potential mechanisms behind coral species-specific PCR amplification performance and the generation of low quality sequences for specific species, emphasizing that all steps involved in sequence generation have a potential to inhibit downstream microbial community analyses. This chapter was instrumental to my development as a scientist; the design and implementation of this experiment taught me the methodological and analytical skills that have only matured in the development of my other thesis chapters.

Chapter three stepped away from the tissue-associated coral microbiome towards investigating the dynamics of reef microorganisms in the seawater surrounding individual coral colonies. In this field-based study, set within the Jardines de la Reina and Canarreos reef-systems in Cuba, we used amplicon- and shotgun metagenomics-based sequencing to show that the

compositions of microorganisms within coral ecospheres, the seawater environment immediately surrounding coral colonies, were both specific to coral species and influenced by the local reef environment. Metagenomic sequencing of all the DNA and genes within pooled samples by coral species demonstrated that coral ecosphere metagenomes were significantly enriched with genes used by microorganisms to attach to and interact with surfaces. Additionally, the relative abundance of *Endozoicomonas* spp., a common coral tissue and mucus symbiont, was higher in the seawater adjacent to *Porites astreoides* corals, potentially suggesting that coral symbionts and/or pathogens may recruit to the coral surface from the coral ecosphere. This chapter has broader significance within the field of coral reef ecology because it demonstrates evidence of a potentially important link between seawater microorganisms and the coral surface, a benthic – pelagic coupling that has been underexplored. The results from this study have generated many questions about the timescales of, and mechanisms behind symbiont or pathogen recruitment to the coral surface and if and how environmental conditions can impact these microbe - coral interactions. Further, the question of how and why different species of corals select for different ecosphere microbial communities is quite interesting, having possible connections with the tissue-associated coral microbiome as well as the potential evolution of coral – microbe specific symbioses. Future studies should focus on species-specific symbiont recruitment and pathogen colonization of corals because they are critical processes that need to be understood, especially as global climate change and other stressors lead to coral bleaching and the spread of coral diseases.

The results of chapter three served as motivation for the field-based research that was conducted in chapter four. In this chapter, I investigated variation in microbial community composition, microbial cell abundances, and inorganic macronutrients over diel, daily, and

spatial scales in proximity to five different *P. astreoides* colonies that were distributed across

Ram Head reef in St. John, U.S. Virgin Islands. In this study, paired, small-volume samples of

seawater were collected 5 cm (ecosphere) and 2 m away (reef-depth seawater) from 5 *Porites*

astreoides coral colonies over the course of three days during the day and night. Surface

seawater samples were also collected. Surprisingly, there were no significant differences in

composition or diversity between ecosphere and reef-depth seawater microbial communities, but

this result could be due to potential colony- or reef-specific differences between Jardines de la

Reina, Cuba and St. John, United States Virgin Islands. Bacterial and archaeal alpha diversity

was significantly higher in reef (reef-depth and ecosphere) seawater compared to surface

seawater, suggesting that benthic organisms and their exudates may be sources of

microorganisms and/or metabolites that can enhance microbial niche availability at reef-depth.

Prochlorococcus and *Synechococcus* cell abundances exhibited predictable diel fluctuations,

with the number of cells increasing at night (doubled for *Prochlorococcus*) compared to daytime

abundances. The abundances of picocyanobacteria also changed over the course of three days

and correlated with changes in temperature and nitrite and silicate concentrations. While

Prochlorococcus and *Synechoccocus* influenced shifts in bacterial and archaeal community

composition through time, there also were a number of bacterial taxa that exhibited significant

diel differences in relative abundance or that were significantly enriched in reef-depth seawater.

In future investigations, it would be useful to study why these bacterial taxa exhibit diel

fluctuations or favor reef-depth seawater. The results of chapter 4 have extended our knowledge

of spatial and temporal microbial variation in coral reef ecosystems, demonstrating that temporal

changes supersede spatial differences in terms of influencing the reef seawater microbial

community across one reef.

Chapter five expands out to a larger spatial scale and returns to reef-systems in Cuba and the Florida Keys to investigate patterns in microbial community composition, function, and diversity, as well as changes in biogeochemical and physicochemical properties of reef seawater along a gradient of human influence. Microbial community ecology of the protected and understudied reef-system of Jardines de la Reina, Cuba was surveyed and compared with the more impacted reef-systems in Canarreos, Cuba and the Florida Keys. This chapter constitutes the first measurements and characterization of microbial communities using 'omics approaches in the relatively healthy reef-system of Jardines de la Reina and provides necessary baseline information about the microbial community ecology of unimpacted Caribbean reefs (a baseline that is almost impossible to obtain for the Caribbean region). Each reef-system was defined by different microbial and biogeochemical signatures. Offshore reefs within Jardines de la Reina had the highest bacterial and archaeal alpha diversity, highest similarities in community composition, and highest abundances of *Prochlorococcus* out of all the reef-systems studied. Also, organic and inorganic macronutrient concentrations were low or undetectable on the offshore JR reefs, demonstrating their oligotrophic nature. In comparison, reefs within Canarreos and the Florida Keys had lower bacterial and archaeal alpha diversity, fewer picocyanobacteria (in terms of cell abundance), and higher concentrations of macronutrients. Nearshore reefs within the Florida Keys also exhibited a few signs of microbialization, demonstrating evidence of a trophic shift favoring microorganisms on these nearshore reefs. Two of the biggest questions emerging from this work and the previous work detailed in chapter four are: 1) what is the role of picocyanobacteria, specifically *Prochlorococcus* and *Synechococcus*, on reefs and 2) are there any direct connections between picocyanobacterial abundance and coral health or is

319

picocyanobacterial abundance just an indicator of suitable environmental conditions for reef formation and growth?

Chapter six studied the forereefs of the protected Jardines de la Reina archipelago using targeted and untargeted metabolomics techniques. Reef-depth, surface reef, and off-reef (bluewater) seawater samples collected across the archipelago were analyzed to identify and quantify known extracellular metabolites and investigate multivariate trends in metabolite composition across the reef-system. These analyses were paired with biogeochemical and physicochemical measurements, estimates of microbial cell abundances, and compositional reef data to investigate trends between bulk measurements and the high resolution metabolomics datasets. Untargeted metabolite feature composition, macronutrient concentrations, and cell abundances were fairly similar across the archipelago, reflecting the homogenous environmental conditions that are likely influenced by input of offshore seawater to the reefs. A variety of nucleosides, amino acids, vitamins, and important metabolic intermediates were identified and quantified using the targeted method, revealing interesting depth-related and biogeographic patterns in the concentrations of riboflavin and methylthioadenosine (MTA). This work marks an important first step towards profiling reef metabolites and provides the groundwork for future research to investigate sources and sinks of metabolites on reefs, metabolite variation across a single reef, associations between reef metabolite compositions and reef health, and signature metabolites of different marine biomes (e.g. seagrass, mangrove, and coral reefs).

Altogether, this dissertation provides important insights into the composition, diversity, functional capacities, and metabolisms of reef seawater microorganisms along different environmental and anthropogenic spatial gradients, as well as different temporal and spatial scales. The results of these analyses will inform future investigations of reef microbial

community ecology and could serve coral restoration managers in their plans for selecting

potential reef locations as outplanting sites for future restoration efforts. Microorganisms are

essential components of coral reef ecosystems and these ecosystems cannot effectively be

protected and/or managed without understanding the conditions that impact the behaviors,

distributions, and functions of microorganisms.

7.2 References

Barbier, E.B., Hacker, S.D., Kennedy, C., Koch, E.W., Stier, A.C., and Silliman, B.R. (2011) The value of estuarine and coastal ecosystem services. *Ecological Monographs* **81**: 169-193.

Brander, L.M., and Van Beukering, P. (2013) The total economic value of U.S. coral reefs: a review of the literature. In. Silver Spring, MD: NOAA Coral Reef Conservation Program, p. 32 p.

Lavides, M.N., Molina, E.P., de la Rosa, G.E., Jr., Mill, A.C., Rushton, S.P., Stead, S.M., and Polunin, N.V. (2016) Patterns of Coral-Reef Finfish Species Disappearances Inferred from Fishers' Knowledge in Global Epicentre of Marine Shorefish Diversity. *PLoS One* **11**: e0155752.

Odum, H.T., and Odum, E.P. (1955) Trophic structure and productivity of a windward coral reef community on Eniwetok Atoll. *Ecological Monographs* **25**: 291-320.

Storlazzi, C.D., Reguero, B.G., Cole, A.D., Lowe, E., Shope, J.B., Gibbs, A.E. *et al.* (2019) Rigorously valuing the role of U.S. coral reefs in coastal hazard risk reduction. In. Reston, Virginia: U. S. Geological Survey open-file report, p. 42 p.