

5-1-2014

Prokaryotic Diversity of the Wastewater Outfalls, Reefs, and Inlets of Broward County

Alexandra Mandina Campbell

Nova Southeastern University, ac1545@nova.edu

Follow this and additional works at: https://nsuworks.nova.edu/occ_stuetd



Part of the [Marine Biology Commons](#)

Share Feedback About This Item

NSUWorks Citation

Alexandra Mandina Campbell. 2014. *Prokaryotic Diversity of the Wastewater Outfalls, Reefs, and Inlets of Broward County*. Master's thesis. Nova Southeastern University. Retrieved from NSUWorks, Oceanographic Center. (7)
https://nsuworks.nova.edu/occ_stuetd/7.

This Thesis is brought to you by the HCNSO Student Work at NSUWorks. It has been accepted for inclusion in HCNSO Student Theses and Dissertations by an authorized administrator of NSUWorks. For more information, please contact nsuworks@nova.edu.

NOVA SOUTHEASTERN UNIVERSITY OCEANOGRAPHIC CENTER

**Prokaryotic Diversity of the Wastewater Outfalls, Reefs, and Inlets of Broward
County**

By

Alexandra Mandina Campbell

**Submitted to the Faculty of
Nova Southeastern University Oceanographic Center
in partial fulfillment of the requirements for
the degree of Master of Science with a specialty in:**

Marine Biology

Nova Southeastern University

May 2014

**Thesis of
ALEXANDRA CAMPBELL**

Submitted in Partial Fulfillment of the Requirements for the Degree of

Masters of Science:

Marine Biology

Nova Southeastern University
Oceanographic Center

April 2014

Approved:

Thesis Committee

Major Professor : _____

Jose Lopez, Ph.D.

Committee Member : _____

Christopher Sinigalliano, Ph.D.

Committee Member : _____

Donald McCorquodale, Ph.D.

TABLE OF CONTENTS

ACKNOWLEDGMENTS.....	5
ABSTRACT.....	6
List of Figures.....	7
List of Tables.....	8
List of Abbreviations.....	8
INTRODUCTION.....	9
Early Microbiology and Marine Microbiology.....	9
Ribosomal RNA.....	11
Metagenomics and microbiomes.....	13
Seawater Characteristics.....	15
Detection of Pathogens.....	18
Southeast Florida coastal waters.....	18
New River and Port Everglades.....	19
HYPOTHESES AND OBJECTIVES.....	19
METHODS.....	20
Seawater Sample Collection.....	20
DNA Extraction.....	23
Polymerase Chain Reaction.....	23
Sequence Analysis.....	24
Statistical Analysis.....	25
Pathogen Screening.....	26
Source Tracker.....	26
RESULTS.....	26
South Florida coastal hydrology.....	26
16S rRNA Community Analysis.....	27
OTU calling with QIIME.....	28
Rarefaction analysis.....	28
Alpha Diversity.....	32
Most abundant taxa.....	32
Rare Microbiota.....	37
Rank Abundance.....	38
Pathogens.....	41
Beta Diversity.....	41
Site Type.....	41
Season.....	43
Depth.....	45
Source Tracker.....	46
Correlation with metadata/environmental parameters.....	50
DISCUSSION.....	52
Sample Collection and Sequencing.....	52
Sequencing Issues.....	52
Illumina.....	53
Rank Abundance.....	53
Pyrosequencing.....	54

Universal Primers.....	54
Amplicon Sequencing.....	55
Florida Waterways.....	56
Wastewater Treatment.....	57
Florida Area Coastal Environment Project (FACE).....	58
Taxonomic Diversity.....	58
Site Type.....	58
Depth.....	60
Season.....	61
Microbial Composition.....	62
Abundant Taxa.....	62
Rank Abundance.....	68
Pathogenic Taxa.....	69
Notable Rare Microbiota.....	71
SourceTracker.....	72
Implications and Significance of study.....	73
CONCLUSIONS.....	73
REFERENCES.....	76
APPENDICES.....	90

Acknowledgements

I thank my advisor, Jose Lopez, for all of the guidance he has given me throughout my thesis research. From my first class with him to my thesis defense, he has supported me not only as an advisor, but as a trusted friend.

I thank my committee members for their support and encouragement through my thesis writing. Thank you to Dr. Christopher Sinigalliano and Dr. Maribeth Gidley for use of their laboratory for the DNA extractions. Thank you to Dr. McCorquodale for serving on my committee and giving me encouragement throughout the process. Thank you also to James White for his assistance with the CloVR software and additional analyses. Thank you also to Brian Walker for providing information on the Port Everglades inlet.

Thank you to the Batchelor Foundation for providing financial support through my thesis research.

Thank you to my family, Vernon, Jessica, Katherine, and Elizabeth Campbell, for your unwavering support throughout my thesis writing and for encouragement, even when I felt like quitting.

I thank Dr. Anja Schulze and former members of the Marine Invertebrate Zoology Laboratory at Texas A&M at Galveston, for the laboratory experience and camaraderie during my undergraduate years. The time spent volunteering in the invertebrate lab gave me vital skills for my thesis work, and I would probably not be where I am today if not for Anja's patience and guidance.

Abstract

We applied culture-independent, next-generation sequencing (NGS) high throughput pyrosequencing, to characterize the microbial communities associated with near shore seawater in Broward County, FL. These waters flow over coral reef communities, which are part of the Florida reef tract, and are close to shore where bathers frequent. Through a close partnership with the NOAA FACE program, 38 total seawater samples were taken from 6 distinct locales -the Port Everglades and Hillsboro Inlets, Hollywood and Broward wastewater outfalls, and the associated reef waters-over the course of one year. Tagged 16S rRNA amplicons were used to generate longitudinal taxonomic profiles of marine bacteria and archaea for one year. 236,322 rRNA quality checked sequences with an average length of 250 base pairs were generated. Sequences were found to vary significantly due to seasonal effects, but depth showed no significant correlation. The most abundant taxa among these samples included *Synechococcus*, Pelagibacteraceae (SAR11), Bacteroidetes, various Proteobacteria, and Archaea, such as Thermoplasmata. Other taxa found, albeit in low numbers, were the Thiotrichales, and some members of which can indicate pollution, the Alteromonadales, a biofilm forming order. Inlet sequences were found to be significantly different from the outfall and reef communities by various analyses. Unifrac analysis of microbial beta diversity showed a significant clustering pattern for the inlet samples. Precipitation during the three days before and after sampling was low meaning there was little to no high terrestrial runoff during the sampling days. Higher levels of turbidity were seen at the inlet sites and significantly affected the growth of surface colonizing and biofilm forming bacterial families such as the Rhodobacteraceae and Flavobacteriaceae. This study represents one of the first to apply NGS analyses for a deep analysis of microbial community dynamics in these S. Florida waters.

16S; ribosomal RNA; Bacteria; Archaea; pyrosequencing; seawater; outfalls; inlets; reefs

List of Figures

1. Figure 1-The structure of ribosomal RNA and location of its hypervariable regions in a prokaryotic organism (Neefs 1993).
2. Figure 2-Map of sampling sites around Broward County. Sampling sites are circled in red
3. Figure 3-Rarefaction plots of the Broward sampling effort
 - a. Raw sampling effort comparing the number of sequences to the number of OTUs
 - b. Sampling effort cut off at 2000 sequences per sample
4. Figure 4-Rarefaction plots of the Hollywood sampling effort
 - a. Raw sampling effort comparing the number of sequences to the number of OTUs
 - b. Sampling effort cut off at 2000 sequences per sample
5. Figure 5-Skiff plot of the class level taxa sorted by outfalls, reefs, and inlets. Ordered as Hollywood then Broward from April to January
6. Figure 6-Summary of taxa abundant (>1%) across all samples.
7. Figure 7- Taxa uniquely abundant at reefs
8. Figure 8- Taxa uniquely abundant at outfalls
9. Figure 9- Taxa uniquely abundant at inlets
10. Figure 10-Rank abundance curve for Broward
11. Figure 11-Rank abundance curve for Hollywood
12. Figure 12-Unweighted Unifrac Emperor graphic by site type
13. Figure 13-Weighted Unifrac Emperor graphic by site type
14. Figure 14-Bray-Curtis distance box plots of site type
15. Figure 15-Unweighted Unifrac Emperor graphic by season
16. Figure 16-Weighted Unifrac Emperor graphic by season
17. Figure 17-Unweighted Unifrac graphic by depth
18. Figure 18-Weighted Unifrac plot by depth

List of Tables

1. Table 1-Latitude, longitude, and site description of sampling sites
2. Table 2-Port Everglades tidal times and collection times
3. Table 3-North Broward tidal times and collection times
4. Table 4-Abundant taxa across all samples
5. Table 5-Abundant taxa across reef samples
6. Table 6-Abundant taxa across outfall samples
7. Table 7-Abundant taxa across inlet samples
8. Table 8- Summary of the percent influence of the Broward and Hollywood outfalls. The outfalls were used as a source and the reefs and inlets were sinks.
9. Table 9- Summary of the percent influence of the Broward and Hollywood inlets. The inlets were used as a source and the reefs and outfalls were sinks.
10. Table 10-Multiple linear regression results from SAS

List of Abbreviations

DNA-Deoxyribonucleic Acid

rRNA-Ribosomal Ribonucleic Acid

gDNA-Genomic Deoxyribonucleic Acid

NOAA-National Oceanic and Atmospheric Administration

AOML-Atlantic Oceanographic and Meteorological Laboratory

MID-Multiplex Identifier

OTU-Operational Taxonomic Unit

QIIME-Quantitative Insights Into Microbial Ecology

CloVR-Cloud Virtual Resource

NCBI-National Center for Biotechnology Information

BLAST-Basic Local Alignment Search Tool

Rnldb-Ribosomal RNA Operon Database

Introduction

Early Microbiology and Marine Microbiology

Microbiology studies a diverse group of invisible life on Earth, microbes. These organisms are a large part of the biological interactions of the world's ecosystems from mammalian guts to hydrothermal vents. Before the use of ribosomal RNA, or rRNA (Woese and Fox, 1977), the use of microscopy and pure culture techniques dominated the science of microbiology. Molecular technology did not exist, or what little did exist, was daunting and expensive to use. Most modern microbiology texts detail information learned from culture-based techniques (Handelsman, 2004). It only takes into account those microbes that can be easily cultured in a laboratory setting (Staley and Konopka, 1985). Most microbial life occurs on a small scale, meaning that they occur below our usual detection limits. Louis Pasteur coined the phrase "everything is everywhere, but the environment selects." He thought that microbes were distributed globally, but some microbial species would be present at different levels of abundance, due to environmental constraints (de Wit and Bouvier, 2006).

Bacteria and Archaea are responsible for biogeochemical cycling, atmosphere component production, and form a large portion of the genetic diversity in the ocean (Whitman et al., 1998). They are important in oceanic systems for the cycling of nutrients and dissolved organic matter (Azam et al., 1983). The continental shelf habitats host about 5×10^5 cells per milliliter of seawater. Most of these are strains of *Prochlorococcus* and other cyanobacteria. Globally, the number of cells per milliliter of seawater is 3.6×10^{28} , with 2.9×10^{27} of those cells being autotrophs (Whitman, 1998).

Early techniques of enumerating bacteria included the use of epifluorescent microscopy and associated dyes, such as 4',6-diamidino-2-phenylindole (DAPI) (Porter and Feig, 1980) and acridine orange (Austin, 1988; Daley and Hobbie, 1975). Epifluorescent microscopy works by exciting the fluorochromes of a fluorescent dye, such as DAPI or acridine orange. It works by attaching to DNA or RNA within the cell and will *fluoresce* under certain wavelengths. DAPI fluoresces bright blue and detritus shows as a faint yellow (Porter and Feig, 1980). Acridine orange fluoresces green for live

matter, but has strong background fluorescence in the presence of organics and detritus. These techniques showed a way to enumerate bacteria (Daley and Hobbie 1975; Porter and Feig 1980), but could not determine the sequences and strong identification. What it did show, though, was that there was a discrepancy in the numbers of bacteria in the environment compared to those that could be cultured under laboratory conditions (Porter and Feig 1980). At the time, most bacteria known were those that were able to be cultured with very little information being available for non-culturable species.

Early microbiology was dependent on culturing bacteria, but this method is not fool proof. The “great plate count anomaly” coined by microbiologists Staley and Konopka (1985), defines culturable bacteria as making up only a very small percentage of the total bacterial diversity. The species that are culturable reflect those that can grow under laboratory conditions. However, they may not have any importance to the environment (Staley and Konopka 1985).

Many uncultivable organisms remained unidentified or vaguely understood, until Carl Woese pioneered ribosomal RNA (rRNA) studies to determine microbial phylogeny. From this work, he found a new domain of life. Archaea were originally thought to be bacteria, but Woese found them to be as different from bacteria as they were from eukaryotes. The rRNA of this new organism was found to be of similar size to that of bacteria, but base modification patterns were found to be different (Woese and Fox, 1977). This eventually led to the three-domain system used today.

Analyzing rRNA bypasses the need for culturing microorganisms. This cultivation independent analysis has led to the discovery of new species of marine microbes, such as SAR11 or *Pelagibacter*, and several species of Archaea (Delong, 2005; Delong and Karl, 2005). Cultivation independent technology now provides a larger view of microbial assemblages in many environments. Many of the approaches to culture-independent analysis include inserting DNA into bacterial artificial chromosomes (BAC) and whole genome shotgun sequencing in which DNA is broken into numerous fragments and sequenced, then assembled *in silico* (Delong 2005). 16S rRNA analysis has even revealed some of the species that occupy the “rare biosphere” (Sogin, 2006) or, species with low abundance levels (Sogin, 2006; Pedrós-Alió, 2012).

Ribosomal RNA

Ribosomal RNA (rRNA) is a commonly applied molecule for microorganismal taxonomy and (to a lesser extent) phylogenies. There are two major rRNA molecules - a small and large subunit (SSU, LSU). The 16S rRNA is a highly conserved region making it ideal for phylogenetic studies (Woese, 1987). It makes up the 30S SSU, which is a part of the over 70S ribosome in prokaryotes (Pei et al., 2009). Phylogenies inferred from other molecules almost always agree with rRNA data to a first approximation (Olsen and Woese, 1993; Pei et al., 2009). A larger ribosomal RNA, such as 23S, may be expected to contain more phylogenetic information than a smaller rRNA, like 16S. Analysis of 23S rRNA could be just as reliable, but due to a lack of universal primers and sequence length, any early sequencing techniques utilizing 23S were challenging (Pei et al., 2009). Defining characteristics at the genome level surpass the phenotypic level, since many genomic sequences may influence phenotype, and the genotype is always changing (Olsen and Woese, 1993). The use of rRNA sequences has increased the knowledge of uncultured microbes and revealed unexpected phylogenies (Rappé and Giovannoni, 2003; White et al., 2012).

Ribosomal RNA taxonomic studies focus on the hypervariable regions (Figure 1) since they can lead to more accurate classifications. Certain variable regions are desirable over others. The V6 region, one used in earlier metagenomics studies, comprise a short base pair region and is flanked by conserved regions, making it less reliable. V1-2 and V4, are flanked by semi-conserved regions, which is more reliable for taxonomic classification (Wang et al., 2007). The V4 hypervariable region shows a high (<90%) classification accuracy up to the family level and around 80% accuracy at the genus level.

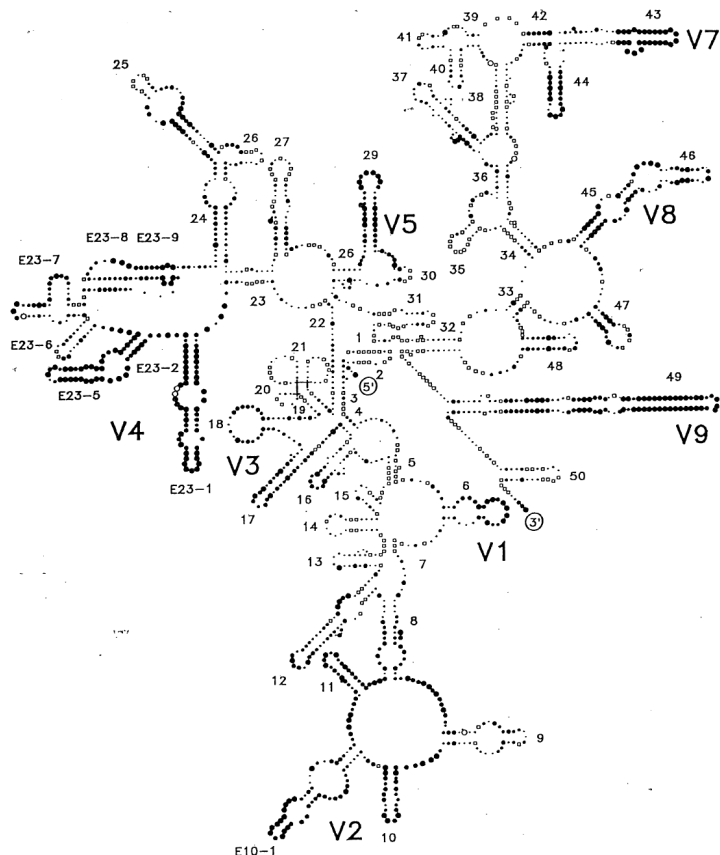


Figure 1 The structure of ribosomal RNA and location of its hypervariable regions in a prokaryotic organism (Neefs et al., 1993).

Currently, quantitative PCR (qPCR) and culture-based data collection make up the majority of microbiological studies, mainly due to the speed of results obtained qPCR and make targeted probes of target organisms for public and environmental health data. The hinderance of this technology, however, is that the majority of microbes are not culturable (Cassler et al., 2008), and qPCR work is dependent on known targets to generate primers and probes. Culture-based studies take at least a day or more and are not a strong indicator of microbial abundance, since most species are uncultivable, and some enter a viable but not culturable state (Oliver, 2005). Deep sequencing data, collected from 454 pyrosequencing and Illumina (formerly Solexa), both of which are next-generation sequencing technologies, can show the rare and uncultivable bacteria and archaea present in environmental samples, which may serve as environmental indicators (Kirchman 2008).

“Next generation DNA sequencing” (NGS) technology is an important advancement in molecular biology since it bypasses the use of clone libraries (Kirchman et al., 2010; Pedrós-Alio, 2012) and can create large datasets from relatively small samples. Pyrosequencing is one type of next generation sequencing platform advanced by the company 454, which utilizes an integration of an unlabelled nucleotide using emulsion PCR on a bead surface. Barcoded DNA samples are placed in a well containing a bead with a fusion primer. The barcoded DNA has a complementary fusion primer and this binds to the bead. Detection of the sequence is determined by the release of inorganic pyrophosphate (Mardis, 2011). The fragment is incubated with DNA polymerase, ATP sulfurylase, firefly luciferase, and a nucleotide degrading enzyme (Ronaghi et al., 2008). The resulting chemical reaction generates light. Typical read length for this technique is about 400 base pairs (Mardis et al., 2011). Pyrosequencing became more important to genomic studies as clone libraries are no longer needed and can generate large datasets from small amounts of DNA (DeLong, 2009; Kirchman, 2010; Pedrós-Alio, 2012). Another method of next generation sequencing technology is Illumina, which utilizes enzymatic amplification on a glass surface, but generates short, about 150bp, sequences (Mardis, 2011). Through the use of next-generation sequencing, it is now possible to view microbes from the “rare biosphere” (Sogin et al., 2006), a group of unculturable, rarely occurring microbes. This new appreciation of unknown microbes may yield clues into the effects of environmental change (Jeraldo et al., 2011). One of the inherent issues with pyrosequencing is the creation of “noise.” The source of pyrosequencing noise is the light generated. It does not always reflect the length of a homopolymer, a sequence of identical bases. Certain softwares, such as PyroNoise, are utilized to remove pyrosequencing noise (Quince et al., 2009).

Metagenomics and the microbiome

Much like the pioneering work of Carl Woese and George Fox revealed unexpected bacterial phylogenies and taxonomies, modern molecular microbiology is beginning to utilize a technique known as metagenomics. Metagenomics is defined as the cultivation-independent analysis of genes taken directly from the environment. It is heavily reliant on molecular biology (e.g. cloning, PCR, DNA sequencing, etc.)

(Handelsman, 2004; Delong, 2009; Gilbert et al., 2010; 2011). The first metagenomics projects utilized clone libraries to separate microbial species out and were sequenced on the Sanger sequencer. In 2000, Béjà et al., cloned the first environmental DNA from seawater into bacterial artificial chromosome (BAC) libraries and found bacteriorhodopsin, which is used by cells to generate ATP from light. This is among one the greatest accomplishments of metagenomics as this phenomenon was thought to only occur in hypersaline pools (Temperton and Giovannoni, 2012). Then, in 2004, J. Craig Venter and colleagues sequenced seawater from the Sargasso Sea (Venter et al., 2004). This study showed the magnitude of metagenomic studies, with 148 novel phylotypes found and 1.2 million novel genes discovered (Temperton and Giovannoni, 2012). This was expanded into the Global Ocean Sampling (GOS) survey in 2007 (Rusch et al., 2007), which aimed at identifying the microbial constituents of the world's oceans. Most of these early studies relied on clone libraries and Sanger sequencing.

DNA sequencing technology has since progressed and now metagenomic and targeted amplicon studies can be done using “massively, parallel next-generation” sequencing on technologies such as Roche 454 pyrosequencing or Illumina (formerly Solexa). Metagenomics, the study of all of the genes in an environment, and 16S rRNA amplicon library analyses, a subset of metagenomic analysis that focuses on targeted gene sequencing to determine the microbial make-up of an environment, is still an emerging field, but recent advances have revealed distributions of phylogenies, genes and metabolisms in an environmental sample (Gilbert, 2010). One of the most recent studies was the Human Microbiome Project (The HMP Consortium, 2012a;2012b), whose aim was to determine the distribution of microorganisms, their functions, and metabolisms within and on the human body. The results showed that microbial composition differs with respect to bacterial quantity, speciation, and location (Anderson et al., 2008; Lazarevic et al., 2009). This suggests that microbes may not show random distribution due to space or time (Petrosino et al., 2009). Newer studies include the Earth Microbiome Project (earthmicrobiome.org), aimed at studying microbial phylogenies from the diverse environments of Earth. A further offshoot of the Human Microbiome Project includes the Hospital Microbiome Project, headed by Jack Gilbert, whose aims are to study the

microbial composition of a hospital from when it is built to when it is finished and how the influx of people can change its microbial makeup.

Seawater characteristics

Seawater covers most of the planet, covering an area of $3.61 \times 10^8 \text{ km}^2$ and containing a total volume of $1.4 \times 10^{21} \text{ L}$ of water. Seawater is chemically characterized by high levels of sodium and chloride ions, but nitrogen, phosphorus, and trace elements like iron are also present. The salinity of open ocean seawater is approximately 32-38 ppt, but in river basins or areas with high terrestrial runoff, salinity may dip as low as 10 (Austin, 1988). River drainage and terrestrial runoff are also tend to be high in limiting nutrients. Freshwater and nutrient input can affect the composition of both microbes and eukaryotic organisms (Austin 1988; Shibata et al., 2004; Brownwell et al., 2007; Futch et al., 2011) by increasing nutrient levels and reducing the salinity. This can allow opportunistic or pathogen species to grow (Brownwell et al., 2007).

Seasonal and spatial differences, often over short distances, have an effect on the community structure of organisms. Alternations between wet and dry weather and temperature are a factor in the community ecology of an area. The addition of nutrients and freshwater from terrestrial environments brings changes in phytoplankton abundance which may also indirectly change the microbial structure of that area. Spatially, the availability of nutrients and organic resources can affect the overall community structure (McArthur, 2006).

In a 2013 study of Kaneohe Bay, HI, by Yeo et al., it was found that during non-storm events, temperature and salinity fluctuated very little and chlorophyll *a* levels remained consistent. Through the use of terminal restriction fragment length polymorphism (T-RFLP), it was found that fluctuations in *Synechococcus*, other photosynthetic organisms, and non-photosynthetic organisms were dependent on environmental changes to change diversity. During storm events, runoff increased, which brought an influx of freshwater and also an increase in inorganic nutrients. This, in turn, increased the chlorophyll *a* content. From this study, it was speculated that storm events are a driver in microbial diversity. The patterns in environmental conditions, along with

the freshwater runoff during storms are important to the dynamics of bacterial communities in response to a disturbance.

A study by Gilbert et al. (2010) in the Western English Channel demonstrated that microbial metagenomes, using 16S rRNA, vary with seasons, with the highest bacterial diversity occurring in the winter months. Temperatures between the winter (Jan-April) and summer (August) months varied by about 5-6°C. Chlorophyll *a* concentrations varied greatly from winter to summer, the highest being in the summer, indicating a possible bloom of photosynthetic organisms. These community levels change in 16S rRNA diversity, meaning, metagenomes can be better explained by seasonal patterns (Gilbert, 2010). Previous work by Murray et al., (1998) at Anvers Island in Antarctica found no significant changes in the distribution of bacteria and archaea with depth, but some seasonal distribution of archaea. Bacterial distribution tended to be stable year round in the Antarctic and a similar distribution was found in the Arctic Ocean (Kirchman et al., 2010). Archaeal diversity tended to be higher in the winter and early spring months, but declined as summer progressed. It is suggested that this may be due to a competition for resources in the water column, supported by a decrease in archaeal rRNA and an increase in chlorophyll *a* concentrations, likely due to phytoplankton abundance (Murray et al., 1998). Increased amounts of phytoplankton create particulate matter to which microbes can attach (Austin, 1988). This correlation between chlorophyll *a* and archaea was seen by Gilbert et al., (2010) in the Western English Channel Study.

Environmental genome shotgun sequencing in the Sargasso Sea has led to insight on the community structure of a subtropical marine environment. The Sargasso Sea is an open ocean environment located on the Gulf Stream. An example of findings from this study includes a large amount of gammaproteobacteria present and SAR11, a bacterial clade distributed throughout the world's oceans (Venter et al., 2004). Work in the Sargasso Sea eventually led to the Global Ocean Sampling (GOS) Expedition which set out to characterize the microbiota of the world's oceans using whole genome shotgun sequencing. Using GenBank and 98% identity cutoff, large assemblages of the SAR11 clade were found throughout the oceans. Other species include *Burkholderia*,

Prochlorococcus, *Synechococcus*, *Shewallena* and alpha- and gammaproteobacteria (Rusch et al., 2007).

Seawater is a diverse environment characterized by mixing from currents, variable nutrient input, and the distribution of microbes (Austin, 1988; Rappe and Giovannani, 2003; Zinger et al., 2011). These advances establish that microbial communities are important to the functions of macroorganisms and to the ecosystem health (Furhman, 2009; Gilbert, 2011). The integration of metagenomics with cultivation studies and environmental surveys has led to insights about genomics, population genetics, and community ecology. This type of study is more likely to be seen in the future and a strong precursor has been the study of human microbiomes (DeLong, 2009). The science of metagenomics is complex and challenged by a vast amount of data.

In order to understand the full potential of metagenomics studies, it is important to first estimate the microbial community structure. Most of this is done through amplicon studies of the 16S rRNA gene and on a designated hypervariable region such as the V4 (Caporaso et al., 2011), or V6 (Gilbert et al., 2010; Sogin et al., 2006). In a study by Huber et al., (2006), they collected water from two hydrothermal vents, Bag City and Marker 52. Each vent had a unique chemical composition and each were only 3 kilometers apart. Marker 52 is noted as being bare rock with low pH, elevated alkalinity, high hydrogen sulfide and a lower temperature than Bag City. There were stark differences between the sites, namely in the epsilonproteobacteria. *Arcobacter* spp. dominated at Bag City, while *Sulfurovum* spp. were dominant at Marker 52.

Previous work on indicator organisms in south Florida waters and beaches have shown seasonal spikes due to non-point source contamination. Non-point sources of contamination include sewage (Shibata et al., 2004), storm water (Shibata et al., 2004; Brownwell et al., 2007), sand resuspension (Hartz et al., 2008), dog feces (Wright et al., 2009) and human shedding (Abdelzaher et al., 2010). Bacteria present due to human shedding include many commensal organisms such as *Enterococci* and *Staphylococcus aureus* (Elmir et al., 2007). While this study was performed without next-generation sequencing technology, it is important since it showed how bacterial communities fluctuate in seawater and sands seasonally due to seasonal rainfall, sewage, dog feces,

and human shedding. It serves as an important baseline for further microbiology studies in south Florida.

Detection of Pathogenic Strains

The detection of pathogenic strains of bacteria is vital to water quality studies. Many of the targeted pathogenic bacteria are enteric bacteria, that is, those bacteria that live in the guts of birds, dogs, humans, etc. The addition of freshwater and nutrients can affect the growth rate of these enteric bacteria in seawater since it affects the amount of osmotic stress on the organisms. Most enteric organisms enter a viable but not culturable state when placed in seawater, and shrink. This makes traditional culturing techniques difficult when assessing water quality in marine ecosystems. In an example from Munro et al., (1989), *Escherichia coli* declines in situ and in laboratory settings when exposed to seawater. There is, however, a chance for bacteria to grow in seawater, if osmotic stress can be overcome. *E. coli* for example, can accumulate potassium ions and activate systems of transport of osmolytes, which prevent the cells from dehydrating in solution (Munro et al., 1989).

Southeast Florida Waters

The coastline of Miami-Dade, Broward, and Palm Beach counties extends 142 kilometers. This expanse of coastline is increasingly developed with contiguous urbanization. The drainage of wetlands has allowed for western urban expansion, however, it created a burden on the drainage system. To deal with the increased demand, the Army Corps of Engineers built drainage canals, causing a change in surface and ground water flows (Sklar et al., 2002), and remains a major carrier of agricultural and urban runoff (SFWMD, 2010).

The climate of southeast Florida, based on the Köppen climate scale (Peel et al., 2007), is a humid tropical savannah climate, which is distinguished by a distinct wet and dry season (Banks et al., 2008). The rainy season occurs during the months of May until the end of September (Bank et al., 2008). Increases in the amount of rainfall were shown to increase the amounts of indicator organisms during the rainy season as a result of more rainfall and warmer temperatures (Brownwell et al., 2007).

The New River and Port Everglades

The Intracoastal Waterway spans from Fernandina Harbor to Miami Harbor and requires periodic dredging. Due to heavy boat traffic, it is likely a source of non-point pollution. The Intracoastal Waterway connects to the Atlantic Ocean via a series of inlets, which are noted as major pollution sources (Lapointe and Bedford, 2010). Port Everglades used to be known as Lake Mabel. In 1913, a cut was made from Lake Mabel to the Atlantic Ocean to allow small boats to enter. In 1924, Joseph Young, mayor of Hollywood, bought land near the lake and created Hollywood Harbor Development Company. In 1928, the Broward County Port Authority was dedicated, and a barrier separating the harbor from the ocean was created. In the 1940s, Port Everglades served as a military base. During the 1950s, it became a major cruise line waterway (<http://www.porteverglades.net/about-us/history/>). The average depth of Port Everglades is between 42-48 feet (personal communication with Brian Walker).

Diverse collections of microbes encode for many biochemical functions that may be beneficial to their environment and neighboring microbes. Bacterial communities are focused on because they are the predominant group in a microbiome and have comprehensively documented phylogenetic datasets and classification systems (Petrosino et al., 2009).

Southeast Florida is characterized by a tropical savannah climate and is the convergence zone for tropical and temperate waters (Banks et al., 2008). The area, which encompasses Martin, Palm Beach, Broward, and Miami-Dade counties, is home to over 5,710,953 (2010 Census). Over 57% of these people rely on centralized sewers, while another 40% rely on inground wastewater disposal (Futch et al., 2011). Wastewater is removed by a series of outfalls, which drain directly into the marine ecosystem (Banks et al., 2008; Futch et al., 2011).

Hypotheses and Objectives

The purpose of this study was to determine the microbial makeup of the seawater of southeast Florida based on season, depth, and site type, using next generation sequencing in lieu of traditional culturing techniques and quantitative PCR. The

overreaching goal of this project was to set a baseline of the microbial life of the outfalls, reefs, and inlets of Broward County, FL, and provide data for further projects in this area.

The following hypotheses about microbial assemblages around Hollywood and Broward County will be tested. First, there will be differences in microbial communities between sites (beta diversity). Second, there will be a difference in microbial distribution based on depth. Third, microbial assemblages will differ with respect to season within one year.

Methods

Seawater Sample Collection

Seawater samples were collected bi-monthly at the Broward and Hollywood outfalls, reef tract, and coastal inlets (Table 1; Figure 2) from varying depths and seasons.

Sample ID	Latitude	Longitude	Habitat Type
BR7	26.2039	-80.0683	Reef track
BR10	26.254	-80.0621	Outfall
BR14	26.2618	-80.0855	Coastal area, inlet
HW4	26.0163	-80.087	Outfall
HW9	26.0673	-80.0851	Reef track
HW14	26.0944	-80.1163	Coastal area, inlet

Table 1-Latitude, longitude, and habitat types for each sample.

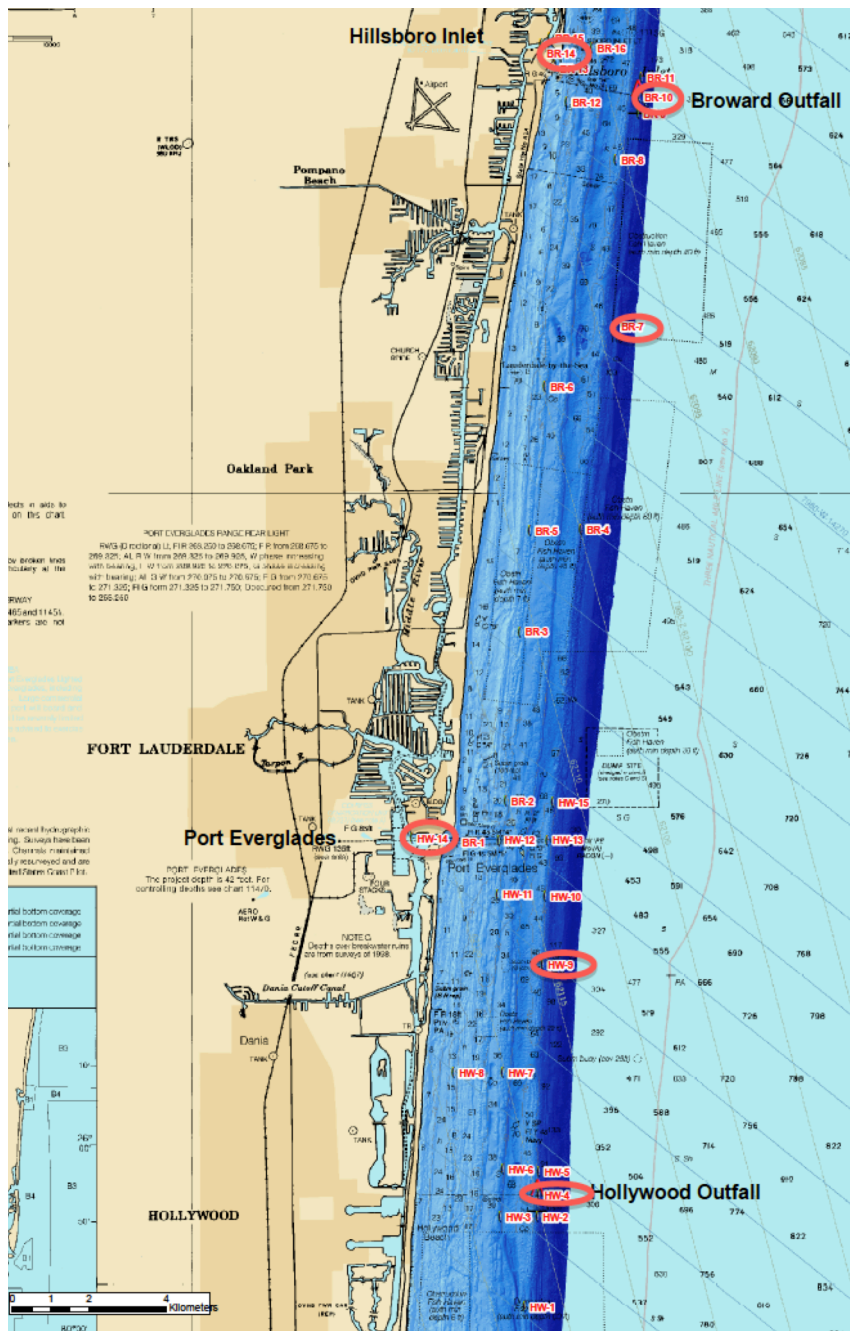


Figure 2-Map of sampling sites. Collection sites are circled in red.

Samples were collected between high and low tides around the Port Everglades (Table 2) and North Broward Outfalls (Table 3).

Date	Tidal Time	Feet	Centimeters	Tides (H/L)	Sample ID	Collection Time	Collection Date	Site Type	Tides
04/19/2011	3:34 AM	-0.3	-9	L	HW14	01:55 PM	04/19/2011	Inlet	Between
04/19/2011	9:42 AM	3.1	94	H	HW9	11:43 AM	04/19/2011	Reef	Between
04/19/2011	3:53 PM	-0.8	-24	L	HW4	10:12 AM	04/19/2011	Outfall	Between
04/19/2011	10:22 PM	3.2	98	H	HW14	03:35 PM	07/06/2011	Inlet	Between
07/06/2011	12:22 AM	2.8	85	H	HW9	01:49 PM	07/06/2011	Reef	Between
07/06/2011	6:32 AM	-0.2	-6	L	HW4	12:17 PM	07/06/2011	Outfall	Between
07/06/2011	12:54 PM	2.7	82	H	HW14	01:49 PM	11/30/2011	Inlet	Between
07/06/2011	6:56 PM	0	0	L	HW9	11:19 AM	11/30/2011	Reef	Between
11/30/2011	5:38 AM	0.3	9	L	HW4	9:56 AM	11/30/2011	Outfall	Between
11/30/2011	11:55 AM	2.9	88	H	HW14	12:30 PM	01/12/2012	Inlet	Between
11/30/2011	6:14 PM	0.5	15	L	HW9	10:34 AM	01/12/2012	Reef	After H
01/12/2012	3:42 AM	-0.4	-12	L	HW4	9:08 AM	01/12/2012	Outfall	Between
01/12/2012	10:08 AM	2.7	82	H					
01/12/2012	4:10 PM	-0.3	-9	L					
01/12/2012	10:30 PM	2.6	79	H					
http://tidesandcurrents.noaa.gov/									

Table 2-Tides and collection times of the Port Everglades area outfalls, reefs, and inlets.

Collection Date	Tidal Time	Feet	Centimeters	Tides (H/L)	Sample ID	Collection Time	Collection Date	Site Type	Tides
04/13/11	4:47 AM	2.4	73	H	BR14	09:37 PM	4/13/11	Inlet	Between
04/13/11	11:09 AM	0.2	6	L	BR7	01:02 PM	4/13/11	Reef	Between
04/13/11	5:16 PM	2.4	73	H	BR10	11:35 AM	4/13/11	Outfall	Low
04/13/11	11:39 PM	0	0	L	BR14	10:21 AM	7/13/11	Inlet	Between
07/13/11	1:33 AM	0.2	6	L	BR7	12:58 PM	7/13/11	Reef	Between
07/13/11	7:31 AM	2.5	76	H	BR10	11:58 AM	7/13/11	Outfall	Between
07/13/11	1:54 PM	-0.4	-12	L	BR14	03:15 PM	11/17/11	Inlet	Between
07/13/11	8:17 PM	2.7	82	H	BR7	12:09 PM	11/17/11	Reef	Between
11/17/11	12:16 AM	2.6	79	H	BR10	01:34 PM	11/17/11	Outfall	Between
11/17/11	6:30 AM	0.6	18	L	BR14	09:52 AM	01/05/12	Inlet	Between
11/17/11	12:52 PM	2.8	85	H	BR7	01:25 PM	01/05/12	Reef	Between
11/17/11	7:09 PM	0.8	24	L	BR10	12:28 PM	01/05/12	Outfall	Between
01/05/12	5:30 AM	2.2	67	H					
01/05/12	11:48 AM	0.5	15	L					
01/05/12	5:30 PM	2.1	64	H					
http://tidesandcurrents.noaa.gov/									

Table 3-Tides and collection times of the Broward area outfalls, reefs, and inlets.

These samples were collected by our collaborator Dr. Christopher Sinigalliano, at the NOAA AOML lab on Key Biscayne, Miami, FL using a Conductivity-Temperature-Depth (CTD) and sterile plastic 2L Niskin bottles, which close at a computer specified depth. The sensor collects conductivity, temperature, depth, pH, dissolved oxygen, and chlorophyll a data. The Broward reef sites were collected close to the bottom and the Hollywood reef sites were collected near the surface and at the bottom. The Broward outfall sites were collected at the bottom, the middle depth, and the surface of the boil, the area where the wastewater reaches the surface. The Hollywood outfall samples were collected at the bottom of the boil and at the surface of the boil. Finally, the Hollywood Inlet samples were collected right in the Port Everglades Inlet. The Broward Inlet samples were collected just north of the Hillsboro Inlet. These were the samples available at the time, closest to the inlet. For the purposes of this study, it is considered part of the Hillsboro Inlet given its close proximity to the inlet site. One liter of seawater was filtered using 0.45 micron filters (Whatman) via vacuum filtration. The filters are then folded with sterile forceps and placed into bead tubes and stored at -80°C until extraction.

DNA Extraction

Genomic DNA was extracted using the FastDNA Soil Kit (MPBiomedicals, Inc). DNA yield and purity was obtained using the NanoDrop 1000 at the Nova Southeastern University Oceanographic Center, and gel electrophoresis was used to cross reference DNA yields and assess purity.

Polymerase Chain Reaction and Sequencing

Following genomic DNA extractions, variable 16S rRNA gene regions were amplified using universal bacterial primers Eco9 and Loop27rc to detect the presence of amplifiable DNA. Samples that amplified with Eco9/Loop27rc were then assigned a multiplex identifier (MID) before proceeding to the final PCR step. Barcoded universal primers, MIDf-515F and 806rc (Caporaso et al., 2011) were utilized for sequencing the V4 hypervariable region of the 16S rRNA. A BioRad MJ Mini thermal cycler (Bio-Rad) was used to amplify the target region. The initial denaturation step is 94°C for 2 minutes. Then the denaturing occurred at 94°C followed by an annealing step at 50°C, and finally

extension at 72°C. This was repeated for 30 cycles, followed by a final extension step at 72°C for ten minutes (Caporaso et al., 2011). To prevent contamination, the lab benches and pipettes were wiped down with 10% bleach followed by 75% ethanol. Filter tips were used to prevent aerosol contamination of the pipette. Fresh tubes and tips were also used to minimize contamination. In the PCR reaction, each MID had a positive and negative control. Amplicons were sent to the University of Kentucky for 454 GS FLX Titanium sequencing, which generated SFF, TEXT, FASTA and QUAL files. These files were then incorporated into Quantitative Insights into Microbial Ecology 1.8.0 (Caporaso et al., 2010).

Sequence Analysis

The sequences were analyzed using Quantitative Insights into Microbial Ecology (QIIME v.1.8.0) (Caporaso et al., 2010). Raw sequences were processed in QIIME, removing the reverse primers and barcodes. The processed sequences denoised using Denoiser (Reeder and Knight, 2010). Operational taxonomic units (OTUs) were assigned using uclust (Edgar, 2010) with 97% similarity, using open reference OTU picking. Taxonomic assignments were made using the uclust taxonomic assigner (Bokulich and Rideout, manuscript in preparation) with an 90% confidence cutoff and the greengenes 13_8 reference database (DeSantis et al., 2006; McDonald et al., 2012). The sequences were aligned using the greengenes reference alignment (DeSantis et al., 2006). Chimeric OTUs were detected and removed using the ChimeraSlayer (Haas et al., 2011) algorithm in QIIME. The resulting alignment was used to create a phylogenetic tree with FastTree (Price, Dehal, and Arkin, 2010) and a resulting OTU table. The sequences were further analyzed using Cloud Virtual Resource (CloVR)(Angiouli et al.,2010).

The rarefaction curves were generated using a MOTHUR (Schloss et al., 2009) rarefaction analysis that is integrated into CloVR (White et al., 2011). Rarefaction analysis is a statistical test, which assesses sequencing effort through a graphical representation of alpha diversity, the diversity within a sample. It plots the number of sequences obtained (x-axis) relative to the number of operational taxonomic units (OTUs) on the y-axis. Rarefaction curves are used to represent the amount of species and

richness of the community sampled. The heatmaps, or skiffs, in CloVR were generated using a custom R script (White et al., 2011).

Beta diversity analysis was conducted using unweighted and weighted Unifrac. The difference between unweighted and weighted Unifrac is that unweighted Unifrac is qualitative and measures how microbial communities differ by their environment (seasonality, site type, etc.). Weighted Unifrac is quantitative and measures changes in abundance (Lozupone et al., 2007). What the clusters show for weighted Unifrac is how abundant each sample is and how it compares to the others. The unweighted Unifrac measure is not a viable method for abundance, but rather on what can live in an environment based on chemistry, depth, salinity, et cetera. Unifrac principal coordinate of analysis (PCoA) plots were generated using Emperor (Vazquez-Baeza, et al. 2013).

Statistical Analysis

In QIIME 1.8.0, the `compare_categories.py` script was used to determine the ANOSIM p-value. A number less than 0.05 would indicate a significant change. The parameters tested were depth, season, and location. The test was nonparametric and was performed in QIIME. Further statistical analysis was also done in CloVR using the program MetaStats (White et al., 2009) to compare seasonal changes in the taxa.

SAS was used to perform a multiple regression analysis. Multiple regression analysis works by looking at two or more independent variables have a relationship to a dependent variable. The independent variables must be determined before starting the model. For this analysis (support.sas.com), the independent variables are the environmental metadata-temperature, salinity, depth, time, oxygen reduction potential (ORP), turbidity, total suspended solids (TSS), chlorophyll a, phaeopigments, nitrates, nitrites, total nitrogen, phosphorus, and silica. The R^2 value shows the variation of the y-value explained by the above independent variables (support.sas.com). The model was run on the four most abundant taxa in the dataset-the families Flavobacteriaceae and Rhodobacteraceae, the genus *Synechococcus*, and the candidate genus, *Candidatus Portiera*. The minimum p-value was set to 0.1 for the SAS analysis. The R^2 value shows how much the Y-variation is explained by the regression equation and how well the

model fits the data. A higher R^2 value means that the model fits the data well (personal communication with Jay Fleischer).

Pathogen Screening

Pathogenic strains of bacteria were detected with the `filter_taxa_from_otu_table.py` script to generate an OTU table of a certain strain, and then converting it into a FASTA file using the `filter_fasta.py` script with the BIOM filter option in QIIME. The resulting FASTA file was then useable for the NCBI BLAST (Atschul et al., 1990) database.

SourceTracker

SourceTracker (Knights et al., 2011) predicts the source of microbial communities in samples using Bayesian analysis. A set of samples, such as a body of water are set as a sink, where the microbes settle, and the source, the original microbial habitat. This script is written in **R language** and is an installed component in QIIME. Samples with multiple depths, the Broward and Hollywood outfalls and the Hollywood reefs, were averaged to get the percent influence of the source. The reefs and inlets were set as the sink environment. Unifrac analysis (Lozupone and Knight, 2005) and relative taxon abundance provide qualitative measures of similarities between samples, but do not show proportions. SourceTracker shows the proportions of a source to a sink sample and provides a standard deviation. Previous iterations of SourceTracker were limited to predetermined indicator species, but as sequencing has grown, it has become a vital tool for comparing similarities between communities. Assessment of community similarity can be performed with Unifrac or abundance of taxa analysis, but neither of these methods gives the proportion of similarity between samples.

Results

Southeast Florida coastlines and hydrology

Thirty-eight seawater samples were collected over one year during each season starting with April 2011 until January 2012. Data collected for this project was directly derived from the Florida Area Coastal Environment (FACE) Program at the National

Ocean and Atmospheric Association in Miami, FL. The samples were collected from the Broward and Hollywood outfalls, inlets, and associated reef tracts (Figure 2; Table 1). FACE has collected and analyzed seawater samples around southeast Florida waters since mid-2004 until January 2012 (FACE website). The goals of the FACE program were to determine the effects of biological, chemical, and physical parameters on water quality and provide a baseline for future observations and to determine the sources of nutrient and microbial contamination into near shore waters using traditional culturing techniques and qPCR probes. More information can be found at <http://www.aoml.noaa.gov/themes/CoastalRegional/projects/FACE/faceweb.htm>.

One of the early FACE projects included a yearlong monitoring project of the Boynton Beach and Delray Beach outfalls, which included preliminary results on levels of indicator microbes using culture techniques, IDEXX, and qPCR. The results show an increase in total enterococci at the outfall site in August 2007 and IDEXX Enterolert showed another spike in viable enterococci in May 2008, though these results do not match with the qPCR results (Carsey et al., 2010).

16S rRNA Community Analysis

The V4 region of the 16S rRNA molecule was sequenced in all the water samples as it displays a high confidence level for identification to genus (Wang et al., 2008) and the primers used also detects Archaea. The total number of raw 16S rRNA amplicon 454 sequences we obtained was 393,545. After splitting the libraries, sequences with an indiscernible reverse primer and multiple primer mismatches were removed, leaving a total of 236,322 sequences. The cutoff quality score for all samples was Q25, the default set in QIIME. The average length of the samples was about 250 base pairs. The number of sequences per sample ranged from 327 to 20,020 pre-chimera checking. Multiplexed samples, tagged samples grouped together, were denoised using Denoiser (Reeder and Knight, 2010) implemented in QIIME. Denoising removes sequencing errors and homopolymers, which can artificially inflate OTU diversity (Quince et al., 2011). The resulting denoised files were run through the OTU picking steps in using the default parameters in QIIME for open reference OTU picking. The resulting files were aligned using PyNASTv1.2.

Further, the program ChimeraSlayer in QIIME further corrected, identified and removed a total of 10,442 chimeric sequences, further reducing the number of sequences to 227,499. Chimera filtered samples were clustered into OTUs using 97% pairwise identity. A total of 4447 OTUs were generated from 227,499 sequences using the Greengenes database incorporated into QIIME (OTU table from Greengenes generated March 2014). Other studies utilizing 454 pyrosequencing have found similar or slightly higher sequence and OTU counts, though these occurred in the Arctic and the temperate North Atlantic (Gilbert et al., 2009). A recent study by Crespo et al. (2013) found 3588 OTUs in their pyrosequencing dataset. The number of OTUs per sample ranged from 321 to 19988. In CloVR, the program UCHIME was used to remove chimeric sequences. The OTU Table Summary is available in the Appendix.

OTU Calling with QIIME The OTU table was generated using the Greengenes 13_8 core set (DeSantis et al., 2006), the default used in QIIME 1.8.0 (Caporaso et al., 2010). Sequences matching known pathogenic bacteria and questionable bacteria/archaea were filtered from the OTU table and made into a FASTA file using the `filter_taxa_from_otu_table.py` and `filter_fasta.py` scripts, respectively. The resulting FASTA file was run through a BLASTN search to further resolve any questionable sequences. The OTU table and taxa assignments were made in March 2014.

Denoised 16S rRNA data was analyzed in CloVR - Cloud Virtual Resource (Angiuoli et al., 2011). After splitting the libraries in QIIME, the script `split_fasta_on_sample_ids.py` was used to isolate each sample site into an individual FASTA file. The FASTA files were then run through CloVR, which produces rarefaction curves (Figures 3 and 4) and a skiff plot, a type of heatmap (Figure 5). For example, the rarefaction analysis compares the diversity of the sample sites by plotting the total sequences per sample by the number of OTUs that were created per sample.

Rarefaction Analysis The rarefaction plots were assembled using CloVR. It shows the number of sequences compared to the number of OTUs. This shows the sampling depth of each sample and the complexity in each sample. AmpliconNoise also has a de novo chimera checker, Perseus. Chimeras are sequences that are the result of incomplete PCR extension. The resulting fragment then acts as a primer for a different sequence, generating a chimera (Quince et al., 2011).

A steep line shows that sampling effort has probably not been saturated and probably inadequate to represent the complete community structure of a sample. Re-sampling or re-sequencing of that sample would be advised. When the rarefaction curve plateaus on the graph, it is an indication that more of the diversity of that sample may have been better determined, and likely. The rarefaction curves generated for Broward and Hollywood reveal startling contrasts in the sequencing depths of each sample and even between sites with Hollywood showing more sequences per sample (Figures 3 and 4).

For the Broward sequencing effort, 15 samples had many sequences (>2500) and high OTU counts. The sequence count for Broward stayed around 16000 and OTU count peaked between 1000-1200. The sample that plateaued was BR10BApr, the Broward outfall sample from April 2011, with over 16000 sequences. It yielded about 800 OTUs. The steepest rarefaction curve was for BR7CApr. It had less than 1000 sequences and under 300 OTUs. BR7CJan and BR7CNov also had low sequence to OTU counts. In Figure 3, the samples were reduced to an even sampling depth of 2000 sequences per sample and an OTU count of 800. Most samples fall within these criteria, however, there are exceptions such as BR10AJuly, BR10BJuly, and BR7CApr. These samples are still included in the data analysis.

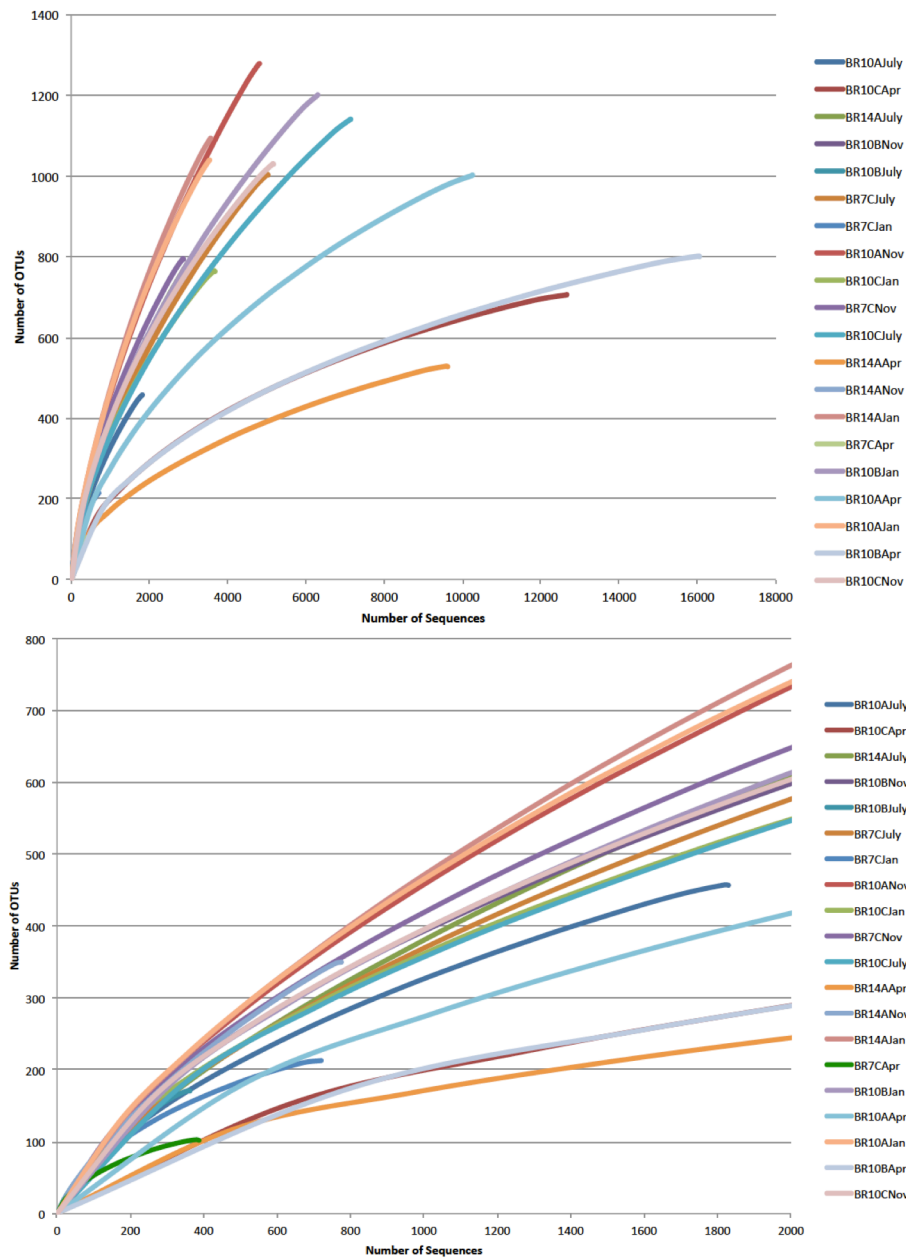


Figure 3-Rarefaction plot of sample IDs for the Broward samples. The plateau of some samples indicates that the sequencing depth is saturated and is a strong measure of the species diversity in a given sample. The top image is the raw data, and the bottom image is the data after evening the sampling depth.

For the Hollywood sequencing effort (Figure 4), 15 samples displayed more than 2500 OTUs. The samples that plateaued were the April samples. The steepest curves occur in the earlier sequencing runs of the November and July samples. The sequence count reached its maximum near 20000 and the maximum OTU count as around 1500. In Figure 4, the even sampling depth was set to 2000 sequences per sample and the OTU cap at 1000. All of these samples exceed 2000 sequences per sample.

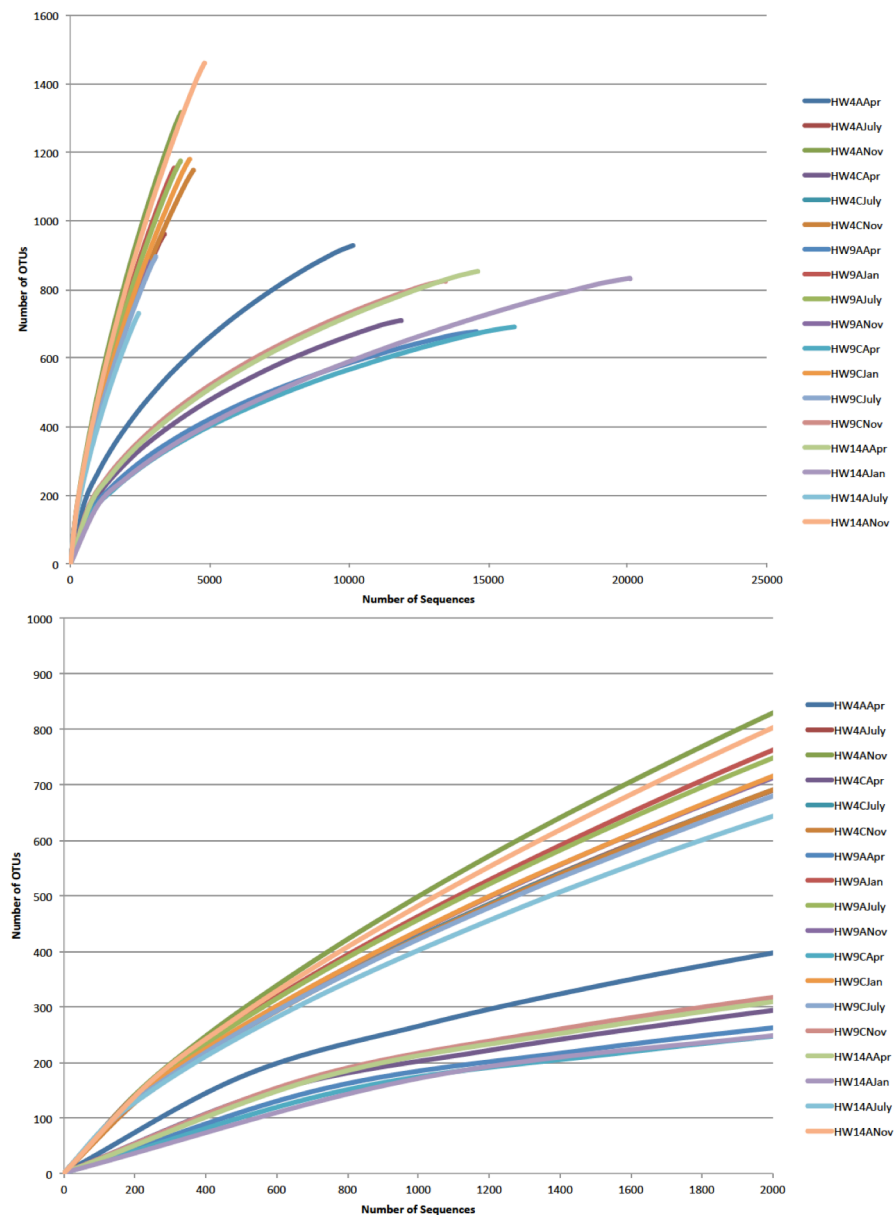


Figure 4-Rarefaction plot of sample IDs for Hollywood. The plateau of some samples very large indicating that the sequence diversity of that sample has been covered. For the other samples, the sequencing depth is low.

Alpha Diversity

Abundant Taxa Alpha diversity is the measure of diversity within a group (Whittaker 1972; Sepkoski, 1988). CloVR was used to determine the abundant taxa associated with each site type. The skiff plot (Figure 5) is a visualization tool, developed to perform unsupervised clustering of samples. The relative microbial abundance is shown by the spectrum of colors on a logarithmic scale. Red indicates more abundant while yellow indicates the less common taxa. The skiff plot also normalizes the data with each sample and calculates Euclidean distance between row and column vectors. Skiff plots are generated using the R packages gplots and RColorBrewer incorporated into CloVR (White et al., 2011). Some taxa are shown to be more present more often than others. The most common classes are the Cyanobacteria, Flavobacteria, Alphaproteobacteria, Gammaproteobacteria, and unidentified bacteria.

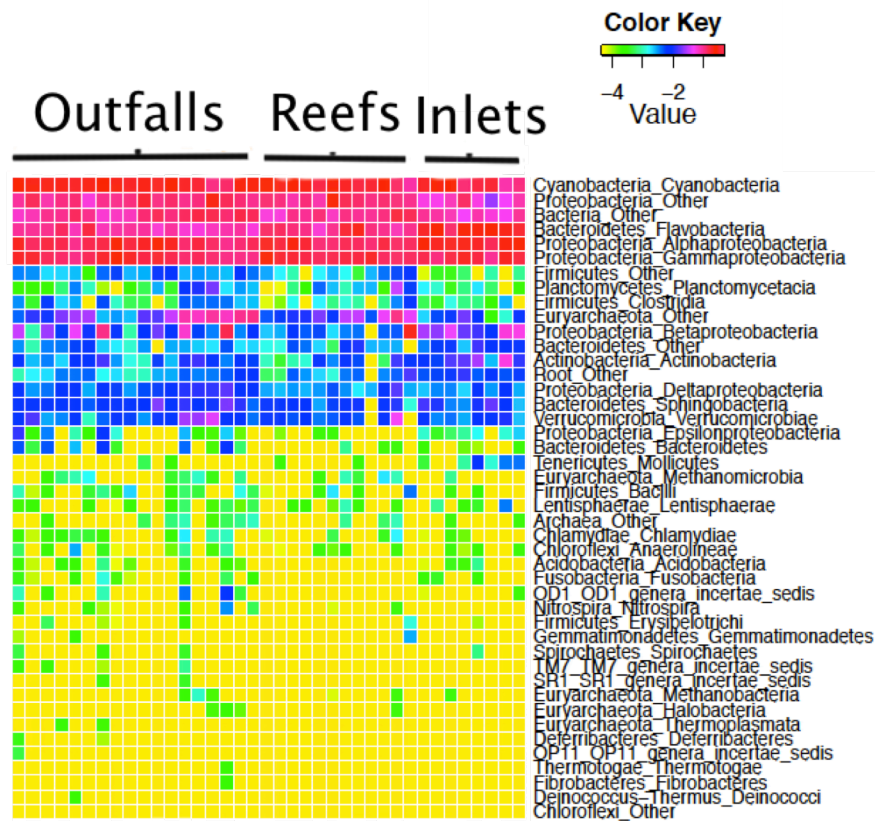


Figure 5-Skiff plot of microbial diversity. Each group is sorted by class. The classes in red are the more abundant and the green/yellow classes occur in low abundance.

The taxa summary bar chart (Figure 6) showed the most abundant members across all samples. The abundant taxa were present in greater than or equal to 1% across all samples. The common taxa include the Gammaproteobacteria, Alphaproteobacteria, Cyanobacteria, Bacteroidetes. Within the abundant taxa, the class Gammaproteobacteria showed seasonal significant changes ($p < 0.001$).

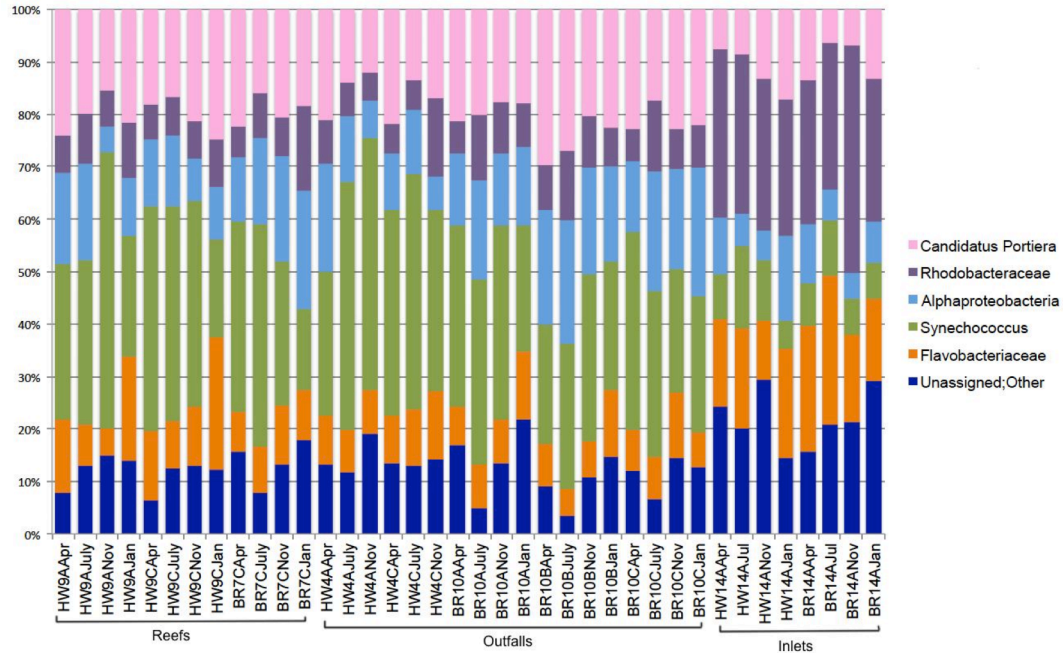


Figure 6-Taxa summary of the abundant taxa to the genus level.

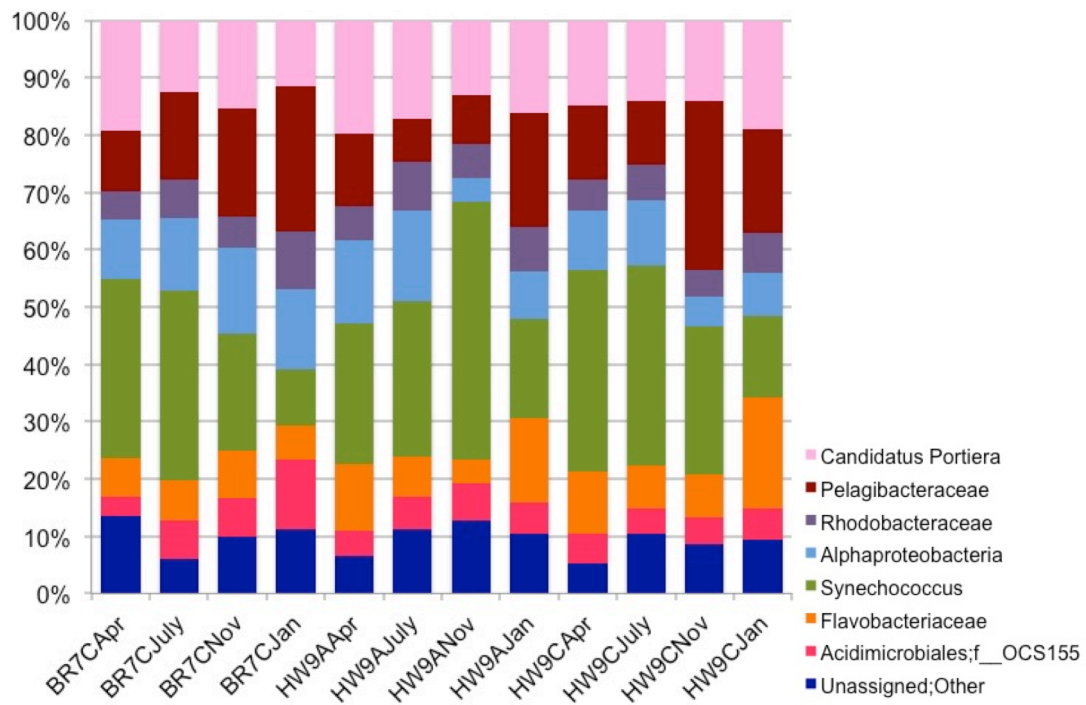
The results of the BLASTN searches showed 70 total methanogenic strains of Euryarcheota. Many are gut symbionts, but there were some matches to deep-sea archaea. A reason for this could be that since Greengenes and RDP Classifier are still being updated, new taxa are not available and the best match was picked.

Each site had its own abundant taxa indicated in Table 4. Numbers generated were based on the matches to the Greengenes 13_8 database. Overall, the most abundant taxa were the families Flavobacteriaceae and Rhodobacteraceae, and the genera *Synechococcus* and *Candidatus Portiera*.

	Flavobacteriaceae	Synechococcus	Alphaproteobacteria	Rhodobacteraceae	Candidatus Portiera
HW4AApr	5.032%	14.832%	11.121%	4.564%	11.416%
HW4AJuly	4.913%	28.322%	7.517%	3.726%	8.375%
HW4ANov	4.292%	24.565%	3.625%	2.697%	6.119%
HW4CApr	5.036%	21.717%	5.945%	3.042%	12.126%
HW4CJuly	6.819%	28.279%	7.782%	3.490%	8.504%
HW4CNov	7.583%	20.317%	3.804%	8.798%	9.934%
BR10AApr	3.505%	16.554%	6.560%	2.851%	10.208%
BR10AJuly	4.412%	18.542%	9.910%	6.522%	10.550%
BR10ANov	3.310%	14.718%	5.430%	3.953%	6.954%
BR10AJan	3.538%	6.601%	4.075%	2.274%	4.927%
BR10BApr	4.583%	13.122%	12.487%	4.904%	17.077%
BR10BJuly	2.804%	14.953%	12.773%	7.165%	14.642%
BR10BNov	2.828%	13.178%	8.331%	4.090%	8.331%
BR10BJan	4.332%	8.326%	6.089%	2.503%	7.634%
BR10CApr	4.816%	23.490%	8.358%	3.866%	14.074%
BR10CJuly	4.054%	15.687%	11.376%	6.714%	8.701%
BR10CNov	4.037%	7.499%	6.119%	2.507%	7.266%
BR10CJan	2.282%	8.949%	8.438%	2.733%	7.628%
HW9AApr	9.094%	19.414%	11.354%	4.636%	15.612%
HW9AJuly	4.820%	18.788%	11.025%	5.830%	11.861%
HW9ANov	3.074%	31.865%	2.908%	4.155%	9.265%
HW9AJan	9.680%	11.264%	5.468%	5.079%	10.577%
HW9CApr	8.758%	28.528%	8.407%	4.382%	12.118%
HW9CJuly	5.705%	25.945%	8.503%	4.651%	10.465%
HW9CNov	5.582%	19.363%	4.017%	3.441%	10.557%
HW9CJan	12.507%	9.236%	4.945%	4.500%	12.271%
BR7CApr	4.607%	21.951%	7.317%	3.523%	13.550%
BR7CJuly	4.817%	22.757%	8.727%	4.618%	8.573%
BR7CNov	3.774%	9.376%	6.834%	2.503%	6.992%
BR7CJan	2.477%	4.025%	5.882%	4.180%	4.799%
HW14AApr	7.505%	3.930%	4.873%	14.549%	3.377%
HW14AJul	9.413%	7.616%	2.980%	14.853%	4.257%
HW14ANov	5.184%	5.412%	2.558%	13.359%	6.166%
HW14AJan	8.495%	2.166%	6.529%	10.526%	6.999%
BR14AApr	10.751%	3.507%	5.004%	12.206%	6.050%
BR14AJul	10.976%	4.017%	2.296%	10.760%	2.439%
BR14ANov	7.054%	2.905%	2.075%	18.257%	2.905%
BR14AJan	7.158%	3.085%	3.548%	12.311%	6.048%

Table 4-Abundant (>1%) of taxa across all sites.

Taxa identified as abundant at the reefs were the family OCS155, which belongs to the order Acidimicrobiales and the family Pelagibacteraceae, in addition to the overall abundant taxa (Figure 7; Table 5). A similar make up was found at the outfalls (Figure 8; Table 6).



	OCS155	Synechococcus	Pelagibacteraceae
BR7CApr	2.439%	21.951%	7.317%
BR7CJuly	4.684%	22.757%	10.694%
BR7CNov	3.139%	9.376%	8.661%
BR7CJan	5.108%	4.025%	10.526%
HW9AApr	3.473%	19.414%	9.985%
HW9AJuly	4.069%	18.788%	5.368%
HW9ANov	4.570%	31.865%	6.107%
HW9AJan	3.526%	11.264%	12.997%
HW9CApr	4.265%	28.528%	10.341%
HW9CJuly	3.161%	25.945%	8.430%
HW9CNov	3.566%	19.363%	22.158%

Table 5-Abundant taxa unique to the Broward and Hollywood reef sites.

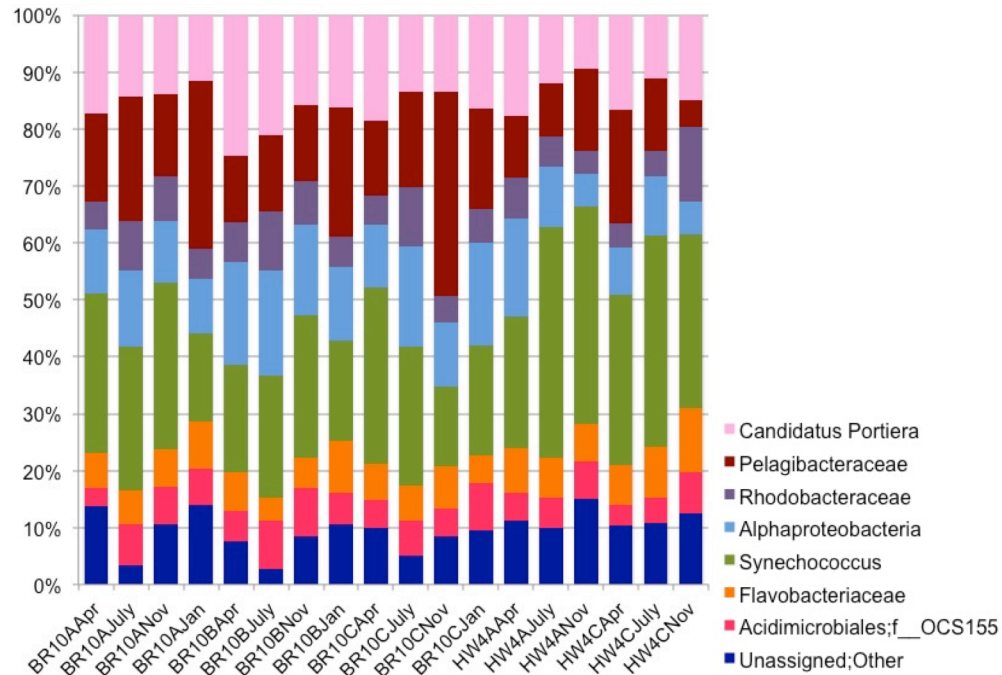


Figure 8-Taxa summary of abundant outfall taxa. Note that the taxa distribution between this and Figure 7 are identical.

	OCS155	Synechococcus	Pelagibacteraceae
BR10AApr	1.931%	16.554%	9.084%
BR10AJuly	5.307%	18.542%	16.113%
BR10ANov	3.286%	14.718%	7.192%
BR10AJan	2.716%	6.601%	12.571%
BR10BApr	3.731%	13.122%	8.135%
BR10BJuly	5.919%	14.953%	9.346%
BR10BNov	4.519%	13.178%	6.993%
BR10BJan	2.645%	8.326%	10.811%
BR10CApr	3.750%	23.490%	9.933%
BR10CJuly	3.926%	15.687%	10.816%
BR10CNov	2.592%	7.499%	19.397%
BR10CJan	3.904%	8.949%	8.228%
HW4AApr	3.111%	14.832%	6.953%
HW4AJuly	3.627%	28.322%	6.462%
HW4ANov	4.176%	24.565%	9.194%
HW4CApr	2.702%	21.717%	14.426%
HW4CJuly	3.450%	28.279%	9.627%
HW4CNov	4.756%	20.317%	3.170%

Table 6-Abundant taxa unique to the outfalls.

The taxonomic make up of the inlets included the overall abundant taxa. Their individual abundant taxa make up included the family Cryomorphaceae, the order Stramenopiles, and the family OM60, which belongs to the order Alteromonadales (Figure 9; Table 7).

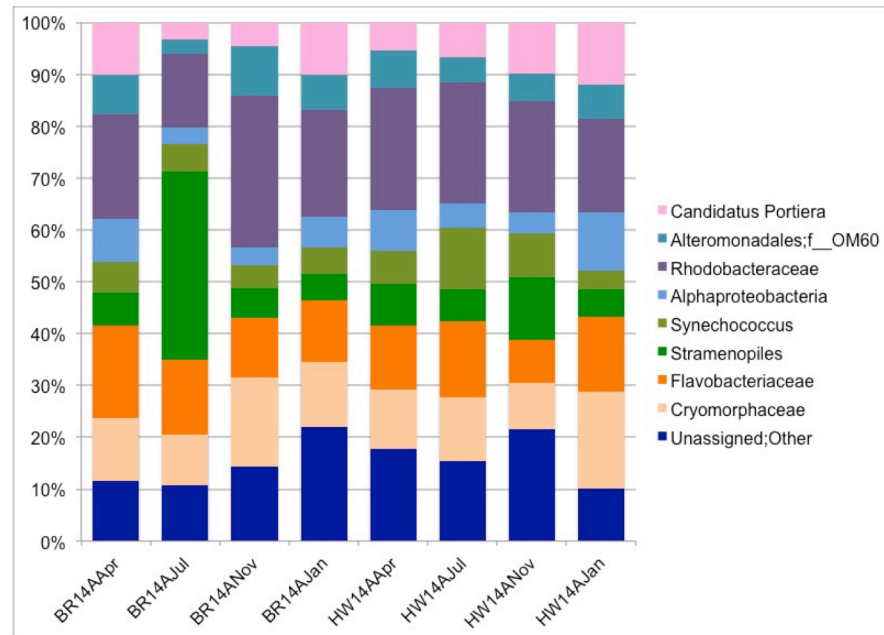


Figure 9-Abundant taxa found at the inlet sites. Aside from the abundant taxa across all samples, note that there is a higher occurrence of the family Cryomorphaceae, which belong the the family Flavobacteriales, and Stramenopiles, which are autrophic eukaryotes.

	Cryomorphaceae	Stramenopiles	Alteromonadales
BR14AApr	7.317%	3.821%	4.554%
BR14AJul	7.389%	27.403%	2.116%
BR14ANov	10.788%	3.596%	5.947%
BR14AJan	7.467%	3.024%	4.042%
HW14AApr	7.129%	4.951%	4.391%
HW14AJul	7.758%	3.832%	3.027%
HW14ANov	5.526%	7.582%	3.403%
HW14AJan	10.837%	3.017%	3.737%

Table 7-Abundant taxa unique to the inlets.

Rare Microbiota At the inlet sites, there was an increase in the number of Tenericutes, which contains the class Mollicutes, share a common ancestor with Firmicutes (Davis et al., 2013). Only recently has Firmicutes been divided into 3 classes:

the Bacilli, Clostridia, and Erysipelotrichia. Mollicutes are part of the Tenericutes phylum since they lack a cell wall, though Mollicutes have also been placed into the Firmicutes phylum (Davis et al., 2013). Lentisphaerae were also found in the inlet samples. Previous studies have found Lentisphaerae in landfill leachate (Chouari et al., 2005; Limam et al., 2010) and are also found in marine datasets (Cho et al., 2004).

At the outfall sites, Bacteroidetes were in higher abundance. Bacteroidetes belongs to the Cytophaga-Flavobacteria-Bacteroidetes (CFB) cluster, which are typically associated with particulate matter, though there are many free-living species (Abdell and Bowman, 2005). Bacteroidetes, particularly the class Flavobacteria, are abundant on organic particles (Abdell and Bowman, 2005).

At the reef sites, there were higher occurrences of Planctomycetes, which are free-living bacteria lacking a peptidoglycan cell wall and have a membrane similar to that of eukaryotic cells (Fuerst et al., 1997). Planctomycetes are collected from various aquatic ecosystems (Fuerst et al., 1997; Pizzetti et al., 2011). Planctomycete numbers appear correlated to higher algal content, suggesting that they feed on algal degradation products (Pizzetti et al., 2011). This is in contrast to the environmental metadata collected by NOAA, as the chlorophyll a and phaeopigments, the degradation product of algal chlorophyll, were not present in any higher numbers at the reefs than at other sites.

Rank Abundance Rank abundance curves compare the abundance of units (species, sequence reads, etc.) being counted and gives them a rank (Figures 10 and 11). For example a score of 1 means that species is the most abundant, followed by 2 as the second abundant, et cetera. On the x-axis is the rank assigned to each taxa and on the y-axis is the richness. The slope determines the sample evenness, with the steeper slope meaning there is low evenness, and the higher ranked species are at higher abundance. Another measure a rank abundance curve can give is the number of singletons in a sample, which affect the evenness of a sample. The evenness of a sample shows how equal each species is represented (Peterson, 1975).

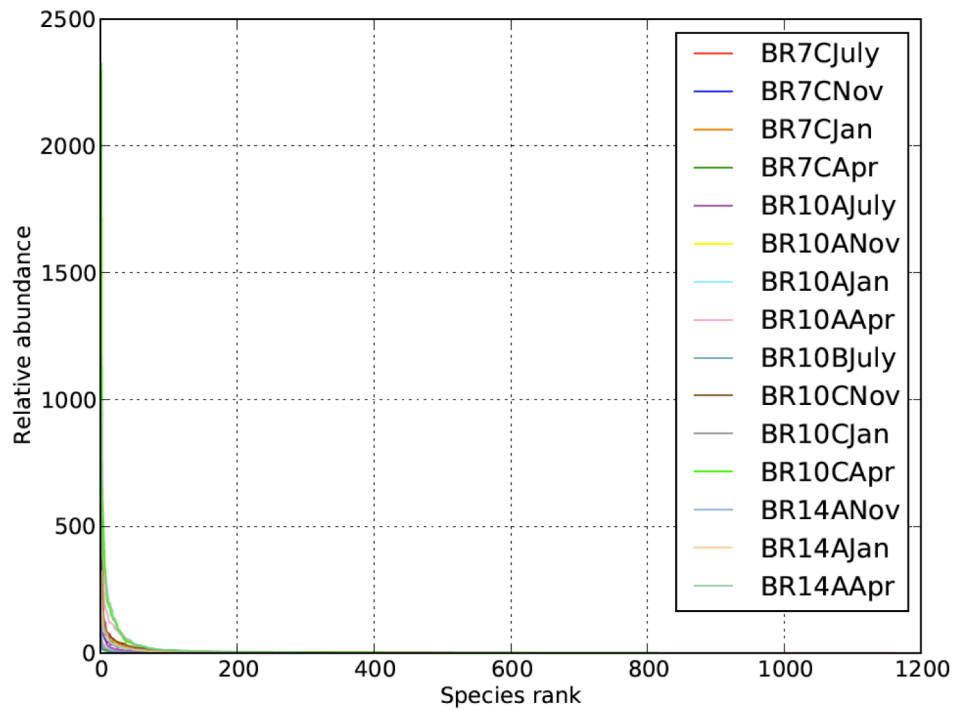


Figure 10-Broward rank abundance plot of all samples.

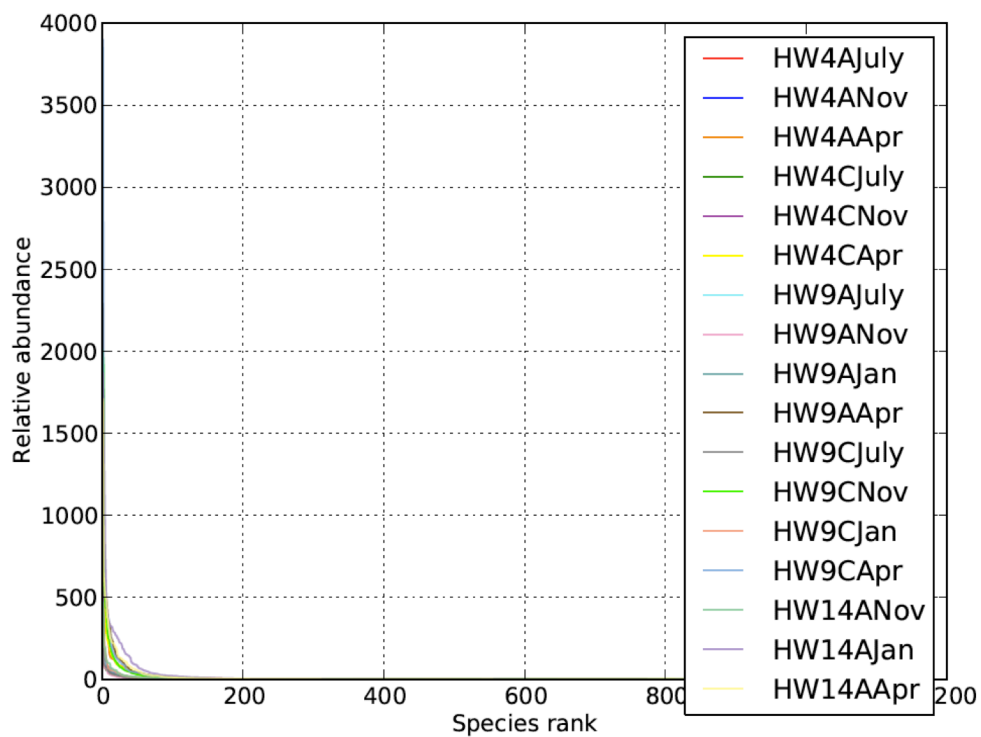


Figure 11-Hollywood rank abundance plot of all samples.

Broward The Broward reef samples had steep slopes and most had low abundance. The April reef sample had a long tail, indicating an abundance of singletons

and/or rare biosphere bacteria. This was the case across July and November. In January, the rare species tail was shorter. The Broward April outfall samples have short slopes, and long tails for the rare biosphere and singletons. A similar curve was seen for July and November. The January Broward outfall samples have a shorter tail at the end, indicating less singletons and rare biosphere bacteria. The Broward inlet samples in April and July had long tails indicating many singletons and rare biosphere bacteria. The November inlet sample had a steep slope and smaller tail.

Hollywood The relative abundances for the Hollywood reef in April had short tails and tall slopes. Reef samples from July had steep slopes and long tails, indicating many singletons and rare biosphere bacteria. In November, the surface reef sample had a long tail, and the bottom had a short tail. In January, the slopes from the surface and bottom were both steep, and they had short tails, indicating many abundant bacteria and few singletons. The slopes of the rank abundance curves were steep and had long tails, indicative of a high abundance of bacteria, and many singletons. In July, the slopes were steep, indicating several high abundance bacteria and many singletons. The November samples had steep slopes and many singletons. The Hollywood inlet sample from April had a steep slope and high singleton counts, similar to the other samples. The samples from July had a steep slope, but a short tail indicating a low singleton count. This may be due to the low sequence count. The sample from November had a steep slope and long tail, which indicates a high singleton count. The sample from January had a steep slope and long tail, indicating many singletons.

Most of the rank abundance plots showed steep slopes and long tails. The long tails are made up of singletons, a single read found in a sample that can either be an artifact of the sequencing platform, or a member of the rare biosphere of bacteria. Similarly, doubletons, double reads, can also occur in next-generation sequencing datasets. Singletons made up a large part of the tail on the rank abundance curve. The total observations between the unfiltered OTU table and the one filtered for singletons decreased from 7595 to 2691, a difference of 4904. Further, after filtering doubletons, the total number of observations between all samples decreased to 2013.

The OTU BIOM summary files are listed in the Appendix.

Pathogens Pathogenic strains of bacteria were picked from the OTU table in QIIME and analyzed using BLASTN. The pathogenic strains found in the water samples were of the classes *Gammaproteobacteria*, the phylum *Firmicutes*, and *Epsilonproteobacteria*. Pathogenic members of *Gammaproteobacteria* include the families *Vibrionales* and *Enterobacteriales*. Disease-causing members of the phylum *Firmicutes* include the class *Clostridia* and the order *Bacillales*. Members of the *Epsilonproteobacteria* class include the order *Campylobacteriales*, which include the genera *Campylobacter* and *Arcobacter*. Pathogenic member of the genus *Vibrio* were also found in water samples. These occurrences of pathogenic bacteria were low among all samples.

Beta Diversity

Beta diversity is the measure of diversity between sites (Whittaker, 1972; Sepkoski, 1988). It was measured using principal coordinates of analysis (PCoA), which compares groups based on counts or phylogenetic analysis. The groups compared were site type and season. The resulting PCoA plots for site type were analyzed using an unweighted Unifrac (Hamady et al., 2010) and a weighted Unifrac distance matrix, a phylogenetic measure. The box plots were generated using a Bray-Curtis distance matrix, a non-phylogenetic measure. The PCoA plots for season were also generated using unweighted and weighted Unifrac. Unweighted Unifrac shows how communities can differ based on what lives in them. Weighted Unifrac looks at how each sample differs based on taxa abundance (Lozupone et al., 2007).

Site Type We examined the changes in beta diversity by site type-reefs, outfalls, inlets. Each site type was divided into Broward (BR) and Hollywood (HW). The results of the Principal Coordinates of Analysis using weighted Unifrac graph show a distinct clustering of the Broward (red) and Hollywood (green) inlets (Figures 12 and 13). The Hollywood reefs and outfalls cluster, and the Broward reefs and outfalls cluster based on site type. Statistically significant differences were found comparing the site type (by ANOSIM, $R=0.4302$; $p=0.001$; 999 permutations).

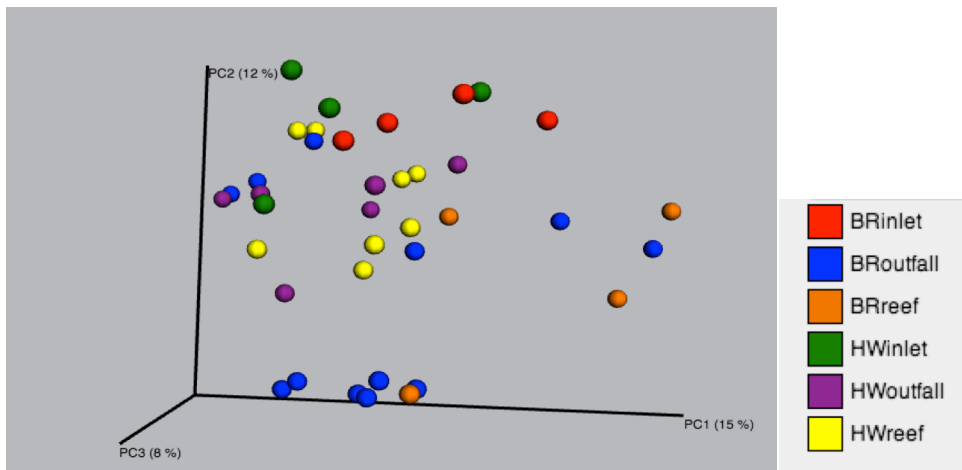


Figure 12-Principal Coordinates of Analysis plot of the different site types split into site types using unweighted Unifrac.

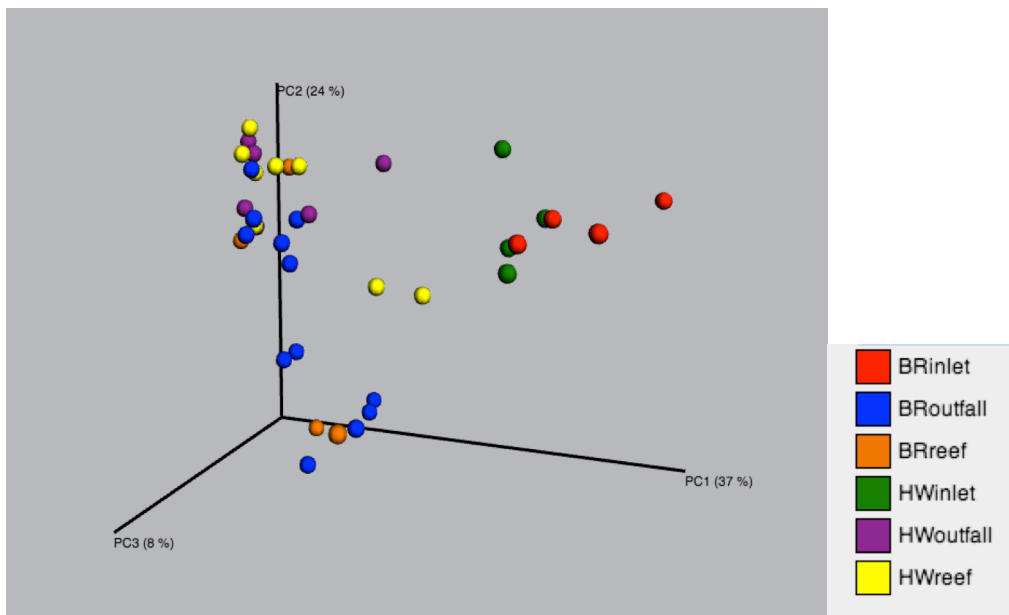


Figure 13-Weighted Unifrac Principal Coordinates of Analysis plot of the site types.

For the Bray-Curtis box plot (Figure 14), a score closer to 0 shows more shared species between each site, and a score closer to 1 would indicate fewer shared species. For most sites, the median dissimilarity score is close to 0.5, an intermediate score. When compared to the Hollywood reef site, the Broward outfall is around 0.6, indicating that there are some similarities in the communities for each site, but also some distinct species. The HW reef site compared to its inlet site shows a score closer to 1, indicating that there are fewer shared species. Similarly, the BR inlet and HW reef also show fewer shared species, as does the BR outfall compared to its own inlet site. When the Broward

inlet is compared to the Hollywood inlet, the dissimilarity score is near the intermediate, which indicates that species composition between the two sites is similar. The same phenomena occur with the outfall and reef sites as well.

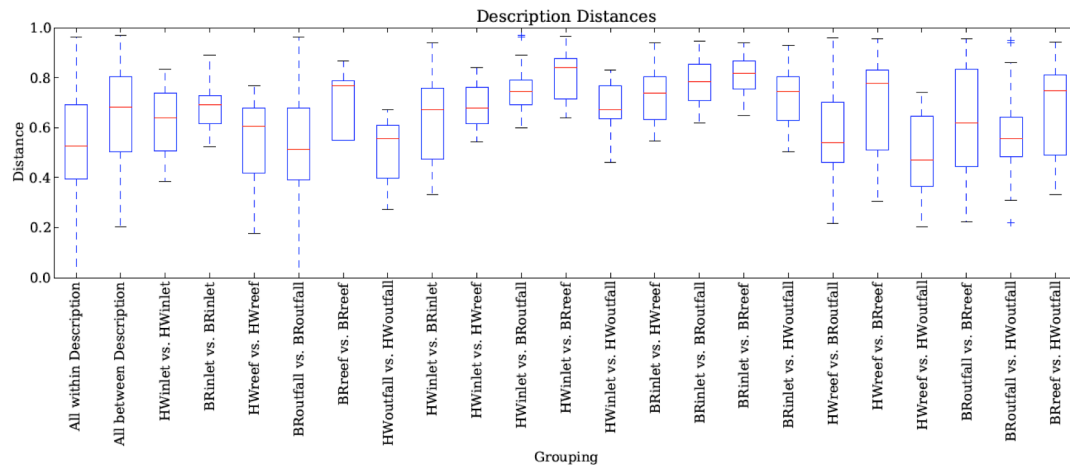


Figure 14-Box plot of the sample sites using Bray-Curtis dissimilarity analysis.

Season Microbial community composition showed significant changes between seasons. In south Florida, the fall and winter months, October-early March, are known as the dry season. The spring and summer months, April-September, are noted as being the rainy season in south Florida. Few bacterial community members showed clustering with the same PCoA plot, and most overall changes remained were not significant (Figure 15-16). However, some classes of bacteria and archaea showed significant seasonal changes between the wet and dry seasons, regardless of site, are the order Acidimicrobiales ($p=0.0423$), the family Balneolaceae of the order Rhodothermales ($p=0.0386$), the chloroplast order Cryptophyta ($p=0.0335$), Alphaproteobacteria ($p=0.002$), the family Rhodospirallaceae ($p=0.0375$), Deltaproteobacteria order Sva0853 ($p=0.0439$), the family Thiohalorhabdaceae ($p=0.0466$), and the Verruco-5 order R76-B128 ($p=0.0375$). Significant differences were shown using ANOSIM analysis ($R=0.2741$; $p\text{-value}=0.001$; 999 permutations).

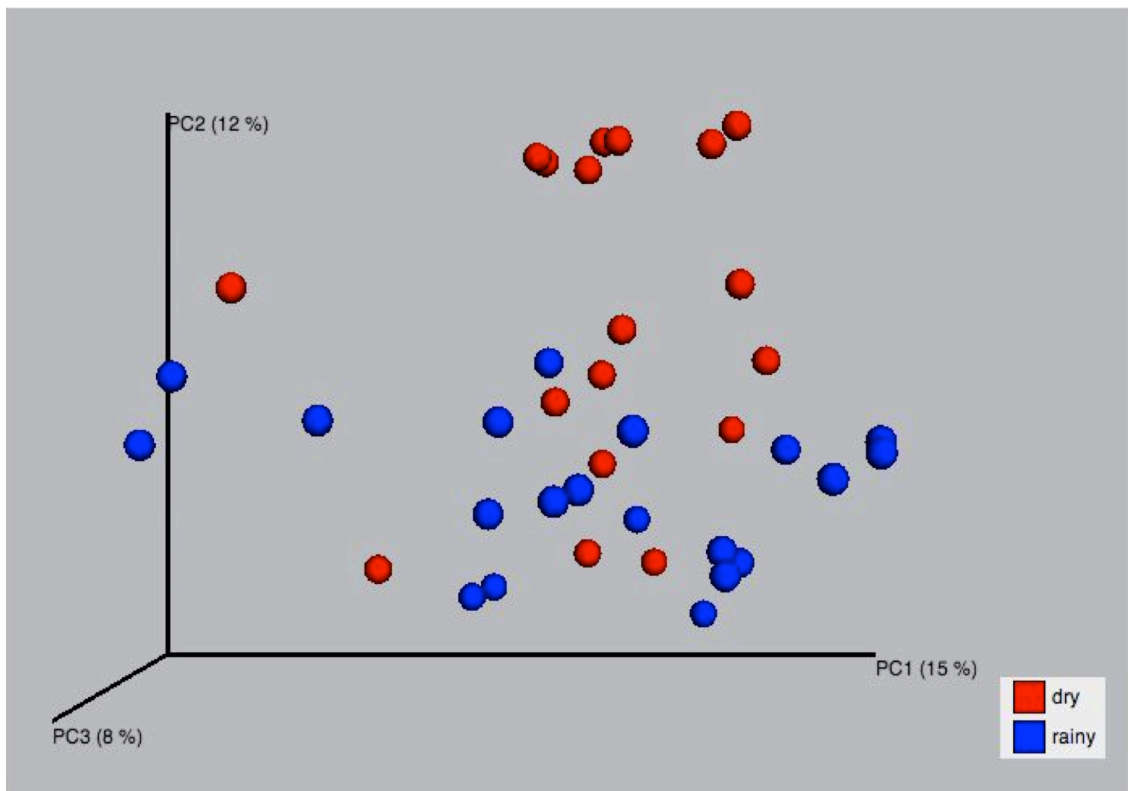


Figure 15-PCoA plot of the seasonal distribution of samples using unweighted Unifrac.

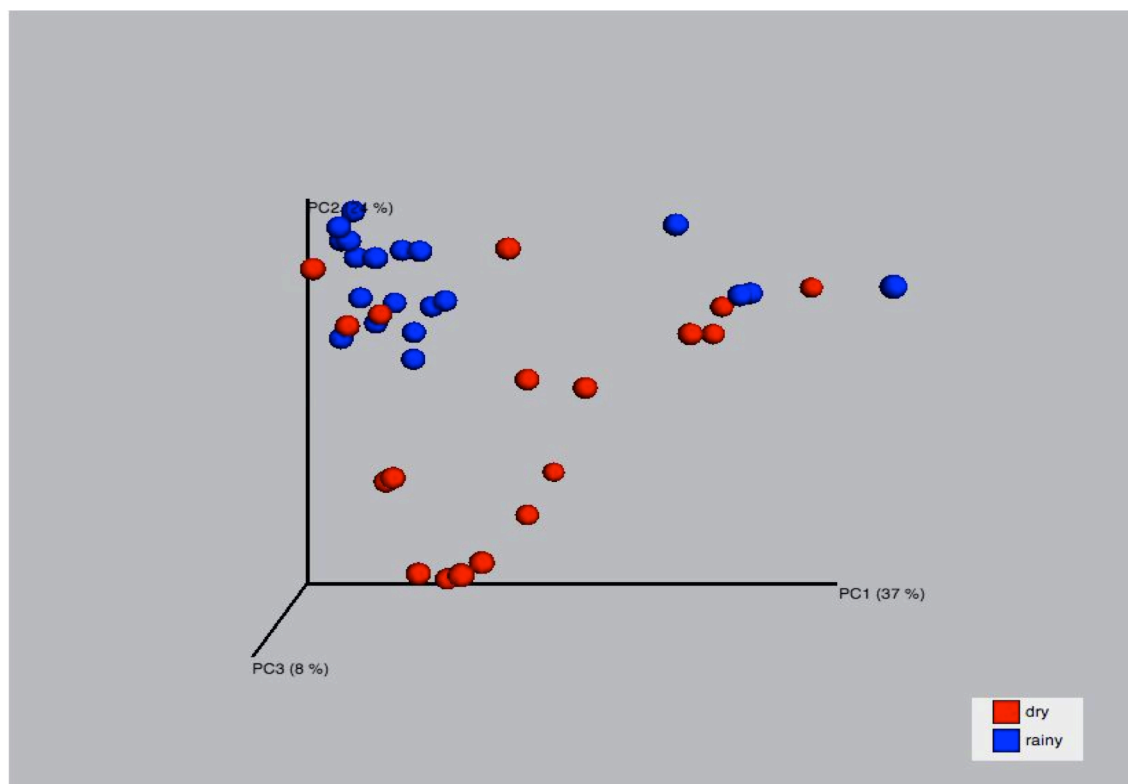


Figure 16-Weighted Unifrac of the Principal Coordinates by season.

Depth There were no significant changes found with depth. Depth was calculated using non-numerical values (surface, mid-depth, bottom). Microbial communities were mostly uniform with depth (by ANOSIM, $R=0.0495$; $p\text{-value}=0.150$; 999 permutations).

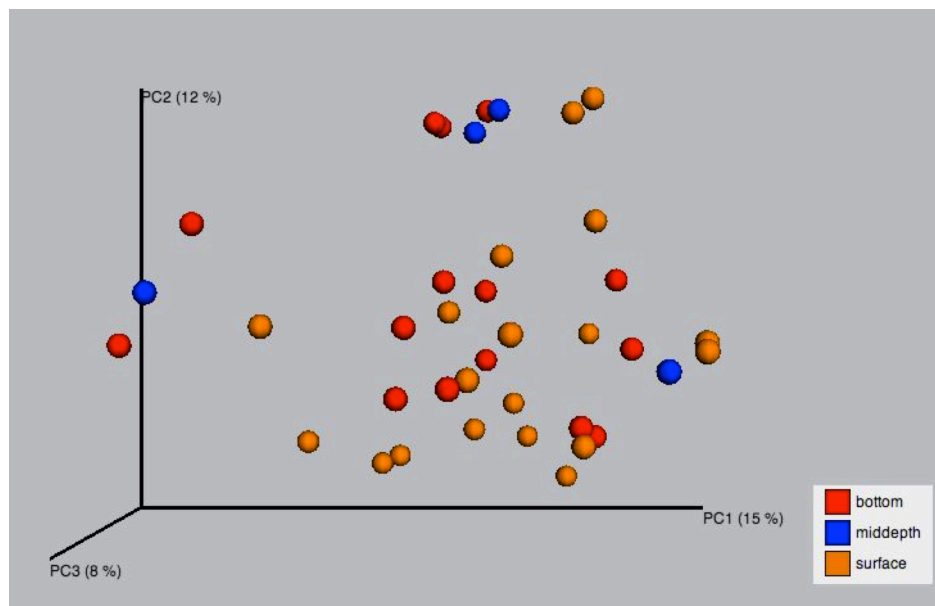


Figure 17-Unweighted Unifrac Principal Coordinates plot by depth.

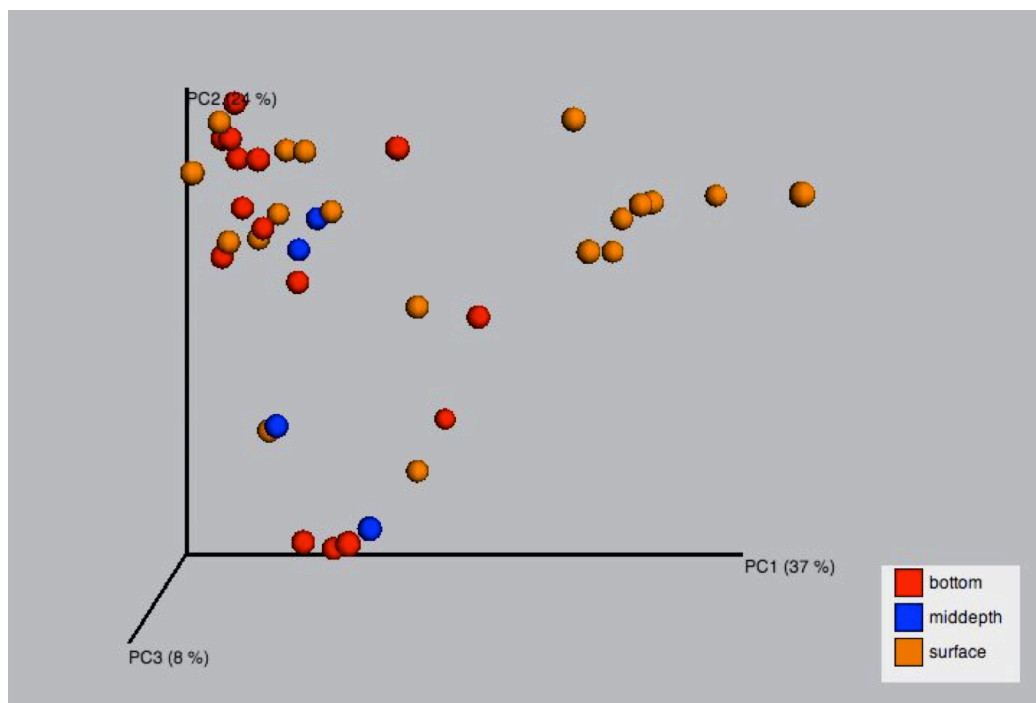


Figure 18-Weighted Unifrac Principal Coordinates of Analysis by depth.

SourceTracker

SourceTracker (Knights et al., 2011) is an R script originally used to determine a possible source of laboratory contamination in 16S rRNA and metagenomic sequencing sets. Previous iterations were limited to predetermined indicator species, but as sequencing has grown, it has become a vital tool for comparing similarities between communities. An investigator assigns an environment as either a source or sink. Assessment of community similarity can be performed with Unifrac or abundance of taxa analysis, but neither of these methods gives the proportion of similarity between samples. Two analyses were performed with SourceTracker, with the outfalls (Table 8) and inlets (Table 9) set as the source, respectively.

Outfalls as a Source

Broward Inlet

In the month of April 2011, the Broward outfall showed 0.00% influence on the Broward Inlet. The Hollywood outfall showed 68.71% influence and there was a 31.29% influence by an unknown source on the Broward Inlet. During July 2011, the influence of the Broward outfall was 0.00%, the Hollywood outfall influence rose to 24.33%, and the unknown source was 75.67%. In November, there was 0.00% influence by the Broward outfall. There was a 52.90% influence found with the Hollywood outfall and a 47.09% influence by an unknown source. In January 2012, there was a similar profile with again, no influence by the Broward outfall, but a 61.09% influence by the Hollywood outfall and a 38.83% influence by an unknown source.

Hollywood Inlet

The Hollywood Inlet was not influenced by the Broward outfall in the month of April 2011, but was influenced by the Hollywood outfall at 61.55% and 38.45% by an unknown source. In July 2011, there was again no influence from the Broward outfall, but and 52.31% influence by the Hollywood outfall and an 47.69% influence by an unknown source. In November, there was a 0.00% influence by the Broward outfall, but the Hollywood outfall's influence remained high at 73.05% and there was still a slight influence by an unknown source at 26.95%. In January 2012, there was a 0.00%

influence by the Broward outfall, but the Hollywood outfall still had the strongest influence at 78.04% and an unknown source at 21.96%.

Broward Reef

At the Broward reef in April 2011, there was a moderate influence by the Broward outfall at 34.73%, an influence of 65.27% by the Hollywood outfall and a 0.00% influence by an unknown source. In July 2011, the microbial communities matched the Broward outfall at 40.7%, 57.07% with the Hollywood outfall, and 2.23% with an unknown source. During November 2011, the Broward outfall showed the highest influence at 99.92%. The lowest influence was with the Hollywood outfall at 0.00%, and there was a 0.00% influence with an unknown source.

Hollywood Reef

At the Hollywood reef site in April 2011, there was an influence of 5.66% (average of surface 11.33% and bottom 0.00%) with the Broward outfall, 94.33% (surface=88.67%, bottom=100%) with the Hollywood outfall, and an overall 0.00% influence by an unknown source. In July 2011, there was a low match of 11.62% (surface=17.47%, bottom=5.78%) with the Broward outfall, a high match of 85.16% (surface=76.17%, bottom=94.14%) with the Hollywood outfall, and a <0.00% match with an unknown source. During November 2011, there was a 54.05% (surface=5.53%, bottom=48.52%) influence by the Broward outfall, a 72.73% (surface=93.97%, bottom=51.48%) influence by the Hollywood outfall, and a 0.00% influence by an unknown source. In January 2012, there was a match of 49.24% (surface=61.00%, bottom=37.48%) with the Broward outfall, 50.31% match (surface=39.00%, bottom=61.62%) with the Hollywood outfall, and a <0.00% influence with an unknown source.

SampleID	BRoutfall	HWoutfall	Unknown
BR7CApr	34.73%	65.27%	0.00%
BR7CJuly	40.70%	57.07%	2.23%
BR7CNov	99.92%	0.00%	0.08%
BR7CJan	98.85%	1.15%	0.00%
HW9AApr	11.33%	88.67%	0.00%
HW9CApr	0.00%	100.00%	0.00%
HW9AJuly	17.47%	76.17%	6.36%
HW9CJuly	5.78%	94.14%	0.08%
HW9AJan	61.00%	39.00%	0.00%
HW9CJan	37.48%	61.62%	0.90%
HW9ANov	5.53%	93.97%	0.50%
HW9CNov	48.52%	51.48%	0.00%
BR14AApr	0.00%	68.71%	31.29%
BR14AJul	0.00%	24.33%	75.67%
BR14ANov	0.00%	52.90%	47.10%
BR14AJan	0.08%	61.09%	38.83%
HW14AApr	0.00%	61.55%	38.45%
HW14AJul	0.00%	52.31%	47.69%
HW14ANov	0.00%	73.05%	26.95%
HW14AJan	0.00%	78.04%	21.96%

Table 8-Summary of the percent influence of the Broward and Hollywood outfalls. The outfalls were used as a source and the reefs and inlets were sinks.

SourceTracker graphical output for the outfalls is listed in the Appendix.

Inlets as a source

Broward Outfalls

Given the extreme difference of the inlets from the reefs and outfalls, the inlets were set as the source, and the outfalls and reefs were set as sinks, given their similarities. Across all seasons, the Hollywood inlet is shown as a major contributor to the microbial communities of the Broward outfalls. The Broward inlet had a 0.00% contribution to the microbial communities for the outfalls. For April 2011, the Hollywood inlet contributed to 57.08%(surface=62.37%, mid-depth=41.67%, bottom=67.19%) of the microbial mass of the Broward outfall. In July 2011, the Hollywood inlet contributed approximately 62.6228% to the Broward outfall. In November 2011, the Hollywood inlet contributed 54.3367% to the Broward outfall. In January 2012, the Hollywood inlet contributed 49.7334%.

Hollywood Outfalls

In April 2011, the Hollywood outfall had a 66.18% similarity (surface=61.42%, bottom=70.93%) on the Hollywood inlet, no match to the Broward outfall, and an average of 33.83% (surface=38.58%, bottom=29.07%) of an unknown source. In July 2011, the Hollywood outfall sample again showed no similarity to the Broward inlet. There was an average of 76.91%(surface=77.26%, bottom=76.55%) to the Hollywood inlet, and a 22.94%(surface=22.74%, bottom=23.41%) match to an unknown source. In November 2011, there was a 90.06% (surface=81.26%, bottom=90.76%) similarity to the Hollywood inlet and none from the Broward inlet. There was also a 13.99% (surface=18.73%, bottom=9.24%) similarity to an unknown source. No Hollywood outfall samples were sequenced for January.

Broward Reefs

At the reef site in April 2011, there was a 62.51% similarity to the Hollywood inlet, none from Broward inlet, and a 37.49% match to an unknown source. In July 2011, there was a 65.76% match to the Hollywood inlet, about 0.00% from the Broward inlet and a 34.24% match to an unknown source. At the reef in November 2011, there was a 38.74% match to the Hollywood inlet, no match to the Broward inlet, and a 61.26% match to an unknown source. In January 2012, there was a 51.03% match to the Hollywood inlet, no match to the Broward inlet, and a 48.97% match to an unknown source.

Hollywood Reefs

In the month of April 2011, there was a 74.51% (surface=64.41%, bottom=84.61%) match to the Hollywood Inlet, none to the Broward inlet, and a 25.49%(surface=35.59%, bottom=15.39%) match to an unknown source. In July 2011, there was no match to the Broward inlet, a 63.29%(surface=56.89%, bottom=69.69%) match to the Hollywood inlet, and a 32.31% (surface=43.07%, bottom=30.31%) to an unknown source. In November 2011, there was a 82.55% (surface=84.37%, bottom=80.72%) match to the Hollywood Inlet, none to the Broward Inlet, and a 17.46% (surface=15.63%, bottom=19.28%) match to an unknown source. For January 2012, there

was a 75.67% (surface=73.07%, bottom=78.16%) match to the Hollywood Inlet, none to the Broward Inlet, and a 24.39% (surface=26.93%, bottom=21.84%) match to an unknown source.

SampleID	BRinlet	HWinlet	Unknown
BR7CApr	0.00%	62.51%	37.49%
BR7CJuly	0.00%	65.76%	34.24%
BR7CNov	0.00%	38.74%	61.26%
BR7CJan	0.00%	51.03%	48.97%
BR10AApr	0.00%	62.37%	37.63%
BR10BApr	0.00%	41.67%	58.33%
BR10CApr	0.01%	67.19%	32.80%
BR10AJuly	0.00%	66.65%	33.35%
BR10BJuly	0.00%	46.98%	53.02%
BR10CJuly	0.00%	52.21%	47.79%
BR10ANov	0.00%	46.89%	53.11%
BR10BNov	0.00%	41.73%	58.27%
BR10CNov	0.00%	43.63%	56.37%
BR10AJan	0.00%	53.46%	46.54%
BR10BJan	0.12%	37.53%	62.35%
BR10CJan	0.00%	41.15%	58.85%
HW4AJuly	0.00%	77.26%	22.74%
HW4CJuly	0.04%	76.55%	23.41%
HW9AJuly	0.04%	56.89%	43.07%
HW9CJuly	0.00%	69.69%	30.31%
HW4ANov	0.01%	81.26%	18.73%
HW4CNov	0.00%	90.76%	9.24%
HW9ANov	0.00%	84.37%	15.63%
HW9AJan	0.00%	73.07%	26.93%
HW9CJan	0.00%	78.16%	21.84%
HW4AApr	0.00%	61.42%	38.58%
HW4CApr	0.00%	70.93%	29.07%
HW9AApr	0.00%	64.41%	35.59%
HW9CApr	0.00%	84.61%	15.39%
HW9CNov	0.00%	80.72%	19.28%

Table 9- Summary of the percent influence of the Broward and Hollywood inlets. The inlets were used as a source and the reefs and outfalls were sinks.

SourceTracker graphical data for the inlets is listed in the Appendix.

Correlation with environmental metadata

A multiple regression analysis was performed on the four most abundant bacterial taxa and various seawater chemical parameters (Table 8). The R^2 value shows how much the Y-variation is explained by the regression equation and how well the model fits the data. A higher R^2 value means that the model fits the data well (personal communication

with Jay Fleischer). A comparison to the percentage of the abundant taxa was made to environmental metadata was performed. The parameters measured were time, depth, temperature, salinity, conductivity, chlorophyll a, phaeopigments, turbidity, oxygen saturation, total suspended solids (TSS), pH, ORP, total nitrogen, nitrates, ammonium, phosphorus, and silica (Table 10).

	Flavobacteriaceae	<i>Synechococcus</i>	Rhodobacteraceae	<i>Candidatus Portiera</i>	<i>Pelagibacteraceae</i>
R-Squared	0.4817	0.5882	0.9347	0.7431	0.2745
Date	NA	0.0015	0.0117	<.0001	NA
Depth	NA	NA	0.0616	NA	NA
Time	0.0099	NA	NA	0.0400	NA
Salinity	0.0088	NA	<0.0001	0.0069	NA
Temperature	NA	NA	<0.0001	0.0099	NA
pH	NA	NA	0.0414	NA	NA
O ₂ Saturation	NA	NA	<0.0001	NA	NA
ORP+	NA	NA	NA	NA	NA
Conductivity	NA	0.0092	NA	NA	NA
Chlorophyll a	NA	NA	NA	0.0632	0.0007
Phaeopigment	<.0001	0.0155	NA	0.0014	NA
TSS!	NA	0.0427	<0.0001	NA	NA
Total Nitrogen	NA	0.0804	NA	0.004	NA
Nitrate	NA	0.0830	0.0015	NA	NA
Ammonium	NA	NA	NA	NA	NA
Phosphorus	NA	NA	NA	NA	NA
Silica	NA	0.0035	NA	NA	NA

Table 10- Results of multiple linear regression analysis from SAS. The p -value threshold was set at 0.1. Any parameter above 0.1 was discarded. Values greater than 0.05 on this table are considered to be approaching statistical significance and are included. No parameter was consistent among all abundant bacteria. The p -values were obtained from SAS.

+ - Oxygen Reduction Potential

! - Total Suspended Solids

The results for the genus *Synechococcus* ($R^2=0.5882$) showed a significant relationship with conductivity ($p=0.0092$) and phaeopigments ($p=0.0155$), total suspended solids (TSS, $p=0.0427$), and silica ($p=0.0035$). The family *Flavobacteriaceae* ($R^2=0.4817$) showed a statistically significant relationship to time ($p=0.0099$), salinity ($p=0.0088$), and phaeopigments ($p<0.0001$). The family *Rhodobacteraceae* ($R^2=0.9347$) had the most variables with a significant relationship. Rhodobacteraceae showed a significant relationship with date ($p=0.0117$), salinity ($p<0.0001$), temperature ($p<0.0001$), oxygen saturation ($p<0.0001$), pH ($p=0.0414$), total suspended solids ($p<0.0001$), and nitrate ($p=0.0015$). The genus *Candidatus Portiera* ($R^2=0.7431$) showed a significant relationship to date ($p<0.0001$), time ($p=0.0400$), salinity ($p=0.0069$), temperature ($p=0.0099$), phaeopigments ($p=0.0014$), and total nitrogen ($p=0.0004$).

The raw SAS data output is listed in the Appendix.

Discussion

Data collected from this study is the first of its kind in southeast Florida waters. Previous water quality monitoring studies have focused on target organisms, usually coliforms and fecal indicator bacteria (Finkl and Charlier, 2003; Shibata et al., 2004; Brownwell et al., 2007; Hartz et al., 2008; Wright et al., 2009; Abdelzaher et al., 2010; Futch et al., 2011), such as the work done by the FACE program. Studies conducted in South Florida focused mainly on indicator bacteria and utilized quantitative PCR techniques (Carsey et al., in press). It is important to focus on the big picture and eventually incorporate genomics into a larger marine water-monitoring program.

Sample Collection and Sequencing Samples were collected by NOAA using Nalgene bottles and subsequently filtered via vacuum filtration onto a 0.45 micron membrane filter and frozen until ready for extraction. Each site was collected once. Ideally, more water should have been used as previous studies like Venter et al., (2004) and Rusch et al., (2007) used 8 liters of water as opposed to our one liter of water. It is, however, not certain if this would have a major effect on the results, but merely an observation for future monitoring studies.

The sequencing strategy used for this project was 454 pyrosequencing on a GS FLX+ Titanium platform. We sequenced different sites collected, bi-monthly, from each season over the course of a year, rather than the same site over the course of a week or month. This approach was followed to establish a baseline for environmental microbes rather than serving as a true monitoring study. This is a pilot study aimed at characterizing microbial (bacteria and archaea) associated with the outfalls, reefs, and inlets.

Sequencing Issues For the Broward and Hollywood rarefaction curves, there are samples with a low sequence count and steep peaks. There were 3 pyrosequencing runs sent at different dates. Most of the samples with low sequence counts were sent for the first two runs. What may have occurred is either these samples are less diverse than the

final run or there are discrepancies in the machine and sequencing analysis. Select samples were rerun to provide an even sequencing depth.

There are still open questions as to whether pyrosequencing reads are suitable for alpha and non-phylogenetic beta diversity analysis (Reeder and Knight, 2010).

Pyrosequencing is prone to errors and tends to have noise introduced in the PCR and sequencing steps (Reeder and Knight, 2010; Quince et al., 2011). This noise artificially inflates the number of OTUs present (Reeder and Knight, 2010). Pre-filtering of the pyrosequencing reads removes some noise, but a considerable amount still remains (Quince et al., 2011). The current method for denoising pyrosequencing data is to cluster flowgrams, files containing the intensity patterns of each read, before conversion to sequence files, though this method tends to be computationally expensive (Reeder and Knight, 2010). Denoising can be seen as a problem of inferring true sequences and abundances of reads (Quince et al., 2011).

Illumina Sequencing A subset of sequences were also sent for Illumina sequencing on the HiSeq2000 platform at the University of Colorado at Boulder. These samples are currently being evaluated. Illumina is sequencing by synthesis. DNA molecules are washed over a slide and attached with a polymerase. Fluorescently labeled nucleotide bases are incorporated onto the DNA strand and a camera captures the molecules on the slide. The maximum amount of data retrievable from an Illumina HiSeq2000 is 6 billion unpaired reads (illumina.com). Pyrosequencing is declining in popularity, with some labs decommissioning their machines as Illumina is generally cheaper to run, and generates more data. Samples with low read numbers and OTU counts from the 454 pyrosequencing effort increased the number of OTUs present when sequenced with Illumina.

Rank Abundance A rank abundance curve with a steep slope indicates that a sample is dominated by a few species, while the long tail represents lower numbers. This steep slope indicates low evenness. All of the samples showed low evenness. Each sample is dominated by only a few taxa, skewing the curve upward. Then there is a drop off with the remaining taxa occurring only a few times. The singletons represent sequence reads that occur only once in the entire sample. The singletons can be one of two possibilities: They are either an actual taxonomic unit, or they are a sequence or

clustering error. Singletons that are taxonomic units may occur as a more abundant species in another sample.

Pyrosequencing Errors and noise occur in pyrosequencing as either homopolymers or chimeras. Determining the difference between a homopolymer or chimeric sequence versus an actual taxonomic unit is tricky. There are algorithms out there that can remove most of this error, such as AmpliconNoise (Quince et al., 2009), which include a denoising algorithm and a de novo chimera checker or Denoiser (Reeder and Knight, 2010), which is also incorporated into QIIME.

Sequence coverage was not as high as it could be in some samples. The mean sequence count was 5987 per sample, however, the minimum count was 321 and the maximum count was 19988. The total number of unique observations was 4447. This is likely due to the inherent errors of a 454 pyrosequencing machine and possibly errors in the PCR. A subset of samples were run on an Illumina HiSeq 2000 platform, and most of the samples that had low OTU count from the 454 run had higher OTU counts with Illumina. This is not to say that Illumina is completely without error. Some samples that had higher OTU counts with 454 were lower with Illumina. The sequencing depth of Illumina, however, is far greater than 454, and likely, with less samples to run, the OTU counts would have been much higher. An inherent error of both platforms was the multiplexing samples. While more samples could be sequenced at one time and costs are kept down, the capacity of the machine is limited and after a certain number of samples, the number of sequences and OTUs drop.

Changes in the abundant taxa were also important. Thermoplasmata became one of the most abundant taxa in the Illumina data. Pelagibacteraceae, while it was not present greater than one percent of the time in the 454 data, is the most abundant in the Illumina data.

Use of Universal Primers The goal of this study and most environmental microbiology studies is to capture as much prokaryotic diversity as possible. Current protocols rely on the universality of primers, though primer independence would be ideal. The primers chosen for this study were the universal 16S rRNA primers 515f and 806r (Caporaso et al., 2011). These primers span the V4 hypervariable region, one of the most phylogenetically informative rRNA regions (Wang et al., 2007). A previous study by

White et al., (2012) found discrepancy in sponge microbial communities using the 27F and 338r primers which span the V2 hypervariable region, another highly informative region. Chloroflexi was found in low numbers in the sponge *Axinella corrugata* and retested using another primer set (White et al., 2012; Schmidt et al., 2012). Further studies could include other hypervariable regions to see how each compares, and also further show how diverse the samples are. Also, a set of archaea-specific primers would be interesting to use. Many of the archaeal OTUs found in this study were part of the rare biosphere of microbes but usage of an archaeal-specific prime may show that they are present in higher numbers. Ideally, an independence from gene-specific, universal primers would be optimal for deep sequencing studies, and as the cost of sequencing decreases, this is closer to a reality.

Amplicon Sequencing The choice to use amplicon sequencing was driven mainly by cost. Ideally, using a primer independent approach, like shotgun sequencing, would have been more ideal to assess not only taxonomic diversity, but also gene diversity and functional potential. One caveat is that some tools are not available for shotgun sequencing as there are for amplicon studies. This is not to say that amplicon studies are not important, but that they only cover a small part of the overall diversity within a sample. Further work should focus not only on the microbial diversity, but also on the functional potential of seawater metagenomes. The effective metagenome size for seawater is 5,000,000,000 base pairs (C. Titus Brown, STAMPS course, unpublished) which can be easily, and cheaply sequenced on the Illumina platform. Functional gene diversity can also be studied using metatranscriptomic sequencing, which will reveal the genes being expressed.

Illumina sequencing is the cheaper technology for metagenomic studies and has surpassed 454 pyrosequencing in the amount of data generated. Millions of base pairs can be generated at once, allowing for deeper genome coverage at a lower price.

South Florida Waterways and the Reef Tract Southeast Florida is at a convergence zone for tropical and temperate waters. The continental shelf is characterized by warm waters, which is a branch of the Gulf Stream. Climate in southeast Florida is characterized as a tropical savanna climate, having distinct wet and dry seasons (Banks et al., 2008). It is a densely populated area with about 5, 710,953 people between Martin, Palm Beach, Broward and Miami-Dade counties (2010 Census). About 57% of the population between these counties rely on centralized sewers, with the remaining 40% dependent on in-ground wastewater disposal which may not be treated (Futch et al., 2011). Wastewater is removed by a series of outfalls, two of which are located in Broward County. These outfalls drain directly into the ecosystem. There is also the influence of the coastal canals which can carry anthropogenic pollutants into the marine environment such as storm water and reckless waste dumping (Futch et al., 2011). These pollutants may carry microbes, some of which may cause human disease. To monitor human pathogens, there are programs in effect to monitor levels of indicator organisms, a term used in environmental health studies for microbes whose presence likely indicates contamination of human pathogens, namely fecal coliforms and enterobacteria. More information from- <http://2010.census.gov/2010census/>

The inlets directly connect the Intracoastal Waterway with the Atlantic Ocean. The inlet habitat and Intracoastal Waterway areas are highly subject to tides and their tidal plumes influence the coastal circulation in areas near the inlets. Depending on the time of year, the salinity of these plumes is different, being less saline in the wet season (June-September). The height of the tides can affect the inlet contribution to reefs (Banks et al., 2008). The coastal canals and inlets carry anthropogenic pollutants into the marine environment such as storm water and products of reckless waste dumping (Futch et al., 2011).

The distance of the Port Everglades Inlet to Hillsboro Inlet is approximately 18.4km (Figure S1). Drainage canals along the Intracoastal Waterway carry agricultural runoff and affect surface and groundwater flow, making this waterway a source of nonpoint pollution. Hillsboro Inlet is 94 meters wide and 3 meters deep. Port Everglades inlet is 137 meters wide and 12.8 meters deep (Carsey et al., in press). The maximum depth of the Port Everglades shipping channel is about 15 meters (Stamates et al., 2013).

The Hollywood and Broward outfalls are both located at 30 feet and are just east of the 3rd reef track. The outfall plume is fresh water at the pipe (Carsey et al., in press) and mixes quickly with the surrounding seawater, keeping the salinity at normal ocean levels. A slight elevation of nutrients is typically observed at the outfall. Plume dilution takes about two minutes to move from the outfall to the surface, with strong mixing occurring as the water moves upward from the outfall. Further mixing occurs with the Florida Current pushing on the effluent horizontally downstream. The effluent eventually rises to the surface about 10 meters away from the outfall pipe (Koopman et al., 2006). With treatment, about 94% of total suspended solids are removed meaning only 136 mg/L are expelled in the Hollywood outfall effluent. At the Broward outfall, 97% of total suspended solids are removed, meaning 217 mg/L are expelled. The only known spikes in nutrients occur at the surface of the outfall boil, but overall, these spikes did not significantly affect the surrounding waters (Carsey et al., in press).

The Florida reef tract extends from southern Martin County to the end of Miami-Dade County before entering the Florida Keys reef system, a distance of about 125 km (Banks et al., 2008). The reefs around south Florida run parallel to the shoreline and are dominated by massive stony corals such as *Montastrea* spp. These reefs form a reef complex composed of an inner, middle, and outer reef system (Banks et al., 2008).

Wastewater Treatment Wastewater treatments ensure that organic material and bacteria are not present when discharged into the ecosystem. Preliminary and primary treatment removes floating debris, suspended solids, and organic material from the wastewater, respectively. Secondary treatment is the removal of organic matter and suspended solids, usually followed by a disinfection step. Further treatment also removes nutrients (Tchobanglous et al., 2003). Secondarily treated wastewater is discharged by a series of outfalls, two of which are located within Broward County, the Broward Outfall and Hollywood Outfall (Futch et al., 2011). Following treatment, the treated water is disinfected using chlorine and followed by a dechlorination step (Tchobanglous et al., 2003). The Broward outfall is located 2.2km offshore at 32.6m of depth. The daily sewage discharge at this site is 36MGD on average. The Hollywood outfall is located 3.1 km offshore, and situated at 28.3 meters depth. It has an average daily discharge rate of 40 MGD of treated sewage water (Carsey et al., 2010). These outfalls are placed in

trenches cutting through reef system with the discharge point on the lower foreslope of the outer reef (Banks et al., 2008).

Florida Area Coastal Environment Project (FACE) The Florida Area Coastal Environment Program (FACE) is a water quality monitoring program under the National Ocean and Atmospheric Association (NOAA) which has collected and analyzed seawater samples around southeast Florida waters since 2004 (FACE website). The goals of the FACE program are to determine the effects of biological, chemical, and physical parameters on water quality. Its purpose is to assess water quality and provide a baseline for future observations and to determine the sources of nutrient and microbial contamination into near shore waters (FACE Website). FACE is a bimonthly sampling program for the waters around Broward County. It mainly focuses on quantitative PCR (qPCR) and culture based techniques to understand the microbial community structure of Broward County and determine the presence of human and land-based sources of pollution. More information can be found at- <http://www.aoml.noaa.gov/themes/CoastalRegional/projects/FACE/faceweb.htm>

Taxonomic Diversity Representative sequences were assigned to taxa using the RDP Bayesian Classifier incorporated in QIIME and CloVR. Dominant classes of bacteria included *Cyanobacteria*, *Alphaproteobacteria*, *Gammaproteobacteria*, and *Flavobacteria*. These classes were abundant across all samples. Individually, there were noticeable differences in rare classes. Broward samples from November and January shared high abundances of mixed *Euryarchaeota*, *Firmicutes*, and *Planctomycetes*. During the cooler months of April, November, and January, *Betaproteobacteria* increase.

Site Type The unweighted Principal Coordinate of Analysis (PCoA) graph shows how some microbial communities group according to site based on a specified parameter from a QIIME mapping file (Figure 12). Both inlet microbial communities appear distinct from the outfalls and reef communities, and surprisingly group together despite Hollywood and Broward inlets being approximately 20 km apart. They also cluster across all four quarterly samples. The inlets are surrounded and fed by land, and thus runoff from heavy rains can affect their microbial content (Brownwell et al 2007), which may explain the clustering of the inlets on the PCoA graph.

The inlets are also close to shore, with the Hollywood and Broward inlets directly adjacent to recreational beaches. The reefs and outfalls show some clustering with each other, but do not cluster with the inlets, suggesting differences in the composition and/or quantity of microbes between them. The western end of the Florida Current is known to meander and revere. This results in counterclockwise rotating fronts, causing upwelling of deep, nutrient-rich water onto the shelf. This also results in the formation of eddies, which can range from 5-30 km in size and last 1-2 days (Carsey et al., 2013). The currents carry treated wastewater to the outfalls and in turn, to the reefs. Land runoff and wastewater tend to be highly diluted stream of the outfalls (Koopman et al., 2008). The Broward North outfall, discharges at a depth of 107 feet and lies 7,300 ft (2,225.04 m) offshore (Koopman et al., 2008), and the Hollywood outfall lies 10,000 ft (3,048 m) offshore and discharges at 93 feet depth (Koopman et al., 2008).

This pattern indicates suggests that the microbial makeup of the inlets is distinct from the reefs and outfalls. One possible caveat to consider is that all samples were collected without regard to tidal stages, but their (Tables S1-S2) influence may be an important missing factor. Further studies need to be done of the tidal fluctuation within these sites to determine what is being flushed from the inlet sites into the open ocean and vice versa.

The skiff, unweighted Unifrac, and weighted Unifrac plots of site types, show a clear distinction between the inlets from the outfalls and reefs. The outfalls are areas where treated wastewater is expelled, carrying potential pathogens (Shibata et al., 2004). The inlets, particularly the Hollywood inlet, are subject to recreational and commercial boat traffic, which release ballast water and deposit oils and other pollutants into the water column. Human and animal activity (Wright et al., 2009) shed large amounts of fecal indicator (FIB) and skin bacteria (Abdelzaher et al., 2010) into the water column in the coastal areas near the beaches.

The way the OTUs cluster determines where they lie on a PCoA graph. A common clustering algorithm is uclust, which is a default clustering algorithm in QIIME. Uclust clusters sequences by comparing them to a centroid sequence. Every sequence in the cluster is similar to a centroid sequence (Edgar, 2010). The longest sequences are representative of each cluster. If there is similarity between the

representative sequence and the others, it is grouped into that cluster (Li et al., 2001;2002;Li and Godzik, 2006).

The significant differences shown in site type suggest that the inlet sites are different from the outfall and reef sites. The outfalls and reefs are subject to current action and are further from shore. The inlets have heavy boat traffic daily, and are coastal areas. The influx of terrestrial runoff and tidal action give the inlets a distinct microbial make up. Also, given that the inlet areas tend to be shallower and not influenced by currents like reefs and outfalls, it is possible that a small change in temperature, salinity, nutrients, etc., would have a more drastic effect on the microbial communities associated with the inlets.

Depth The depth samples were collected from water less than thirty feet. These shallow depths receive much light and mixing for any depth stratification. Only the Broward samples had the mid-depth sampled, which were composed of 4 samples. Previous work by Caro-Quintero and Konstantinidis (2012) found changes in the microbial structure below the photic zone. Since these samples were found well within the photic zone, this explains why there was no significant change with depth.

The outfall samples were collected at the boil. Water from the pipe must ascend about 60 meters to the surface, and if there is strong current action, the effluent plume will not be near the pipe. Sampling of the outfalls was determined based on the location of the surface boil.

No significant changes in microbial composition were found between depth. Although it was hypothesized that there would be a depth stratification of microbial communities, the data showed otherwise. Likely, there was no sediment resuspension, or the samples were collected to far above the benthos that resuspended materials were not captured. A future study could focus on the waters immediately above the benthos. The collection of more samples, might establish a depth stratification of bacteria. There is also the likelihood that samples from the outfalls were not collected directly from the outfall. Sampling location for the outfall was dependent on the location of the surface boil, and sampled directly below it for the mid-depth and bottom samples. Activity from currents can change the location of the boil, and angle of the wastewater plume. Further studies of

the wastewater outfalls should focus on collecting water directly from the plume as it emerges from the pipe to get a more accurate picture of what is in the water.

Season The moderate clustering of seasons (Figure 15) indicates that there is some overlap between seasons and perhaps only small effects of temperature. This is due to SE Florida having distinct wet (June-September) and dry seasons (Banks et al., 2008). This explains the lack of clustering, as it is not a matter season, but more a factor of rainfall. Rainfall carries soil and nutrients into the ocean, which may influence the clustering of the inlets. Total rainfall from June 2011-September 2011 was 19.34 inches. Rainfall during the months of October 2011-January 2012 was 17.48 inches (NOAA Climatic Data-www.ncdc.noaa.gov). Land runoff due to rainfall is more dilute near the outfalls and reefs. During the dry season, the lack of heavy rainfall and cooler temperatures can influence the microbial makeup (Brownwell et al., 2007). NOAA Climatic Data is listed in the Appendix.

A similar study by Gilbert et al. (2010) found that there was seasonal effects on water samples from the Western English Channel, with water surface temperatures fluctuating between 10-16°C between summer and winter. The difference between Gilbert et al. (2010) and this study is that the water temperature of the sampling sites were greater than 25°C but not less than 30°C throughout the year in southeast Florida. The overall range for the sampling period was 21.02°C-30.08°C (NOAA metadata). Since the water temperature in southeast Florida is warmer, microbial communities will look different than that of a temperate ocean.

There are significant changes in microbial composition across seasons. The inlets are coastal areas, and receive large inputs of terrestrial runoff due to heightened rainfall in the rainy season. The outfalls and reefs are subject to currents, which keeps the microbial communities mostly uniform. The influence of the inlets onto the outfalls and reefs would mainly be due to tidal action, but given the proximity to shore, this is not likely the case.

A previous project looking at the microbial makeup of southeast Florida seawater was the Florida Area Coastal Environment (FACE) Project. The main focus was to determine water quality using culture techniques for indicator bacteria and qPCR of these same indicator bacteria. This previous study created a base line for microbial data. To

further enhance this project, the next-generation sequencing techniques used to generate this data can be used not only to reaffirm the presence of these indicator bacteria, but also establish a baseline of the non-target microbes. Resampling over the course of 2-5 years can further establish true seasonal fluxes. Work completed for this project only spanned for one year, so to say that microbial spikes and drops are seasonally determined is premature.

Microbial composition

Abundant Taxa The abundant phyla in all samples are Cyanobacteria, Bacteroidetes, and Proteobacteria. Cyanobacteria occur frequently in the upper ocean layers, and certain species of Alphaproteobacteria, such as Rhodobacterales, also occur. In the taxa chart, many results are listed as “other”. This means that there are still microbes in the Greengenes and RDP databases that need to be characterized. Some of the taxa were listed as “other,” meaning that the database has no designation for particular taxa matching that sequence. Unknown taxa are not talked about due to their ambiguity. The abundant taxa are defined as being present at greater than or equal to 1% across all samples. It should also be kept in mind that there is inherent variability in the copy number of 16S rRNA molecules between bacterial genomes and sequence variability between closely related phyla and even within genomes (Větrovský and Baldrain, 2013), such as *Prochlorococcus* (Martiny et al., 2009). Data for the 16S ribosomal RNA copy numbers was obtained from the Ribosomal RNA Operon Copy Number Database (rrndb) by Michigan State University (Klappenbach et al., 2001) - <http://rrndb.umms.med.umich.edu/>

Prochlorococcus *Prochlorococcus* was first discovered by Dr. Sallie Chisolm at Woods Hole Oceanographic Institute in 1986 and published in Nature magazine in July 1988 (Chisolm et al., 1988). *Prochlorococcus* is found throughout the world’s oceans and is an abundant, open ocean cyanobacterium, as seen in the environmental shotgun-sequencing project done in the Sargasso Sea by the J. Craig Venter Institute in 2004 (Venter et al., 2004). These bacteria are small ($>0.6\ \mu\text{m}$). From this study, *Prochlorococcus* occurrence appeared highest in the warmer months of April and July, and decreased in the cooler months. This reflects the possibility of environmental

parameters shaping microbial communities (Furhman et al., 2006). *Prochlorococcus* has two ecotypes, one adapted to high light conditions, and one to lower light conditions. The ecotypes have a 3% difference in their ribosomal DNA (Rocap et al., 2003), but they exhibit different preferences for light, pigments, nitrogen usage, and cyanophage resistance. Ribosomal RNA analysis, however, generally lumps these ecotypes together. It should also be noted that there is a slight difference in the copy number of rRNA molecules in the *Prochlorococcus* ecotypes. *Prochlorococcus* belonging to the eMIT9313 clade tend to have two ribosomal RNA molecules, while the remaining clades have one (Martiny et al., 2009).

In the 2007 Global Ocean Sampling (GOS) Expedition (Rusch et al., 2007), they compared bacterial samples from the Pacific to the Atlantic. Both had similar ribotype composition, but there were genes more common to the Atlantic than to the Pacific. The majority of genes were for phosphate transport and utilization. Phosphate is important for microbial growth, especially for *Prochlorococcus* whose fitness is determined by the level of phosphorus in the water column. Phosphorus limitation is a common occurrence in the ocean and is a component for cyanobacterial blooms (Venter et al., 2004). *Prochlorococcus* is one of the most dominant cyanobacteria in the ocean, and higher than *Synechococcus* (Venter et al., 2004; Rusch et al., 2007). Similarities in the samples collected on the GOS Expedition showed that sample similarities had a correlation with changes in the environment, particularly with temperature and salinity (Rusch et al., 2007).

Synechococcus *Synechococcus* is adapted to oligotrophic environments and is globally distributed throughout tropical and temperate seas (Palenik et al., 2003). A unique feature of *Synechococcus* is that it can utilize organic nitrogen and phosphorus and has strategies to conserve iron by using nickel and cobalt in enzymatic reactions (Palenik et al., 2003). Compared to *Prochlorococcus*, *Synechococcus* is less abundant in oligotrophic environments, but has a larger, uniformly global distribution (Moore et al., 1995; Palenik et al., 2003), and tends to be found in shallower depths than *Prochlorococcus* (Moore et al., 1995). *Synechococcus* was not initially found to be of high abundance, but picking the OTUs with an open reference actually flipped the previous data to show *Prochlorococcus* as being the less dominant of the two. The 16S

ribosomal RNA is not well resolved between *Synechococcus* and *Prochlorococcus*, especially the high-light adapted *Prochlorococcus* (Rocap et al., 2002).

Pelagibacteraceae Another abundant microbe was Pelagibacteraceae, formerly known as SAR11. This group was among the first groups to be identified using cultivation-independent techniques (Rappe et al., 2002). These are a cosmopolitan species of bacteria, found throughout the world's oceans and have been found in every 16S clone library since their discovery. Pelagibacter contains one copy of the 16S rRNA molecule. They are thought to account for over 18% abundance in photic waters (Morris et al., 2002). Pelagibacter are a sister group to the order Rickettsiales, which are thought to be the precursor to the modern mitochondria (Thrash et al., 2011). Pelagibacter is not yet considered an official species, and is instead referred to as a candidate species. What is clear from the data is that it occurred in high numbers across all water samples. Pelagibacteraceae showed a correlation with chlorophyll a.

Candidatus Portiera The candidate genus *Candidatus Portiera* is an endosymbiont of the whitefly (Thao and Baumann, 2004; Santo-Garcia et al., 2012; Bing et al., 2013). *Candidatus Portiera* is a symbiotic bacterium. It is commonly associated with whiteflies, insects closely related to aphids and other psyllid flies. These insects suck sap out of plants and are a common agricultural pest. *Candidatus Portiera* is a symbiont for many psyllid insects and provides amino acids and carotenoids that the insect would not be able to obtain from their diet (Thao and Baumann, 2004; Santos-Garcia et al., 2012). Insects, various arthropods, and birds can carry bacteria into an aquatic ecosystem (McArthur, 2006). A possible reason for the occurrence of an insect commensal is due to how well updated the Greengenes database was at the time the OTU table was generated. It is likely that *Candidatus Portiera* is not the part of the oceanic microbiome, but due to the information available at the time, the database matched to the insect symbiont. The rRNA copy number for *Candidatus Portiera* is not yet known, but other members of the family Halomonadaceae are known to have between 4-5 copies of the 16S rRNA molecule (rrndb), which could explain why *Candidatus Portiera* is part of the top 1% of taxa found in the water column, despite being an insect symbiont. *Candidatus Portiera* showed a significant change with the date, time of day collected, salinity, temperature, phaeopigments, and total nitrogen. Little is known of this genus and

its occurrence in this dataset was likely the best match to the GREENGENES database at the time of analysis. At the beginning of 2014, QIIME underwent a large update and GREENGENES was updated. Subsequent analysis showed a reduction in the occurrence of *Candidatus Portiera*, but it was still present among the abundant taxa. According to Phil Hugenholtz, the Greengenes database will be updated by mid-year 2014 to account for the *Candidatus Portiera* hits (unpublished).

Rhodobacteraceae The family Rhodobacteraceae belong to the class Alphaproteobacteria (Lee et al., 2005). These bacteria comprise a large portion of the microbial life in the oceans. Rhodobacteraceae are rapid surface colonizers (Dang et al., 2008) and are purple, photosynthetic bacteria. Roseobacter is a well known genus of the Rhodobacteraceae. They are exclusively marine. Roseobacter synthesizes bacteriochlorophyll a. It does not utilize sunlight for growth, but for the production of ATP under aerobic conditions (Giovannoni and Rappé, 2000). One of the significant changes in Rhodobacteraceae was temperature. Stratil et al. (2013) found that OTU richness increased as a result of temperature on a brown macroalgal biofilm community dominated by alphaproteobacteria, namely those of the family Rhodobacteraceae. Rhodobacteraceae increased two-fold between temperatures of 5° and 25°C. Ribosomal RNA operon copy number varies greatly between 1-5 among the various genera in the family Rhodobacteraceae (rrndb). Rhodobacteraceae showed a significant relationship with salinity, temperature, oxygen saturation, pH, total suspended solids, and nitrate concentration.

Chloroplasts Stramenopiles (heterokonts) form one of eight major eukaryotic groups. Oceanic stramenopiles are higher than in coastal regions (Massana et al., 2004). Across all samples the chloroplasts coincide with Stramenopiles, and also other members, such as Haptophyceae. Chloroplasts are present in phytoplankton and land plants and are responsible for photosynthesis. Similarities occur between chloroplast and cyanobacteria, the only difference being that cyanobacteria are free-living. Chloroplasts are well within the cyanobacterial radiation (Giovannoni et al., 1988) and have their own 16S rRNA similar to prokaryotes (Zablen et al., 1975). They are also closely related to unicellular, nitrogen-fixing cyanobacteria (Falcon et al., 2010). Stramenopiles were found to be across all samples, its highest abundance occurring at the inlets. The percent abundance

did not fluctuate beyond 7% with the exception of the Broward July inlet sample where it rose to over 27%. This high abundance was not seen in any other inlet sample. Ribosomal RNA copy number varies between 2-4 for cyanobacteria and chloroplast (rrndb).

Flavobacteriaceae The family Flavobacteriaceae were found through all of the samples. Flavobacteriaceae is common in environmental samples. In a study by Campbell et al. (2011), it was found that Flavobacteriaceae were a group that was present greater than 50% of the time across their samples and sometimes cycled below 50%, but were still always present. Flavobacteria belong to a consortium of bacteria that make up the Cytophaga-Flavobacteria-Bacteroides (CFB) group. This group consists of some of the most common bacterial taxa in seawater. These organisms are important to the marine environment as they break down large molecules like chitin and DNA (Giovannoni and Rappé, 2000). Our regression analysis showed that phaeopigment concentrations had an effect on the family Flavobacteriaceae, though there was no significant correlation with total suspended solids or turbidity. Flavobacteria are often associated with higher primary production, namely in colder waters, and they are often the dominant class of Bacteroidetes in marine picoplankton communities (Gomez-Pereira et al., 2010). Flavobacteriaceae OTU abundance showed a significant correlation with phaeopigments, an algal degradation product, (Abdell and Bowman, 2005), likely because they break down algal degradation products (Gomez-Periera et al., 2010). The number of rRNA operon copies varies between 1-4 for Flavobacteriaceae (rrndb). The family *Flavobacteriaceae* ($R^2=0.4817$) showed a statistically significant relationship to time of day, salinity, and phaeopigments.

Cryomorphaceae Cryomorphaceae are a group within the Flavobacteriaceae (Alonso et al., 2007) which were found in high abundance at the inlets. These organisms are strict aerobes and cannot utilize carbohydrates as an energy source (Bowman et al., 2003; Lau et al., 2005). Instead, like other Flavobacteriaceae, they can hydrolyze complex molecules like casein and gelatin (Lee et al., 2010).

Thiotrichales Thiotrichales are a group of filamentous, sulfur oxidizing microbes. They show an increased growth response to high molecular weight dissolved organic material. Due to its sulfur oxidizing nature, it utilizes sulfur as an electron source. Notable genera from this group include *Beggiatoa*, a sulfurous mat forming microbe

which may indicate organic pollution (Chet and Michell, 1975; Fenchel and Bernard, 1995; Elliot et al., 2006), *Piscirickettsia*, which is a pathogen to salmonid fishes (Fryer et al., 1992; Fryer and Hedrick, 2003), and *Thiomargarita*, one of the largest bacteria isolated from marine sediments in Namibia (Schulz et al., 1999) and the Gulf of Mexico (Kalanetra et al., 2005). *Thiomargarita* is an anaerobic bacterium, but can tolerate atmospheric oxygen unlike its relative *Beggiatoa* (Schulz et al., 1999).

Methylophilales Methylophilales are a group of microbes typically associated with freshwater ecosystems, however, there is one strain, HTCC2181, which is present in marine datasets. A unique feature of this organism is that it has one of the smallest bacterial genomes, about 1.3 million base pairs. These organisms cannot oxidize methane. Instead, they were shown to utilize methanol and formaldehyde (Giovannoni et al., 2008) as energy sources. Methylophilales blooms have been known to occur along with phytoplankton blooms. It is suggested that during the phytoplankton bloom (Morris et al., 2006), the photosynthetic organisms release methanol and other methyl compounds like dimethylsulfoniopropionate, which the Methylophilales, and other methylotrophs utilize for energy (Neufeld et al., 2008).

Alteromonadales The order Alteromonadales are a particle attached bacteria that form biofilms on the particulate matter. A feature of these organisms is that they create antagonistic molecules (Long and Azam, 2001), molecules that prevent other bacteria from settling on or near them. An advantage to their biofilm growth pattern and antagonistic molecule production is that they are immune to the antagonistic molecules of other organisms (Long and Azam, 2001). Alteromonadales are also shown to increase in the presence of oil in mangrove sediments (dos Santos et al., 2011). Alteromonadales are also important in the uptake of dissolved organic nitrogen (Wawrik et al., 2012).

Thermoplasmata An archaeal class, Thermoplasmata, were found in fairly high abundance across all samples. After many rRNA studies, archaea were found in a variety of habitats (DeLong and Pace 2001). A recent study by Pires et al. (2012) found Thermoplasmata in mangrove sediments, indicating this class of Archaea is not bound to deep-sea hydrothermal vents. DeLong (2006) actually found that the class Euryarcheota, which Thermoplasmata belongs to, is abundant at the surface and declines with increasing depth. Archaea are as different from bacteria as they are from eukaryotes

(DeLong and Pace 2001), but share similar rRNA molecules to bacteria. The lack of highly abundant archaeal sequences is possibly due to their poor understanding. Archaea are notoriously hard to culture and were for the longest time thought to occur only in extreme conditions. Recent rRNA phylogeny studies have shown, however, that archaea can be found in a variety of habitats (DeLong and Pace 2001). The total rRNA copy number for the class Thermoplasmata is one amongst all known genera (rrndb).

Abundant taxa were also found for each site. The outfalls and reefs had similar abundant taxa, especially with *Prochlorococcus* and *Candidatus Portiera*, though the Broward reefs had abundant occurrences of Acidomicrobiales, of the family OCS155.

Acidomicrobiales Acidomicrobiales are heterotrophic organisms. These organisms produce secondary compounds to keep antagonistic organisms away. The strain OSC155 is common in this dataset, but it is not well understood. It was recently shown by Needham et al. (2013), that this particular clade fluctuates greatly with time, an increase of about six-fold over the course of 10 days before reducing in equal proportions over 4 days.

Rank Abundance

The Broward samples, BR7CApr and BR10BJuly, with the lower relative abundance and higher ranks were also the samples with the lowest OTU counts. They did, however, have better evenness. This could be due to a small sampling depth or the samples could have an even and highly abundant microbial community. Some of the samples with a steeper slope, BR10BApr, BR10CApr, and BR14AApr, likely have an abundance of singletons, sequences that only occur once indicating one of a particular organism. The samples with the good diversity and highly ranked were BR7CJuly, BR7CNov, BR7CJan, BR10AJuly, Br10CNov and BR10CJan. These samples are also likely the most evenly distributed of samples since their abundance curves are not as steep.

HW14AJan and Apr, HW9CApr, and HW4CApr had 19980, 13990, 15196, and 11353 OTUs respectively. These were the highest counts of OTUs in each sample for the Hollywood dataset. These samples also had the steeper relative abundance slopes, which likely indicates a high number of singletons, which results in many taxa with a lower

rank. The samples HW4ANov and HW14ANov had 3427 and 4331 OTUs, respectively. This likely indicates that these samples are made up mostly of singletons and pyrosequencing noise.

While it is possible to have one organism in a sample, there is also the possibility of sequencing error, which is common with pyrosequencing. A way to work around this is to denoise a dataset and run chimera checks. The number of sequences per sample will drop, but the dataset will have less errors.

Pathogenic Taxa

Among the samples, several pathogenic bacterial sequences were found. These pathogens were made up of three classes: *Epsilonproteobacteria*, *Firmicutes*, and *Gammaproteobacteria*. Within the class Gammaproteobacteria were the family Enterobacteraceae and the genus, *Vibrio*. Members of these classes encompass taxa that usually cause severe infections and some opportunistic infections in both humans and oceanic species. Detection of pathogenic species is important because not only do they pose a threat to human health, but also DNA from dead bacteria can be incorporated into the genome of other bacteria, conferring mutations such as antibiotic resistance (McDaniel et al., 2010).

Epsilonproteobacteria Within the *Epsilonproteobacteria* are the genera *Campylobacter* and *Arcobacter*. Members of both genera cause diarrheal disease in humans, mainly in very young, very old, and immunocompromised individuals. Species of *Campylobacter* and *Arcobacter* have also been detected as part of the microbial consortia present in coral black band disease (Frias-Lopez et al.2002). *Campylobacter jejuni* is one of the leading causes of bacterial gastroenteritis (Altekrus et al., 1999) and subsequently, is also linked to the development of Guillan-Barre syndrome (Alos, 1997; Altekruse et al., 1999). Occurrences of Campylobacteraceae were below 0.001% across all samples. It should be noted, however, that *Arcobacter* was found in slightly higher numbers. In Broward it was found in the outfall surface site in April 2011 at 0.008%, 0.002% at the outfall surface in November 2011, at 0.002% in the outfall surface in January 2012, and at the mid-depth at 0.001% in April 2011. At the Broward inlet sites, it was found at 0.001% in April, 0.001% in November, and 0.001% in January. In

Hollywood, *Arcobacter* was found at the outfall surface in April 2011 at 0.005% and in July at 0.004%. There was also a hit for *Arcobacter* at the Hollywood inlet site at 0.001%. These numbers should be no cause for alarm as *Arcobacter* has non-pathogenic strains that are responsible for nitrogen fixation.

Firmicutes Within the Firmicutes, the class that stands out is Clostridia. Notable members of the Clostridia class are *Clostridium perfringens*, *C. difficile*, *C. tetani*, and *C. botulinum*. *Clostridium* species have been found within the microbial consortia of coral black band disease (Frias-Lopez et al. 2002) and are also responsible for human diseases such as colitis (Larson et al., 1978), tetanus, and botulism. The occurrence of *Clostridia* in seawater likely came from terrestrial input, either from farms or soil (Saif and Brazier, 1996). Clostridia are spore-forming bacteria, so they can survive many harsh environments, and may actually survive the water treatment process (Romano et al., 2012). *Clostridium* was found in low levels (less than 0.001%) across all samples.

Enterobacteraceae Within the family Enterobacteraceae, the only pathogen occurrence was *Serratia marcescens*. *Serratia marcescens* rarely causes infection in humans, with the exception of severe UTIs in immunocompromised individuals, and catheter infections. It is typically a nosocomial, or hospital-acquired, infection that, when it does cause disease, causes a variety of problems including wound infections, respiratory tract infections, and meningitis. It is also a cause of endocarditis and will contaminate contact lenses and cases (Parment 1997; Shanks et al., 2007). It is an enteric bacterium and also a causative agent of white pox disease in Caribbean acroporid corals (Patterson et al., 2002). *Serratia* was not a direct hit in the OTU table. It was lumped under the family Enterobacteriaceae and searched for using the NCBI database. The family Enterobacteriaceae was found at less than 0.001-0.0000% across most sites, except for the Broward reef in January 2011, which was at 0.0015%, the Broward outfall surface sample in July 2011 at 0.0006%. For the Hollywood sites, Enterobacteraceae was found at the surface outfall sample in April 2011 at 0.0005% and at the inlet sample in January 2012 at 0.0002%. Given the small percentage of Enterobacteraceae present, this should be little cause for concern.

Vibrio Members of the genera, *Vibrio*, are of great concern for human health. Many cause serious wound infections and gastrointestinal illness. *Vibrio* species are of

great concern for those with compromised livers and diabetes. The most pathogenic members include *V. parahaemolyticus* and *V. vulnificus*, both common in seawater and shellfish. Another species of concern is *V. cholerae*, which is found attached to copepods along with *V. parahaemolyticus* (Huq et al., 1983). Other *Vibrio* species have been speculated to cause coral disease, mainly *Vibrio shiloi*, a bacterium thought to be spread by bearded fireworms as they fed on coral (Kushmaro et al., 2001; Sussman et al., 2003). Other species of *Vibrio* have also been implicated in incidences of coral disease in both the tropical Atlantic and tropical Pacific (Cervino et al., 2004). It should be noted, though that incidences of *Vibrio vulnificus* have occurred on the west coast of Florida in the early fall of 2013. This species is commonly found in estuarine to marine waters (Farmer, 1991). Members of the family Vibrionaceae, which *Vibrio spp.* belongs to were found at 1% at the Broward reef sites in the month of April 2011 and January 2012. There was also a 1.000% incidence at the Broward inlet site in November 2011. The remaining sites were had *Vibrio* hits at less than 0.001% and some were at 0.000%. *Vibrio spp.* are common to marine waters and their incidence is to be expected.

The outfalls are treated with chlorine so there should be no live pathogens, but the DNA may still be intact, which possibly explains their presence. Since these bacteria are in marine waters, the high salinity shrinks them, making it difficult to culture them and most are not targeted by the qPCR used by the FACE project. Further work on understanding the presence of these microbes in seawater would have a strong impact on human and ecosystem health. FACE currently targets indicator bacteria, but should also target known pathogens such as *Campylobacter spp.*, *Arcobacter spp.*, *Vibrio parahaemolyticus*, *V. vulnificus*, etc., as these organisms can have a detrimental effect on human health. It should be noted, however, that the risk of illness from an outfall pipe is low since the wastewater is treated and the effluent mixes quickly with the surrounding seawater quickly (Koopman et al., 2006).

Notable Rare Microbes

Betaproteobacteria Previous clone and culturing experiments showed that betaproteobacteria were not present in the open ocean, but in freshwater ecosystems. Culture-independent study, however, has found an occurrence of betaproteobacteria in

open-ocean waters. A notable microbe found was the genus *Zoogloea*, which is often associated with sludge and wastewater (Giovannoni and Rappé, 2000). Low levels of *Zoogloea* were found at the wastewater outfalls. No hits for *Zoogloea* were found at the inlet and reef sites. *Zoogloea* are common components to activated sludge in wastewater treatment plants.

Planctomycetes According to the CloVR skiff results, Planctomyces were found in elevated levels at the reefs. Planctomyces have been isolated from various aquatic ecosystems (Fuerst et al., 1997) and seem to have a correlation with algal blooms, likely subsisting on the algal degradation products (Pizzetti et al., 2011). Interestingly, chlorophyll a and phaeopigment, a pigment of algal degradation, were not higher at reefs, making this result suspicious. Also, planctomyces are the one of the only bacterial cells lacking a peptidoglycan cell wall, and have a membrane similar to that of the eukaryotic “nuclear membrane” (Pizzetti et al., 2011). Planctomycetes have been found previously at cold water coral *Lophelia* reefs (Neulinger et al., 2008; Kellog et al., 2009).

SourceTracker

Understanding the impacts of microbial communities from different sites on one another is crucial to this study. One of the main objectives was to determine if one site was affecting the other. SourceTracker works by comparing the communities present in a sample and seeing how similar they are to each other. From there, conclusions can be drawn as to whether a site is having some sort of effect on the other. Using SourceTracker, we were able to see the effects of a particular site on another. According to the SourceTracker results, the inlets had a moderate similarity to the outfalls, though they cluster separately in the Unifrac graphs. The Hollywood inlet had the same communities as both the Broward and Hollywood reef site. This is strange given their proximity to each other.

Using the outfalls as a source, the Broward reef showed strong similarities to the Broward outfall and the Hollywood reef showed strong similarities to the Hollywood outfall. A similar profile is seen when compared to the Unifrac results, where the reefs and outfalls cluster together.

Significance

Currently, qPCR and culture-based data collection make up the majority of microbiological studies. The majority of microbes are not culturable, and qPCR work is dependent on already known sequences to generate primers. Deep sequencing data, collected from 454 pyrosequencing, can amplify some of the rarer species, which may serve as indicators of environmental change, as they will likely be affected first (Kirchman 2008).

Metagenomics projects, to date, have mostly focused on the human body, phytoplankton, and soil communities. Work by Gilbert (2009; 2010), Delong (2006; 2009) and efforts from the Global Ocean Sampling Expedition have paved the way for further work on microbial community structure using metagenomics analysis. These efforts have expanded into the Human Microbiome Project and the Earth Microbiome Project each aimed at assessing microbial communities in different biomes within an environment. An understanding of the microbial world will greatly benefit monitoring projects by showing the composition of microbes in the environment, how they interact, and the impacts humans have on the environment around them. More information can be found at <https://commonfund.nih.gov/hmp/>

Knowing the microbial composition and how it relates to environmental and anthropogenic factors is an important aspect for environmental and public health studies. Seasonal and location changes in species composition may provide a baseline for water quality studies and baseline data for environmental studies, and in particular, how microbes interact with the marine environment during healthy and disease states. Data collected from this project may likely serve as a preliminary project for further endeavors, such as the Earth Microbiome Project.

Summary and Conclusions

Microbial life is essential to the health and dynamics of an ecosystem. Long-term monitoring studies not only reflect the health of an ecosystem, but also how human impacts, storm events, and natural phenomenon shape community structure. A longer sampling time is needed to fully assess how microbial community structure is influenced by depth, season, and location. 454 pyrosequencing is currently being phased out in favor

of Illumina, which can generate millions of high-quality reads (Illumina.com) for a lower cost. More samples can be multiplexed on an Illumina run making long-term environmental monitoring studies more cost effective. Furthermore, long term monitoring of tidal impacts on community composition is also vital, as this was not touched on as much for this study.

Current projects utilize quantitative PCR (qPCR) and culturing, but this is limited to known microbes. By utilizing next generation sequencing platforms, water quality studies can pick up on viable but no culturable microbes and genes that may have survived the treatment process. Next generation sequencing platforms can also assess the transcriptomic diversity of water column microbes and determine the presence of expressed genes in the water column, as amplicon sequencing does not distinguish between alive or dead microbes. Further long term monitoring should include samples taken pre- and/or post treatment from the wastewater treatment plant to more accurately portray what the outfalls are discharging. Future studies could include a metatranscriptome study to round out some of the amplicon data. It would detail the expressed gene distribution in SE Florida waters and show changes in gene expression. Some genetic responses could indicate pollution, such as oil or gas from illegal dumping, and some environmental changes such as pH, temperature, salinity, et cetera.

Related projects should emphasize the microbial communities of the sediments of the outfalls, inlets, and reefs. Sediments do not have the diluting factor water has, so nutrients and microbes will likely accumulate in high numbers (Austin, 1988). The only time this would be disturbed is during a storm event when the sediments would be resuspended or when tidal or current action is at its highest (Austin, 1988;Hartz et al., 2008).

Another focus could be the microbial diversity of the organisms associated with the outfalls, reefs, and inlets. A goal of this study was to establish a baseline of the microbial life in southeast Florida waters. Marine organisms play a vital role in the health of an ecosystem and are directly affected by changes in water quality, particularly the filter and suspension feeding organisms such as sponges, corals, and shellfish, especially if they are of commercial interest. Like sediments, these organisms tend to concentrate

bacteria and nutrients and will show a far greater amount of diversity than the surrounding seawater as there is no diluting effect in these organisms.

Finally, a focus on the eukaryotic microorganism structure is important but beyond the scope of this project. Microbial eukaryotic diversity, mainly protists, in the oceans is also a burgeoning topic, as most are also difficult to maintain in pure culture. Microbial eukaryotes were for the longest time ignored and known as “small flagellates” (Massana and Pedrós-Alió, 2008). Next-generation sequencing technologies and molecular techniques have made it possible for microbial eukaryotes to be understood as vital components to the oceanic ecosystem (Massana and Pedrós-Alió, 2008). These organisms were not touched on for this study, mostly due to the cost of sequencing, but also that no standards were set forth by the Earth Microbiome Project for eukaryotic data at the time. Since June 2013, the Earth Microbiome has set up standards for 18S rRNA amplicon sequencing. Such advances would establish the microbial eukaryote diversity of southeast Florida waters. A study including the eukaryotes is vital as many microbial eukaryotes have implications on public health.

The NGS technology used for this study has revealed the most common seawater microbes of southeast Florida’s wastewater outfalls, reefs, and inlets. The microbes identified in this bypassed standard culturing techniques and showed the changes and similarities in community structure over the course of a year. It revealed the presence of common seawater microbes, such as *Prochlorococcus*, the families Rhodobacteraceae and Flavobacteraceae, and the candidate genus, *Portiera*.

Overall, this project revealed an environmental baseline for further sequencing studies, and shows the microbial communities present at different sites and seasons. Further, it establishes next-generation sequencing technology as a vital tool for environmental monitoring in addition to the culturing and PCR techniques utilized by the NOAA FACE program and similar entities.

References

- Abdell, G.C. and Bowman, J.P. (2005) Ecological and biogeographic relationships of class Flavobacteria in the Southern Ocean. *FEMS Microbial Ecology*. 51 (2) 265-277
- Abdelzaher, A.M., Wright, M.E., Ortega, C., Solo-Gabriele, H.M., Miller, G., Elmir, S., Newman, X., Shih, P., Bonilla, J.A., Bonilla, T.D., Palmer, C.J., Scott, T., Lukasik, J., Harwood, V.J., McQuaig, S., Sinigalliano, C., Gidley, M., Plano, L.R.W., Zhu, X., Wang, J.D., and Fleming, L.E. (2010) Presence of Pathogens and Indicator Microbes at a Non-Point Source Subtropical Recreational Marine Beach. *Appl. Environ. Microbiol.* February 1, 2010 76:724-732; published ahead of print 4 December 2009, doi:10.1128/AEM.02127-09
- Al Saif, N., and Brazier, J.S. (1996) The distribution of *Clostridium difficile* in the environment of South Wales. *Journal of Medical Microbiology*, 45(2), 133-137.
- Allos, B.M. (1997) Association between *Campylobacter* infection and Guillain-Barré syndrome. *J Infect Dis.* 176:S125-8
- Alonso, C., Warnecke, F., Amann, R., and Pernthaler, J., (2007) High local and global diversity of *Flavobacteria* in marine plankton. *Environmental Microbiology*. 9 (5) 1253-1266 doi:10.1111/j.1462-2920.2007.01244.x
- Altekruse, S.F., Stern, N.J., Field, P.I., and Swerdlow, D.L. (1999) *Campylobacter jejuni*-An Emerging Foodborne Pathogen. *Emerging Infectious Diseases*. 5(1) 28-35.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol* 215(3):403-410
- Andersson, A.F., Lindberg, M., Jakobsson, H., Bäckhed, F., Nyrén, P., et al., (2008) Comparative Analysis of Human Gut Microbiota by Barcoded Pyrosequencing. *PLoS ONE* 3(7): e2836. doi:10.1371/journal.pone.0002836
- Angiuoli, S.V., Matalka, M., Gussman, G., Galens, K., Vangala, M., Riley, D.R., Arze, C., White, J.R., White, O., and Fricke, W.F. (2011) CloVR: A virtual machine for automated and portable sequence analysis from the desktop using cloud computing. *BMC Bioinformatics*. 12:356.
- Austin, B. (1988) *Marine microbiology*. Cambridge University Press. Vol. 1: 1, 4-6, 31-35, 45.
- Azam, F., Fenchel, T., Field, J.G., Gray, J.S., Meyer-Reil, L.A., and Thingstad, F. (1983) The Ecological Role of Water-Column Microbes in the Sea. *Mar. Ecol. Prog. Ser.* (10) 257-263.

Banks, K.E., Riegl, B.M., Shinn E.A., Piller, W.E., and Dodge, R.E. (2007) Geomorphology of Southeast Florida continental reef tract (Miami-Dade, Broward, and Palm Beach Counties, USA). *Coral Reefs*. 26:617-623. doi:10.1007/500338-007-0231-0

Banks, K.E., Riegl, B.M., Richards, V.P., Walker, B.E., Helmle, K.P., Jordan, L.K.B., Phipps, J., Shivji, M., Spieler, R.E., and Dodge, R.E. (2008) Chapter 5: The reef tract of continental Southeast Florida (Miami-Dade, Broward, and Palm Beach Counties, USA). In: Riegl, B., Dodge, R.E. (eds) *Coral Reefs of the USA*. Springer-Verlag, Dordrecht, 175-220.

Béjà, O., Aravind, L., Koonin, E.V., Suzuki, M.T., Hadd, A., Nguyen, L.P., *et al.* (2000) Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* **289**: 1902–1906.

Bing, X.L., Yang, J., Fein, E.Z., Wang, X.W., and Liu, S.S. (2013) Characterization of a Newly Discovered Symbiont of the Whitefly *Bemisia tabaci* (Hemiptera: Aleyrodidae)). *Applied and Environmental Microbiology*. 79 (2) 569-575. doi:10.1128/AEM.03030-12

Bowman, J. P., Nichols, C. M., & Gibson, J. A. (2003). *Algoriphagus ratkowskyi* gen. nov., sp. nov., *Brumimicrobium glaciale* gen. nov., sp. nov., *Cryomorpha ignava* gen. nov., sp. nov. and *Crocinitomix catalasitica* gen. nov., sp. nov., novel flavobacteria isolated from various polar habitats. *International Journal of Systematic and Evolutionary Microbiology*, 53(5), 1343-1355.

Bowman, J.S., Rasmussen, S., Blom, N., Deming, J.W., Rysgaard, S., and Sicheritz-Ponten, T. (2012) Microbial community structure of Arctic multiyear sea ice and surface seawater by 454 sequencing of the 16S RNA gene. *International Society for Microbial Ecology*. 6: 11-20. doi:10.1038/ismej.2011.76

Brownwell, M.J., Harwood, V.J., Kurz, R.C., McQuaig, S.M., Lukasik, J., and Scott, T.M. (2007) Confirmation of putative stormwater impact on water quality at a Florida beach by microbial source tracking methods and structure of indicator organism populations. *Water Research*. (41) 3747-3757.

Campbell, B.J., Yu, L., Heidelberg, J.F., and Kirchman, D.L. (2011) Activity of abundant and rare bacteria in a coastal ocean. *Proceedings of the National Academy of Science*. 108:12776–12781.

Caporaso, J.G., Bittinger, K., Bushman, F.D., DeSantis, T.Z., Andersen, G.L., Knight, R. (2010) PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* 26:266-267.

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

Caro-Quintero, A., and Konstantinidis, K.T. (2012) Bacterial species may exist, metagenomics reveal. *Environmental Microbiology*. 14 (2) 347-355.

Carsey, T., Casanova, H., Drayer, C., Featherstone, C., Fischer, C., Goodwin, K., Proni, J., Saied, A., Sinigalliano, C., Stamates, J., Swart, P., and Zhang, J.Z. (2010) NOAA OAR AOML-38. FACE Outfalls Survey Cruise: October 6-19, 2006. NOAA Technical Report.

Carsey, T et al.,(in press) NOAA Technical Report, OAR AOML: BROWARD COUNTY COASTAL OCEAN: WATER QUALITY CRUISES 2010-2012. AOML

Cassler, M., Winegar, R. , McCarthy, P.J., Pomponi, S.A., Peterson, C.L, Wright, A.E, Lopez, J.V. (2008) Use of Real-Time qPCR to quantify members of the unculturable heterotrophic bacterial community in a deep sea marine sponge, *Vetulina* sp. *Microbial Ecology* 55:384-394.

Cervino, J.M., Hayes, R.L., Polson, S.W., Polson, S.C., Goreau, T.J., Martinez, R.J., and Smith, G.W. (2004) Relationship of *Vibrio* species Infection and Elevated Temperatures to Yellow Blotch/Band Disease in Caribbean Corals. *Appl. Environ. Microbiol.* 70(11): 6855-6864 doi: 10.1128/AEM.70.11.6855-6864.2004

Chet, I., and Mitchell, R., (1975) Bacterial attack of corals in polluted seawater. *Microbial Ecology*. 2 (3) 227-233

Chisholm, S.W., Olson, R.J., Zettler, E.R., Waterbury, J., Goericke, R. and Welschmeyer, N. (1988). A novel free-living prochlorophyte occurs at high cell concentrations in the oceanic euphotic zone. *Nature* 334 (6180): 340–343. doi:10.1038/334340a0

Crespo, B.G., Pommier, T., Fernández-Gómez, B., and Pedrós-Alió, C. (2013) Taxonomic composition of the particle-attached and free-living bacterial assemblages in the Northwest Mediterranean Sea analyzed by pyrosequencing of the 16S rRNA. *MicrobiologyOpen*. 2(4) 541-552. doi: 10.1002/mbo3.92

Daley, R.J., and Hobbie, J.E. (1975) Direct Counts of Aquatic Bacteria by a Modified Epifluorescence Technique. *American Society of Limnology and Oceanography*. 20 (5) 875-882.

Dang, H., Li, T., Chen, M., and Huang, G. (2008) Cross-Ocean Distribution of *Rhodobacterales* Bacteria as Primary Surface Colonizers in Temperate Coastal Marine Waters. *Applied and Environmental Microbiology*. 79 (1) 52-60. doi:10.1128/AEM.01400-07

de Wit, R., and Bouvier, T. (2006) ‘Everything is everywhere, but the environment selects’: what did Baas Becking and Beijerinck really say? *Environmental Microbiology*. 8_ 755-758.

- DeLong, E.F. and Pace, N.R. (2001) Environmental Diversity of Bacteria and Archaea. *Syst. Biol.* 50 (4) 470-480.
- DeLong, E.F. (2005) Microbial community genomics in the ocean. *Nature.* (3) 459-469.
- DeLong, E.F. and Karl, D.M. (2005) Genomic perspectives in microbial oceanography. *Nature.* (437) 336-342.
- DeLong, E. F. (2006) Archaeal mysteries of the deep revealed. *Proc Natl Acad Sci U S A.* 103:6417-6418.
- DeLong, E.F. (2009) The microbial ocean from genomes to biomes. *Nature.* (459) 200-206.
- DeSantis, T. Z., P. Hugenholtz, N. Larsen, M. Rojas, E. L. Brodie, K. Keller, T. Huber, D. Dalevi, P. Hu, and G. L. Andersen. (2006) Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. *Appl Environ Microbiol* 72:5069-72.
- dos Santos, H.F., Cury, J.C., do Carmo, F.L., dos Santos, A.L., Tiedje, J., van Elsas, Rosado, A.S., and Peixoto, R.S. (2011) Mangrove Bacterial Diversity and the Impact of Oil Contamination Revealed by Pyrosequencing: Bacterial Proxies for Oil Pollution. *PLoS ONE* 6(3): e16943. doi:10.1371/journal.pone.0016943
- Edgar, R.C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26(19):2460-2461.
- Elliot, J.K., Spear, E., and Wylie-Echeverria, S. (2006) Mats of *Beggiatoa* bacteria reveal that organic pollution from lumber mills inhibits the growth of *Zostera marina*. *Marine Ecology.* (27) 372-380 doi:10.1111/j.1439-0485.2006.0100.x
- Elmir, S.M., Wright, M.E., Abdelzahar, A., Solo-Gabriele, H.M., Fleming, L.E., Miller, G., Rybolowik, M., Shih, M.T.P., Pillai, S.D., Cooper, J.A., and Quaye, E.A. (2007) Quantitative evaluation of bacteria released by bather in a marine water. *Water Research.* 41(1):3-10. Epub 2006 Nov 17.
- Falcón, L. I., Magallón, S., and Castillo, A. (2010). Dating the cyanobacterial ancestor of the chloroplast. *The ISME journal*, 4(6), 777-783.
- Farmer, III, J. J., and Hickman-Brenner, F. W. (2006). The genera *Vibrio* and *Photobacterium*. In *The Prokaryotes* (pp. 508-563). Springer New York.
- Fenchel, T., and Bernard, C. (1995) Mats of colourless sulphur bacteria. I. Major marine processes. *Marine Ecology Progress Series.* 178: 161-170.
- Fera, M.T., Maugeri, T.L., Gugliandolo, C., Beninati, C., Giannone, M., La Camera, E., and Carbone, M. (2004) Detection of *Arcobacter* spp. in the Coastal Environment of the

Mediterranean Sea. *Applied and Environmental Microbiology*. 70 (3) 1271-1276
doi:10.1128/AEM.70.3.1271-1276.2004

Finkl, C.W. and Charlier, R.H. (2003) Sustainability of Subtropical Coastal Zones in Southeastern Florida: Challenges for Urbanized Coastal Environments Threatened by Development, Population, Water Supply and Storm Hazards. *Journal of Coastal Research*. 19 (4) 934-943

Florida Inland Navigation District (2000), *Long Range Dredged Material Management Program for the Atlantic Intracoastal Waterway in Florida*, retrieved 23 February 2012 from <http://www.aicw.org/pdfs/dmmp.pdf>.

Frias-Lopez, J., Zerkle, A.L., Bonheyo, G.T., and Fouke, B.W. (2002) Partitioning of Bacterial Communities between Seawater and Healthy, Black Band Diseased, and Dead Coral Surfaces. *Applied and Environmental Microbiology*. 68 (5) 2214- 2228
doi:10.1128/AEM.68.5.2214-2228.200

Fryer, J.L., Lannon, C.N., Giovannoni, S.J., and Wood, N.D. (1992) *Piscirickettsia salmonis* gen. nov., sp. nov., the Causative Agent of and Epizootic Disease in Salmonid Fishes. *International Journal of Systematic Bacteriology*. 42 (1) 120-126

Fryer, J.L., and Hedrick, R.P. (2003) *Piscirickettsia salmonis*: a Gram-negative intracellular bacterial pathogen of fish. *Journal of Fish Diseases*. 26:251-262.

Fuerst, J.A., Gwilliam, H.A., Lindsay, M., Lichanska, A., Belcher, C., Vickers, J.E., and Hugenholtz, P. (1997) Isolation and Molecular Identification of Planctomycete Bacteria from Post Larvae of the Giant Tiger Prawn. *Applied and Environmental Microbiology*. 63 (1) 254-262.

Fuhrman, J.A., Hewson, I., Schwalbach, M.S., Steele, J.A., Brown, M.V., and Naeem, S. (2006) Annually reoccurring bacterial communities are predictable from ocean conditions. *Proceedings of the National Academy of Sciences*. 103 (35) 13104-13109.
doi:10.1073/pnas.0602399103

Fuhrman, J.A. (2009) Microbial community structure and its functional implications. *Nature*. (459) 193-199.

Futch, J.C., Griffin, D.W., Banks, K., and Lipp, E.R. (2011) Evaluation of sewage source and fate on southeast Florida coastal reefs. *Marine Pollution Bulletin*. (62) 2308-2316.

Gilbert, J.A., Field, D., Swift, P., Newbold, L., Oliver, A., Smyth, T., Somerfield, P.J., Huse, S. and Joint, I. (2009) The seasonal structure of microbial communities in the Western English Channel. *Environmental Microbiology*. 11 (12) 3132-3139.
doi:10.1111/j.1462-2920.2009.02017.x

Gilbert, J.A., Field, D., Swift, P., Thomas, S., Cummings, D., 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(11): e15545.

Gilbert, J.A. and DuPont, C.L. (2011) Microbial Metagenomics: Beyond the Genome. Annu. Rev. Marine Sci. (3) 347-371.

Giovannoni, S.J., Turner, S., Olsen, G.J., Barns, S., Lane, D.J., and Pace, N.R. (1988) Evolutionary Relationships among Cyanobacteria and Green Chloroplasts. Journal of Bacteriology. 170 (8) 3584-3592.

Giovannoni, S.J., and Rappé, M. (2000) Chapter 3: Evolution, Diversity, and Molecular Ecology of Marine Prokaryotes. In Kirchman D.L. (ed.) Microbial Ecology of the Oceans. Wiley-Liss, Inc. pgs. 47-84.

Giovannoni, S.J., Hayakawa, D.H., Tripp, H.J., Stingl, U., Givan, S.A., Cho, J., Oh, H., Kitner, J.B., Vergin, K.L., and Rappé, M. (2008) The small genome of an abundant coastal ocean methylotroph. Environmental Microbiology. 10 (7) 1771-1782
doi:10.1111/j.1462-2920.2008.01598.x

Gómez-Pereira, P.R., Fuchs, B.M., Alonso, C., Oliver, M.J., van Beusekom, J.E., and Amann, R. (2010) Distinct flavobacterial communities in contrasting water masses of the north Atlantic Ocean. ISME Journal. 4(4):472-87. doi: 10.1038/ismej.2009.142

Haas, B.J., Gevers, D., Earl, A.M., Feldgarden, M., Ward, D.V., Giannoukos, G., et al. (2011) Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. Genome Research 21:494-504.

Hamady, M., Lozupone, C., and Knight, R. (2010) Fast Unifrac: facilitating high-throughput phylogenetic analyses of microbial communities including analysis of pyrosequencing and PhyloChip data. International Society for Microbial Ecology (4) 17-27.

Handelsman, J. (2004) Metagenomics: Application of Genomics to Uncultured Microorganisms. Microbiology and Molecular Biology Reviews. 4 (68) 669-685.

Harris, J. K., Kelley, S. T., and Pace, N. R. (2004) New perspective on uncultured bacterial phylogenetic division OP11. Applied and Environmental Microbiology, 70(2), 845-849.

Hartz, A., Cuvelier, M., Nowosielski, K., Bonilla, T.D., Green, M., Esiobu, N., McCorquodale, D.S., and Rogerson, A. (2008) Survival Potential for *Escherichia coli* and Enterococci in Subtropical Beach Sand: Implications for Water Quality Managers. J. Environ. Qual. 37:898-905. doi:10.2134/jeq2007.0312

Hezaji, A., and Falkiner, F.R. (1997) *Serratia marcescens*. Journal of Medical Microbiology. 46: 903-912.

Huber, J.A., Welch, D.B.M., Morrison, H.G., Huse, S.M., Neal, P.R., Butterfield, D.A., and Sogin, M.L. (2007). Microbial Population Structures in the Deep Marine Biosphere. *Science*, 318:97-100. doi: 10.1126/science.11466689

Huq, A., Small, E.B., West, P.A., Huq, M.I., Rahman, R., and Colwell, R.R. (1983) Ecological relationships between *Vibrio cholerae* and planktonic crustacean copepods. *Appl. Environ. Microbiol.* 45:275–283.

Jeraldo, P., Chia, N. and Goldenfeld, N. (2011), On the suitability of short reads of 16S rRNA for phylogeny-based analyses in environmental surveys. *Environmental Microbiology*, 13: 3000–3009. doi: 10.1111/j.1462-2920.2011.02577.x

Kalanetra, K.M., Joye, S.B., Sunseri, N.R., and Nelson, D. (2005) Novel vacuolated sulfu bacteria from the Gulf of Mexico reproduce by reductive division in three dimensions. *Environmental Microbiology*. 7 (9) 1451-1460 doi:10.1111/j.1462-2920.2005.00832.x

Kirchman, D.L., Cottrell, M.T., and Lovejoy, C. (2010) The structure of bacterial communities in the western Arctic Ocean as revealed by pyrosequencing of 16S rRNA genes. *Environmental Microbiology*. 12 (5) 1132-1143.

Knights, D., Kuczynski, J., Charlson, E.S., Zaneveld, J., Mozer, M.C., Collman, R.G., Bushman, F.D., Knight, R., and Kelley, S.T. (2011) Bayesian community-wide culture-independent microbial source tracking. *Nature Methods*. 8 (9) 761-763. doi:10.1039/nmeth.1650

Koopman, B., Heaney, J.P., Cakir, F.Y., Rembold, M., Indeglia, P., and Kini, G. (2006) Ocean Outfall Study Final Report. Florida Department of Environmental Protection.

Kushmaro, A., Banin, E., Loya, Y., Stackebrandt, E., and Rosenberg, E. (2001) *Vibrio shiloi* sp. nov., the causative agent of bleaching of the coral *Oculina patagonica*. *Int. J. Syst. Evol. Microbiol.* 51:1383–1388.

Lapointe, B.E., and Bedford, B.J. (2010) Ecology and nutrition of invasive *Caulerpa bractyus* f. *parvifolia* blooms on coral reefs off southeast Florida, U.S.A. *Harmful Algae* 9, 1-12.

Lau, K. W., Ng, C. Y., Ren, J., Lau, S. C., Qian, P. Y., Wong, P. K., and Wu, M. (2005). *Owenweeksia hongkongensis* gen. nov., sp. nov., a novel marine bacterium of the phylum ‘Bacteroidetes’. *International journal of Systematic and Evolutionary Microbiology*, 55(3), 1051-1057.

Lazarevic, V., Whiteson, K., Huse, S., Hernandez, D., Farinelli, L., Osteras, M., Schrenzel, J., and Francois, P. (2009) Metagenomic study of the oral microbiota by Illumina high-throughput sequencing. *Journal of Microbiological Methods*. (79) 266-271.

Lee, D. H., Choi, E. K., Moon, S. R., Ahn, S., Lee, Y. S., Jung, J. S., and Kahng, H. Y. (2010). *Wandonia haliotis* gen. nov., sp. nov., a marine bacterium of the family Cryomorphaceae, phylum Bacteroidetes. *International Journal of Systematic and Evolutionary Microbiology*, 60(3), 510-514.

Lee, K., Liu, C., Anzai, Y., Kim, H., Aono, T., and Oyaizu, H. (2005) The hierarchal system of the '*Alphaproteobacteria*': description of *Hyphomonadaceae* fam. nov. *Xanthobacteraceae* fam. nov. and *Erythrobacteraceae* fam. nov. *International Journal of Systematic and Evolutionary Microbiology*. (55) 1907-1919. doi: 10.1099/ijs.0.63663-0

Long, R.A., and Azam, F., (2001) Antagonistic Interactions among Marine Pelagic Bacteria. *Applied and Environmental Microbiology*. 67 (11) 497-4983
doi:10.1128/AEM.67.11.4975-4983.2001

Lozupone, C. and Knight, R. (2005). "UniFrac: A New Phylogenetic Method for Comparing Microbial Communities". *Applied and Environmental Microbiology* 71 (12): 8228–8235. doi:[10.1128/AEM.71.12.8228-8235.2005](https://doi.org/10.1128/AEM.71.12.8228-8235.2005).

Lozupone, C., Hamady, M., Kelley, S.T., and Knight, R. (2007) Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. *Appl. Environ. Microbiol.* 73 (5) 1576-1585.
doi:10.1128/AEM.01996-06

Mardis, E.R. (2011) A decade's perspective on DNA sequencing technology. *Nature*. 470 (7333):198-203.

Massana, R., Castresana, J., Balagué, V., Guillou, L., Romari, K., Groisillier, A., Valentin, K., and Pedrós-Alió, C. (2004) Phylogenetic and ecological analysis of novel marine stramenopiles. *Applied and Environmental Microbiology* 70 (6) 3528-3534.

Massana, R., and Pedrós-Alió, C. (2008) Unveiling new microbial eukaryotes in the surface ocean. *Current Opinion in Microbiology*. 11:213–218.
doi:10.1016/j.mib.2008.04.004

McArthur, J.V. (2006) *Microbial Ecology: An Evolutionary Approach*. Elsevier Academic Press. 3-4; 41-45, 139, 149-151, 156-160, 171-172, 175, 179, 255.

McDaniel, L.D., Young, E., Delaney, J., Ruhnau, F., Ritchie, K. B., and Paul, J.H. (2010) High Frequency of Horizontal Gene Transfer in the Oceans. *Science*. 330: 50. doi 10.1126/science.1192243

McDonald, D., Price, M.N., Goodrich, J., Nawrocki, E.P., DeSantis, T.Z., Probst, A., Andersen, G.L., Knight, R., Hugenholtz, P. (2012) An improved Greengenes taxonomy

with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. ISME J 6(3): 610–618.

Moore, L.R., Goericke, R., and Chisholm, S.W. (1995) Comparative physiology of *Synechococcus* and *Prochlorococcus*: Influence of light and temperature for growth, pigments, fluorescence, and absorptive properties. Marine Ecology Progress Series. 116:259-275

Morris, R. M., Rappé, 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 (6917), 806-810.

Morris, R.M., Longnecker, K., and Giovannoni, S.J. (2006) *Pirellula* and OM43 are among dominant lineages identified in an Oregon coast diatom bloom. Environmental Microbiology. 8 (8) 1361-1370 doi:10.1111/1462-2920.2006.01029.x

Munro, P.M., Gauthier, M.J., Breittmayer, V.A., and Bongiovanni, J. (1989) Influence of Osmoregulation Processes on Starvation Survival of *Escherichia coli* in Seawater. Applied and Environmental Microbiology. 55 (8) 2017-2024.

Murray, A.E., Preston, C.M., Massana, R., Taylor, L.T., Blakis, A., Wu, K., and DeLong E.F. (1998) Seasonal and Spatial Variability of Bacterial and Archaeal Assemblages in the Coastal Waters near Anvers Island, Antarctica. Applied and Environmental Microbiology. 7 (64) 2585-2595.

Needham, D.M., Chow, C.T., Cram, J.A., Sachdeva, R., Parada, A., and Furhman, J. (2013) Short-term observations of marine bacterial and viral communities: patterns, connections and resilience. ISMEJ. (7) 1274-1285

Neefs, J.M., Van de Peer, Y., Rijk, P.D., Chapelle, S., and Wachter, RD. (1993) Compilation of small ribosomal subunit RNA structures. Nucleic Acids Research. (21) 3025-3049.

Neufield, J.D., Bodon, R., Moussard, H., Schäfer, H., and Murrell, J.C., (2008) Substrate-Specific Clades of Active Marine Methylophs Associated with a Phytoplankton Bloom in a Temperate Coastal Environment. Applied and Environmental Microbiology. 74 (23) 7321-7328 doi:10.1128/AEM.01266-08

Oliver, J.D. (2005) The Viable but Nonculturable State in Bacteria. The Journal of Microbiology. 43 (S) 93-100.

Olsen, G.J., and Woese, C.R. (1993) Ribosomal RNA: the key to phylogeny. FASEB. (7) 113-123.

Palenik, B., Brahmisa, B., Larimer, F.W., Land M., Hauser, L., Chain, D., Lamerin, J., Regala, W., Allen, E.E., McCarren, J., Paulsen, I., Dufresne, A., Partensky, F., Webb,

E.A., and Waterbury, J. (2003) The genome of a motile marine *Synechococcus*. *Nature*. 424:1037-1042

Parment, P.A. (1997) The role of *Serratia marcescens* in soft contact lens associated ocular infections. *Acta Ophthalmologica Scandinavica*, 75: 67–71. doi: 10.1111/j.1600-0420.1997.tb00253.x

Patterson, K.L., Porter, J.W., Ritchie, K.B., Polson, S.W., Peters, E.C., Santavy, D.L., and Smith, G.W. (2002) The etiology of white pox, a lethal disease of the Caribbean elkhorn coral, *Acropora palmata*. *Proceedings of the National Academy of Sciences* (99) 8725-8730.

Pedrós-Alió, C. (2012) The Rare Bacterial Biosphere. *Annu. Rev. Marine Sci.* (4) 449-466.

Peel, M.C., Finlayson, B.L., McMahon, T.A. (2007). Updated world map of the Köppen-Geiger climate classification. *Hydrol Earth Syst Sci* **11**: 1633–1644.

Pei, A., Nossa, C.W., Chokshi, P., Blaser, M.J., Yang, L., et al.,(2009) Diversity of 23S rRNA Genes within Individual Prokaryotic Genomes. *PLoS ONE* 4(5): e5437. doi:10.1371/journal.pone.0005437

Petrosino, J.F., Highlander, S., Luna, R.A., Gibbs, R.A., and Versalovic, J. (2009) Metagenomic Pyrosequencing and Microbial Identification. *Clinical Chemistry*. 55:5 856-866.

Pires, A.C.C., Cleary, D.F.R., Almeida, A., Cunha, A., Dealtry, S., Mendonça-Hagler, L.C.S., Smalla, K., and Gomes, N.C.M. (2012) Denaturing gradient gel electrophoresis and barcoded pyrosequencing reveal unprecedented archaeal diversity in mangrove sediment and rhizosphere samples. *Applied and Environmental Microbiology* 78 (16) 5520-5528.

Pizzetti, I., Fuchs, B.M., Gerdt, G., Wichels, A., Wiltshire, K.H., and Amann, R., (2011) Temporal variability of coastal Planctomyces Clades at Kabeltonne Station, North Sea. *Applied and Environmental Microbiology*. 77: 5009-5017. doi: 10.1128/AEM.0293170

Porter, K.G., and Feig, Y.S. (1980) The Use of DAPI for Identifying and Counting Aquatic Microflora. *American Society for Limnology and Oceanography*. 25 (5) 943-948.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2010) FastTree 2-Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS One* 5(3).

Puglise, K.A. and Kelty, R. (eds.). (2007) NOAA Coral Reef Ecosystem Research Plan for Fiscal Years 2007 to 2011. Silver Spring, MD: NOAA Coral Reef Conservation Program. NOAA Technical Memorandum CRCP 1.

- Pusceddu, A., Fiordelmondo, C., and Danovaro, R. (2005) Sediment Resuspension Effects on the Benthic Microbial Loop in Experimental Microcosms. *Microbial Ecology*. 50 (4) 602-613.
- Quince, C., Lanzén, A., Curtis, T.P., Davenport, R.J., Hall, N., Head, I.M., Read, L.F., and Sloan, W.T. (2009) Accurate determination of microbial diversity from 454 pyrosequencing data. *Nature Methods*, 6(9):639-641. doi:10.1038/NMETH.1361
- Quince, C., Lanzen, A., Davenport, R.J., and Turnbaugh, P.J. (2011) Removing Noise from Pyrosequencing Amplicons. *BMC Bioinformatic*. 12 (38).
- Rappé, M.S., and Giovannoni, S.J. (2003) The Uncultured Microbial Majority. *Annual Reviews of Microbiology*. (57) 369-394.
- Reeder, J., and Knight, R. (2010) Rapid denoising of pyrosequencing amplicon data: exploiting the rank abundance distribution. *Nature Methods*. 7 (9) 668-669. Doi:10.1038.nmeth0910-668b.
- Rocap, G., Distel, D.L., Waterbury, J., and Chisholm, S.W. (2002) Resolution of *Prochlorococcus* and *Synechococcus* Ecotypes by Using 16S-23S ribosomal DNA Internal Transcribed Spacer Sequences. *Applied and Environmental Microbiology*. 68 (30) 1180-1191. doi:10.1128/AEM.68.3.1180-1191.2002
- Rocap, G., Larimer, F.W., Lamerdin, J., Malfatti, S., Chain, P., Ahlgren, N.A., Arellano, A., Coleman, M., Hauser, L., Hess, W.R., Johnson, Z.I., Land, M., Lindell, D., Post, A.F., Regala, W., Shah, M., Shaw, S.L., Steglich, C., Sullivan, M.B., Ting, C.S., Tolonen, A., Webb, E.A., Zinser, E.R., and Chisholm, S.W. (2003) Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature*. 424:1042-1047.
- Romano, V., Pasquale, V., Krovacek, K., Mauri, F., Demarta, A., and Dumontet, S. (2012) Toxigenic *Clostridium difficile* PCR ribotypes from wastewater treatment plants in Southern Switzerland. *Applied and Environmental Microbiology*, 78(18), 6643-6646.
- Ronaghi, M. (2001) Pyrosequencing sheds light on DNA sequencing. *Genome Res* 11:3.11.
- Rusch, D.B., Halpern, A.L., Sutton, G., Heidelberg, K.B., Williamson, S. et al., (2007) The *Sorcerer II* Global Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific. *PLoS Biol* 5(3): e77. doi:10.1371/journal.pbio.0050077.
- Santos-Garcia, D., Farnier, P., Beitia, F., Zchori-Fein, E., Vavre, F., Mouton, L., Moya, A., Latorre, A., and Silva, F.J. (2012) Complete genome sequence of “*Candidatus Portiera aleyrodidarum*” BT-QVLC, an obligate symbiont that supplies amino acids and carotenoids to *Bemisia tabaci*. *Journal of Bacteriology* (194)23: 6654-6655.

- Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., Sahl, J.W., Stres, B., Thallinger, G.G., Van Horn, D.J., and Weber, C.F. (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75:7537–7541.
- Seposki, J.J. (1988) Alpha-, Beta, or Gamma: Where does all diversity go? *Paleobiology*. 14 (3) 221-234.
- Shanks, R.M.Q., Stella, N.A., Kalivoda, E.J., Doe, M.R., O'Dee, D.M., Lathrop, K.L., Guo, F.L., and Nau, G.J. (2007) A *Serratia marcescens* OxyR Homolog Mediates Surface Attachment and Biofilm Formation. *Journal of Bacteriology* (189) 7262-7272.
- Shibata, T., Solo-Gabriele, H.M., Fleming, L.E., and Elmir, S. (2004) Monitoring marine recreational water quality using multiple microbial indicators in an urban tropical environment. *Water Research*. (38) 3119-3131.
- Sklar, F., C. McVoy, R. VanZee, D. E. Gawlik, K. Tarboton, D. Rudnick, S. Maiao, and T. Armentano, *The Effects of Altered Hydrology on the ecology of the Everglades*, p 39-82, in J. Porter and K. Porter (eds.), *The Everglades, Florida Bay, and Coral Reefs of the Florida Keys, An Ecosystem Sourcebook*, CRC Press, New York.
- Sogin, M.L., Morrison, H.G., Huber, J.A., Welch, D.M., Huse, S.M., Neal, P.R., Arrieta, J.M., and Herndl, G.J. (2006) Microbial diversity in the deep sea and the unexplored “rare biosphere.” *Proc. Natl. Acad. Sci.* 103(32) 12115-12120.
- Staley, J.T. and Konopka, A. (1985) Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial environments. *Annu. Rev. Microbiol.* 39:321-340.
- Stamates, S. J., Bishop, J.R., Carsey, T.P., Craynock, J.F., Jankulak, M.L., Lauter, C.A., and Shoemaker, M.M. (2013) Port Everglades flow measurement system. NOAA Technical Report, OAR-AOML-42, 1-22.
- Stratil, S.B., Neulinger, S.C., Krecht, H., Friedrichs, A.K., and Wahl, M. (2013) Temperature driven shifts in the epibiotic bacterial community composition of the brown macroalgae, *Fucus vesiculosus*. *MicrobiologyOpen*. 2 (2) 338-349. doi: 10.1002/mbo3.79
- Sussman, M., Loya, Y., Fine, M., and Rosenberg, E. (2003), The marine fireworm *Hermodice carunculata* is a winter reservoir and spring-summer vector for the coral-bleaching pathogen *Vibrio shiloi*. *Environmental Microbiology*, 5: 250–255. doi: 10.1046/j.1462-2920.2003.00424.x
- Tchobanoglous, G., Burton, F.L., and Stensel, H.D. (2003) Wastewater engineering: Treatment and reuse. MANAGEMENT. 4th Ed. New York: McGraw-Hill Book Co. pg. 1-24

Temperton, B., and Giovannoni, S.J. (2012) Metagenomics: microbial diversity through a scratched lens. *Current Opinion in Microbiology* **15**:605–612.

Thao, M.L., and Baumann, P. (2004) Evolutionary Relationships of Primary Prokaryotic Endosymbionts of Whiteflies and Their Hosts. *Applied and Environmental Microbiology*. 70 (6) 3401-3406. doi: 10.1128/AEM.70.6.3401–3406.2004

The Human Microbiome Consortium (2012a) A framework for human microbiome research. *Nature*. E-pub ahead of print. doi:10.1038/nature11209

The Human Microbiome Consortium (2012b) Structure, Function and Diversity of the Human Microbiome in an Adult Reference Population. *Nature*. E-pub ahead of print. doi:10.1038/nature11234.

Thrash, J. C., Boyd, A., Huggett, M. J., Grote, J., Carini, P., Yoder, R. J., ... and Giovannoni, S. J. (2011). Phylogenomic evidence for a common ancestor of mitochondria and the SAR11 clade. *Scientific reports*, 1.

Vazquez-Baeza, Y., Pirrung, M., Gonzalez, A., Knight, R. (2013) Emperor: A tool for visualizing high-throughput microbial community data. *GigaScience* 2(1):16.

Venter, J.G., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., Wiu, D., Paulsen, I., Nelson, K.E., Nelson, W., Fouts, D.E., Levy, S., Knap, A.H., Lomas, M.W., Nealson, K., White, O., Peterson, J., Hoffman, J., Parsons, R., Baden-Tillson, H., Pfannkoch, C., Rogers, Y.H., and Smith, H.O. (2004) Environmental Genome Shotgun Sequencing of the Sargasso Sea. *Science*. 304: 66-74.

Wang, Q., Garrity, G.M., Tiedje, J.M., and Cole, J.R. (2007) Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. (73) 5261-5267.

Wawrik, B., Boling, W.B., Van Nostrand, J.D., Xie, J., Zhou, J., and Bronk, D.A. (2012) Assimilatory nitrate utilization by bacteria on the West Florida Shelf as determined by stable isotope probing and functional microarray analysis. *FEMS Microbiological Ecology*. 79 (2) 400-411 doi:10.1111/j.1574-6941.2011.01226.x

Werner, J.J., Koren, O., Hugenholtz, P., DeSantis, T.Z., Walters, W.A., Caporaso, J.G., Angenent, L.T., Knight, R., and Ley, R.E. (2012) Impact of training sets on classification of high-throughput bacterial 16S rRNA gene surveys. *ISME J* 6:94-103.

Whitaker, J.H. (1972) Evolution and Measurement of Species Diversity. *Taxon*. 21 (2/3) 213-251

White, J.R., Arze, C., Matalka, M., CloVR Team, White, O., Angiuoli, S.V., and Fricke, W.F. (2011) CloVR-16S: Phylogenetic microbial community composition analysis based

on 16S ribosomal RNA amplicon sequencing-standard operating procedure, version 2.0. Nature Proceedings. Version 2.0

White, J.R., Nagarajan, N., and Pop, M. (2009) Statistical Methods for Detecting Differentially Abundant Features in Clinical Metagenomic Samples. PLoS Computational Biology 5(4): e1000352. doi:10.1371/journal.pcbi.1000352

White, J.R., Patel, J., Ottesen, A., Arce, G., Blackwelder, P., and Lopez, J.V. (2012) Pyrosequencing of bacterial symbionts within *Axinella corrugata* sponges: diversity and seasonal variability. PLoS One 7:e38204

Whitman, W.B., Coleman, D.C., and Wiebe, W.J. (1998) Prokaryotes: The unseen majority. Proceedings of the National Academy of Science. 95: 6578-6583.

Woese, C.R., and Fox, G.E. (1977) Phylogenetic structure of the prokaryotic domain the primary kingdoms. Proc. Natl. Acad. Sci. USA. 74 (11) 5088-5090.

Woese, C.R. (1987) Bacterial evolution. Microbiological Reviews. 51 (2) 221-271.

Wright, M.E., Solo-Gabriele, H.M., Elmir, S., and Fleming.... (2009) Microbial load from animal feces at a recreational beach. Marine Pollution Bulletin. (58) 1649-1656.

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(2):e56207. doi:10.1371/journal.pone.0056207

Zablen, L.B., Kissil, M.S., Woese, C.R., and Buetow, D.E. (1975) Phylogenetic Origin of the Chloroplast and Prokaryotic Nature of Its Ribosomal RNA. Proceedings of the National Academy of Science. 72 (6) 2418-2422.

Zinger, L., Amaral-Zettler, L.A., Fuhrman, J.A., Horner-Devine, M.C., Huse, S.M., et al.,(2011) Global Patterns of Bacterial Beta-Diversity in Seafloor and Seawater Ecosystems. PLoS ONE 6(9): e24570. doi:10.1371/journal.pone.0024570

APPENDICES

APPENDIX 1-Table of Sample IDs, collection dates, DNA yields and site description.

Sample ID	Collection Date	ng/uL	Description
BR7C_Apr	04/13/11	61.68	Bottom, reef track
BR10A_Apr	04/13/11	88.49	Surface, outfall
BR10B_Apr	04/13/11	37.02	Mid-depth, outfall
BR10C_Apr	04/13/11	47.05	Bottom, outfall
BR14A_Apr	04/13/11	67.28	Surface, inlet
BR7C_July	07/13/11	35.20	Bottom, reef track
BR10A_July	07/13/11	45.70	Surface, outfall
BR10B_July	07/13/11	41.90	Mid-depth, outfall
BR10C_July	07/13/11	38.20	Bottom, outfall
BR14A_July	07/13/11	63.2	Surface, inlet
BR7C_Nov	11/17/12	25.26	Bottom, reef track
BR10A_Nov	11/17/12	30.00	Surface, outfall
BR10B_Nov	11/17/12	27.81	Mid-depth, outfall
BR10C_Nov	11/17/12	30.4	Bottom, outfall
BR14A_Nov	11/17/12	69.3	Surface, inlet
BR7C_Jan	01/30/12	36.83	Bottom, reef track
BR10A_Jan	01/30/12	42.05	Surface, outfall
BR10B_Jan	01/30/12	34.87	Mid-depth, outfall
BR10C_Jan	01/30/12	31.38	Bottom, outfall
BR14A_Jan	01/30/12	56.13	Surface, inlet
HW4A_Apr	04/19/11	88.54	Surface, outfall
HW4C_Apr	04/19/11	43.00	Bottom, outfall
HW9A_Apr	04/19/11	45.98	Surface, reef
HW9C_Apr	04/19/11	69.42	Bottom, reef
HW14A_Apr	04/19/11	69.16	Surface, inlet

HW4A_July	07/06/11	76.3	Surface, outfall
HW4C_July	07/06/11	29.1	Bottom, outfall
HW9A_July	07/06/11	30.6	Surface, reef
HW9C_July	07/06/11	27.8	Bottom, reef
HW14A_July	07/06/11	53.5	Surface, inlet
HW4A_Nov	11/30/11	50.35	Surface, outfall
HW4C_Nov	11/30/11	44.89	Bottom, outfall
HW9A_Nov	11/30/11	49.76	Surface, reef
HW9C_Nov	11/30/11	45.14	Bottom, reef
HW14A_Nov	11/30/11	68.05	Surface, inlet
HW9A_Jan	01/12/12	36.83	Surface, reef
HW9C_Jan	01/12/12	38.84	Bottom, reef
HW14A_Jan	01/12/12	48.16	Surface, inlet

APPENDIX 2-OTU Table Summary pre-chimera checking

Num samples: 38

Num observations: 2869

Total count: 235592

Table density (fraction of non-zero values): 0.188

Table md5 (unzipped): 3b2e87d6e7955588e9e938d7593b33e1

Counts/sample summary:

Min: 327.0

Max: 20020.0

Median: 3998.500

Mean: 6199.789

Std. dev.: 5154.124

Sample Metadata Categories: None provided

Observation Metadata Categories: taxonomy

Counts/sample detail:

BR10BJuly: 327.0

BR7CApr: 388.0

BR7CJan: 643.0

BR14ANov: 742.0

BR10AJuly: 1583.0

HW14AJul: 2243.0

BR7CNov: 2514.0

HW9ANov: 2598.0

HW4CJuly: 2679.0

BR14AJul: 2901.0

HW9CJuly: 3000.0

BR10AJan: 3130.0

HW4AJuly: 3230.0

BR10CJan: 3319.0

BR14AJan: 3376.0

HW9AJan: 3556.0

HW9AJuly: 3745.0

HW4ANov: 3747.0

BR10BNov: 3962.0

HW9CJan: 4035.0

HW4CNov: 4054.0

BR10ANov: 4161.0

BR7CJuly: 4613.0

HW14ANov: 4643.0

BR10CNov: 4704.0

BR10BJan: 5642.0

BR10CJuly: 6253.0

BR14AApr: 9572.0
HW4AApr: 10132.0
BR10AApr: 10278.0
HW4CApr: 12051.0
BR10CApr: 12685.0
HW9CNov: 13491.0
HW14AApr: 14647.0
HW9AApr: 14742.0
BR10BApr: 16084.0
HW9CApr: 16102.0
HW14AJan: 20020.0

APPENDIX 3-OTU table summary post chimera removal

Num samples: 38
Num observations: 4447
Total count: 227499
Table density (fraction of non-zero values): 0.109

Counts/sample summary:

Min: 321.0
Max: 19988.0
Median: 3803.500
Mean: 5986.816
Std. dev.: 4994.995
Sample Metadata Categories: None provided
Observation Metadata Categories: taxonomy

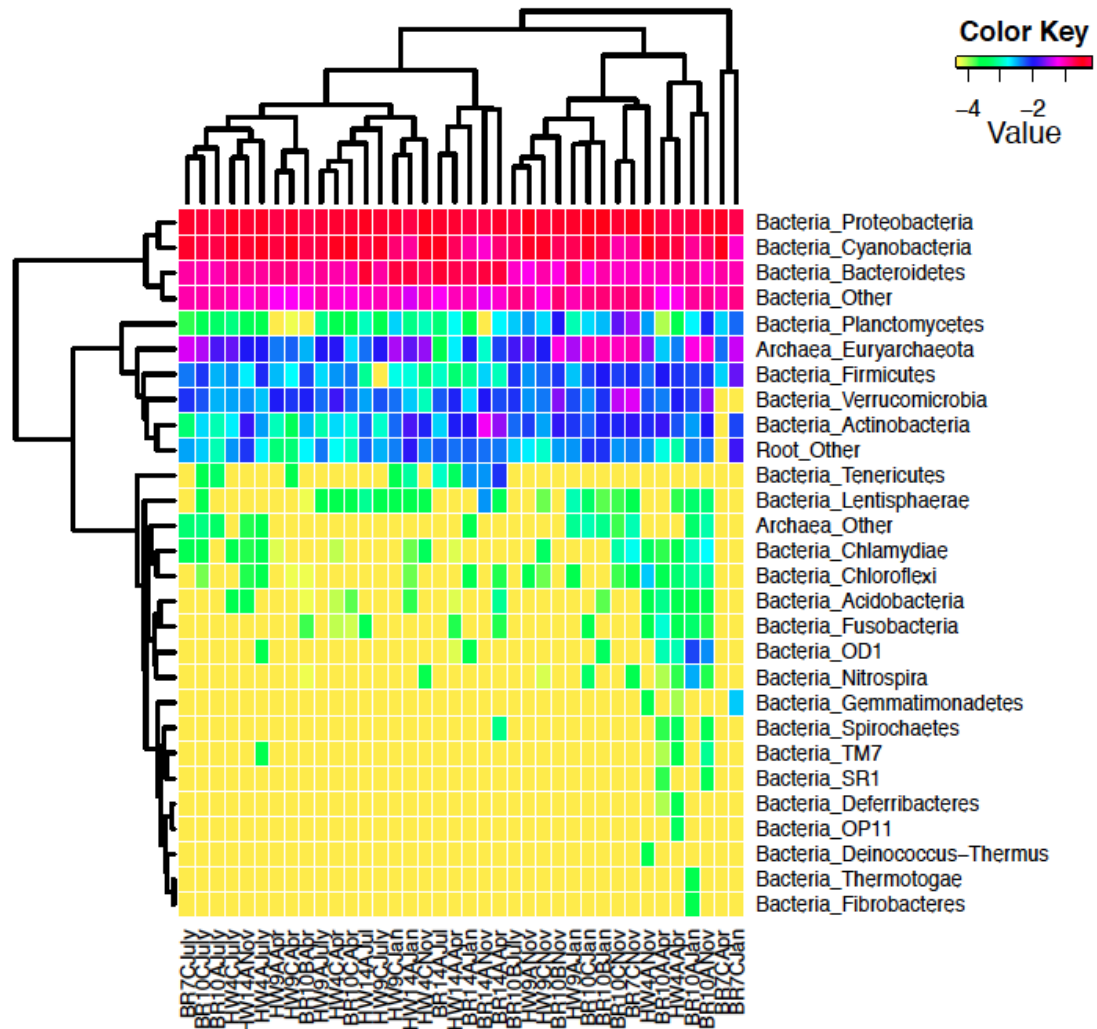
Counts/sample detail:

BR10BJuly: 321.0

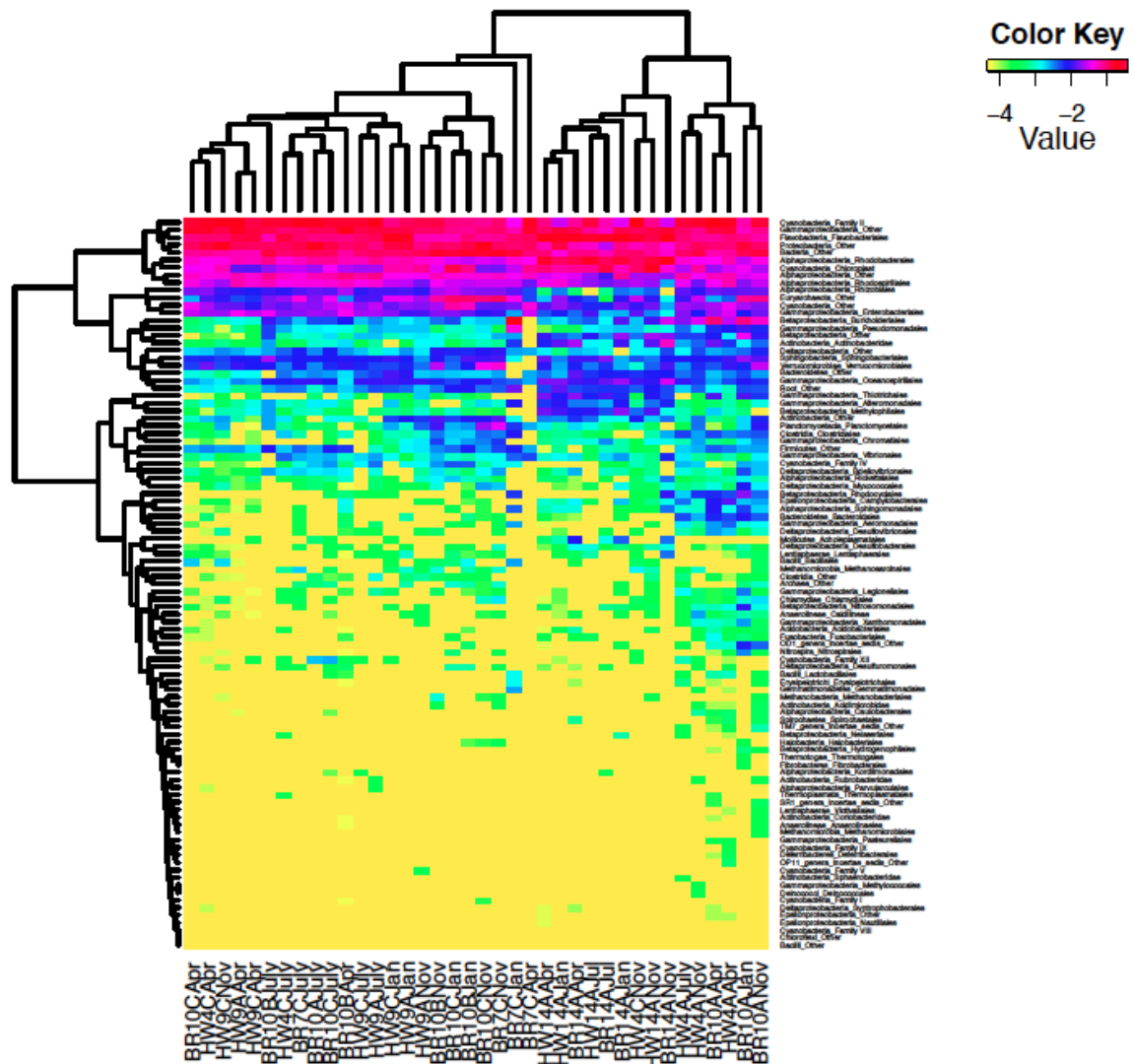
BR7CApr: 369.0
BR7CJan: 646.0
BR14ANov: 723.0
BR10AJuly: 1564.0
HW14AJul: 2114.0
HW9ANov: 2407.0
HW4CJuly: 2493.0
BR7CNov: 2517.0
HW9CJuly: 2752.0
BR14AJul: 2788.0
HW4AJuly: 3033.0
BR10AJan: 3166.0
BR14AJan: 3241.0
BR10CJan: 3330.0
HW9AJan: 3347.0
HW4ANov: 3448.0
HW9AJuly: 3465.0
HW4CNov: 3785.0
HW9CJan: 3822.0
BR10BNov: 3961.0
BR10ANov: 4199.0
HW14ANov: 4379.0
BR7CJuly: 4526.0
BR10CNov: 4707.0
BR10BJan: 5633.0
BR10CJuly: 6241.0
BR14AApr: 9553.0
BR10AApr: 9786.0
HW4AApr: 9837.0
HW4CApr: 11438.0
BR10CApr: 12001.0

HW9CNov: 12844.0
HW9AApr: 14021.0
HW14AApr: 14097.0
HW9CApr: 15357.0
BR10BApr: 15600.0
HW14AJan: 19988.0

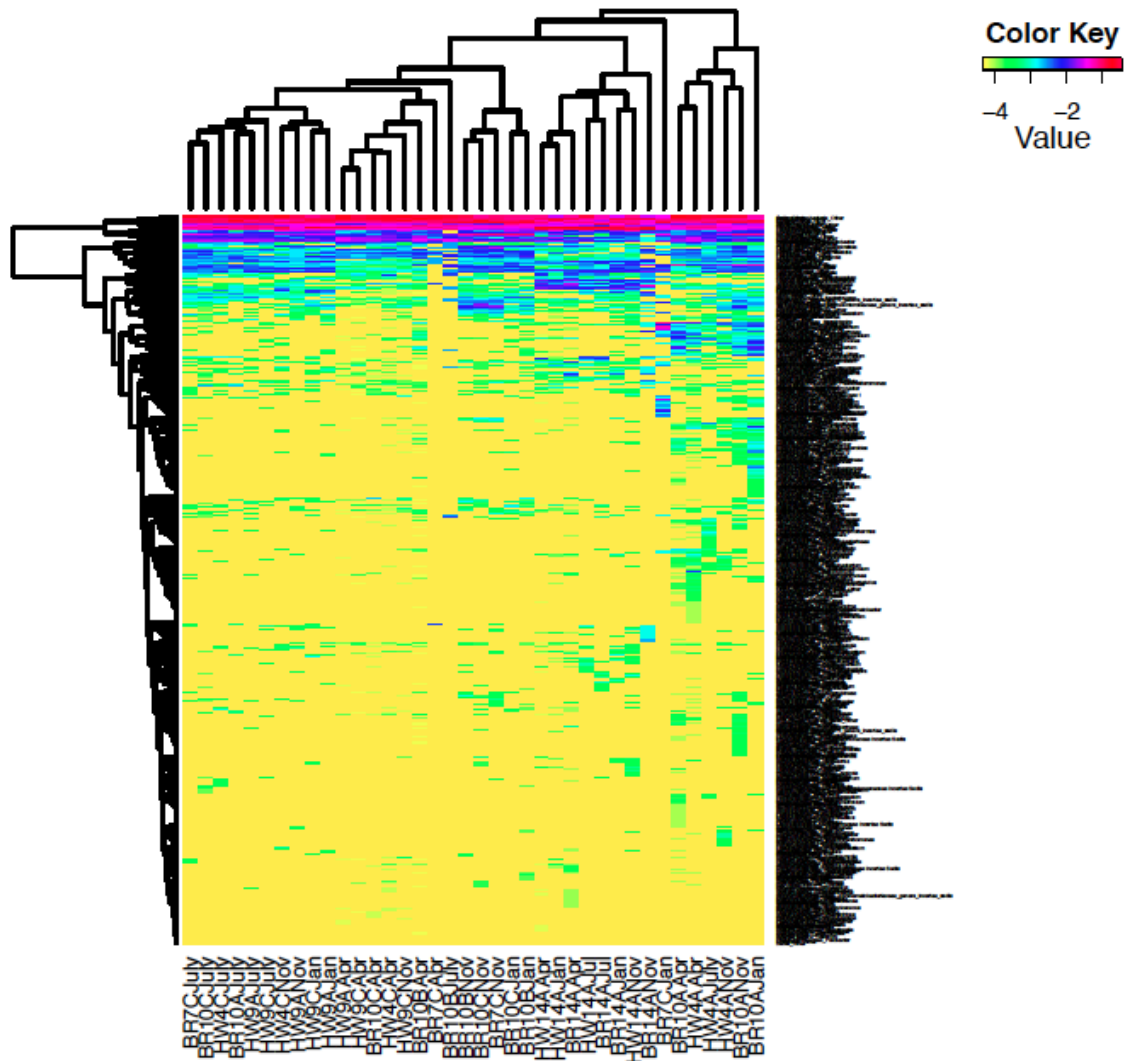
APPENDIX 4-Phylum level skiff plot



APPENDIX 5-Order level skiff plot



APPENDIX 7-Genus level skiff plot



APPENDIX 8-Raw SAS multiple linear regression analysis output

Flavobacteriaceae

R-Square = 0.4817

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	0.01213	0.00404	10.53	<.0001
Error	34	0.01305	0.00038390		
Corrected Total	37	0.02518			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-0.40162	0.18897	0.00173	4.52	0.0409
Time	-0.00000166	6.077094E-7	0.00286	7.46	0.0099
Salinity	0.01414	0.00509	0.00297	7.72	0.0088
Phaeopigments	0.11077	0.02363	0.00844	21.98	<.0001

Synechococcus

R-Square = 0.5882

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	0.16102	0.02300	6.12	0.0002
Error	30	0.11274	0.00376		
Corrected Total	37	0.27376			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	8.42788	2.31302	0.04989	13.28	0.0010
Date	-0.00041579	0.00011868	0.04612	12.27	0.0015
Conductivity	-0.07079	0.02543	0.02913	7.75	0.0092
Phaeopigments	-0.27645	0.10771	0.02476	6.59	0.0155
TSS	0.16866	0.07970	0.01683	4.48	0.0427
Total_Nitrogen	-0.09435	0.05213	0.01231	3.28	0.0804
NO3_N	0.14870	0.08291	0.01209	3.22	0.0830
Si	-0.01899	0.00599	0.03779	10.06	0.0035

Rhodobacteraceae

R-Square = 0.9347

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	0.05933	0.00742	51.87	<.0001

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Error	29	0.00415	0.00014298		
Corrected Total	37	0.06348			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-12.18465	1.90332	0.00586	40.98	<.0001
Date	-0.00006772	0.00002518	0.00103	7.23	0.0117
Depth	-0.00038859	0.00019988	0.00054040	3.78	0.0616
Salinity	0.04264	0.00739	0.00476	33.29	<.0001
Temperature	0.13457	0.01979	0.00661	46.22	<.0001
O2_Saturation	1.22128	0.17931	0.00663	46.39	<.0001
pH	0.04830	0.02263	0.00065132	4.56	0.0414
TSS	0.08893	0.00927	0.01316	92.02	<.0001
NO3_N	-0.01007	0.00287	0.00176	12.34	0.0015

Candidatus_Portiera

R-Square = 0.7431

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	0.03461	0.00494	12.40	<.0001
Error	30	0.01197	0.00039885		
Corrected Total	37	0.04658			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	4.92203	0.87282	0.01268	31.80	<.0001
Date	-0.00019756	0.00004073	0.00938	23.53	<.0001
Time	-0.00000201	9.384402E-7	0.00184	4.61	0.0400
Salinity	-0.02313	0.00797	0.00336	8.42	0.0069
Temperature	-0.00552	0.00201	0.00303	7.59	0.0099
Chlorophyll_a	0.03764	0.01951	0.00148	3.72	0.0632
Phaeopigments	-0.25274	0.07162	0.00497	12.45	0.0014
Total_Nitrogen	-0.01259	0.00315	0.00638	16.00	0.0004

Pelagibacteraceae

R-Square = 0.2745

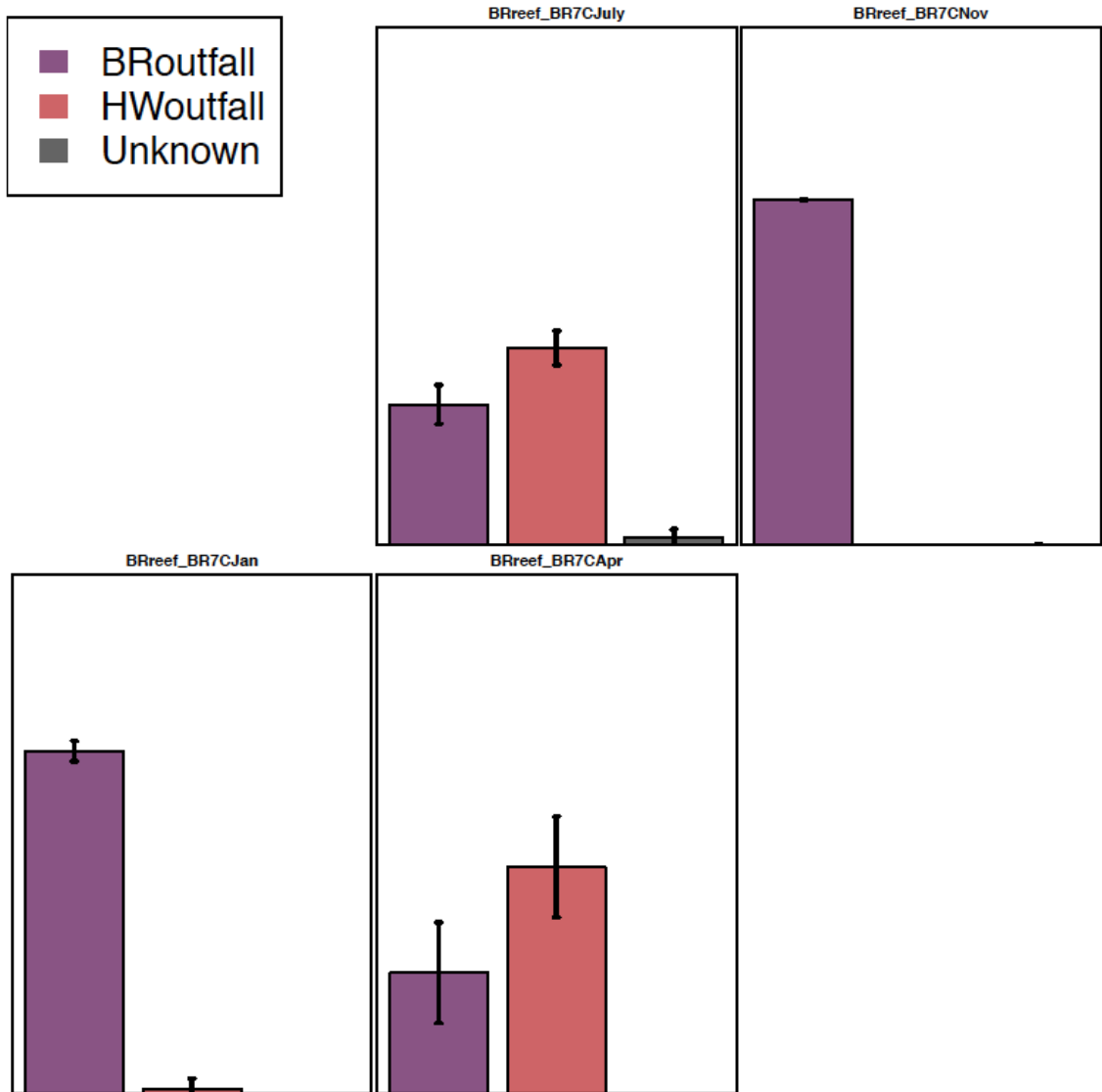
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.01967	0.01967	13.62	0.0007
Error	36	0.05200	0.00144		
Corrected Total	37	0.07167			

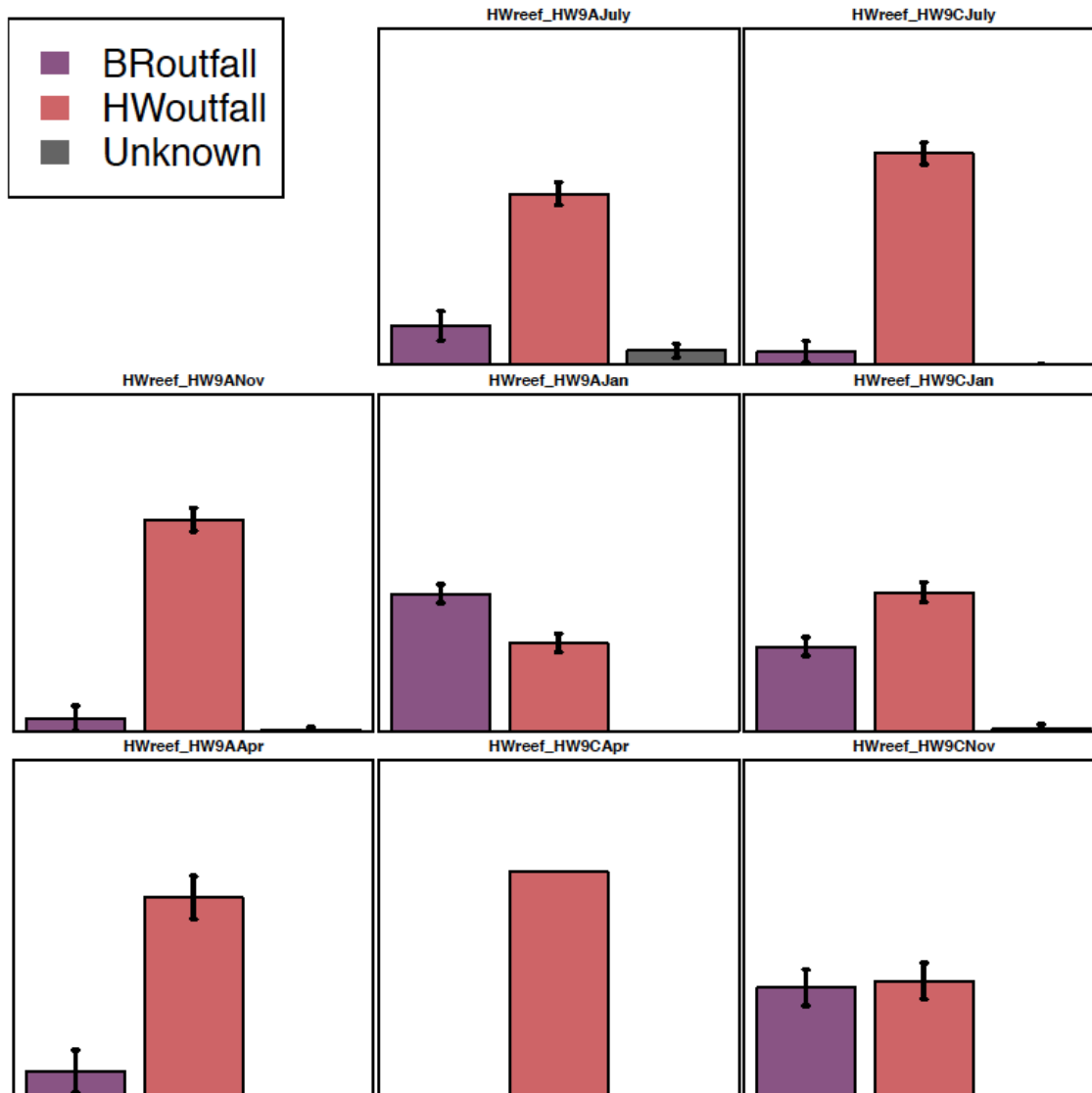
Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	0.11520	0.00886	0.24438	169.19	<.0001
Chlorophyll_a	-0.03257	0.00883	0.01967	13.62	0.0007

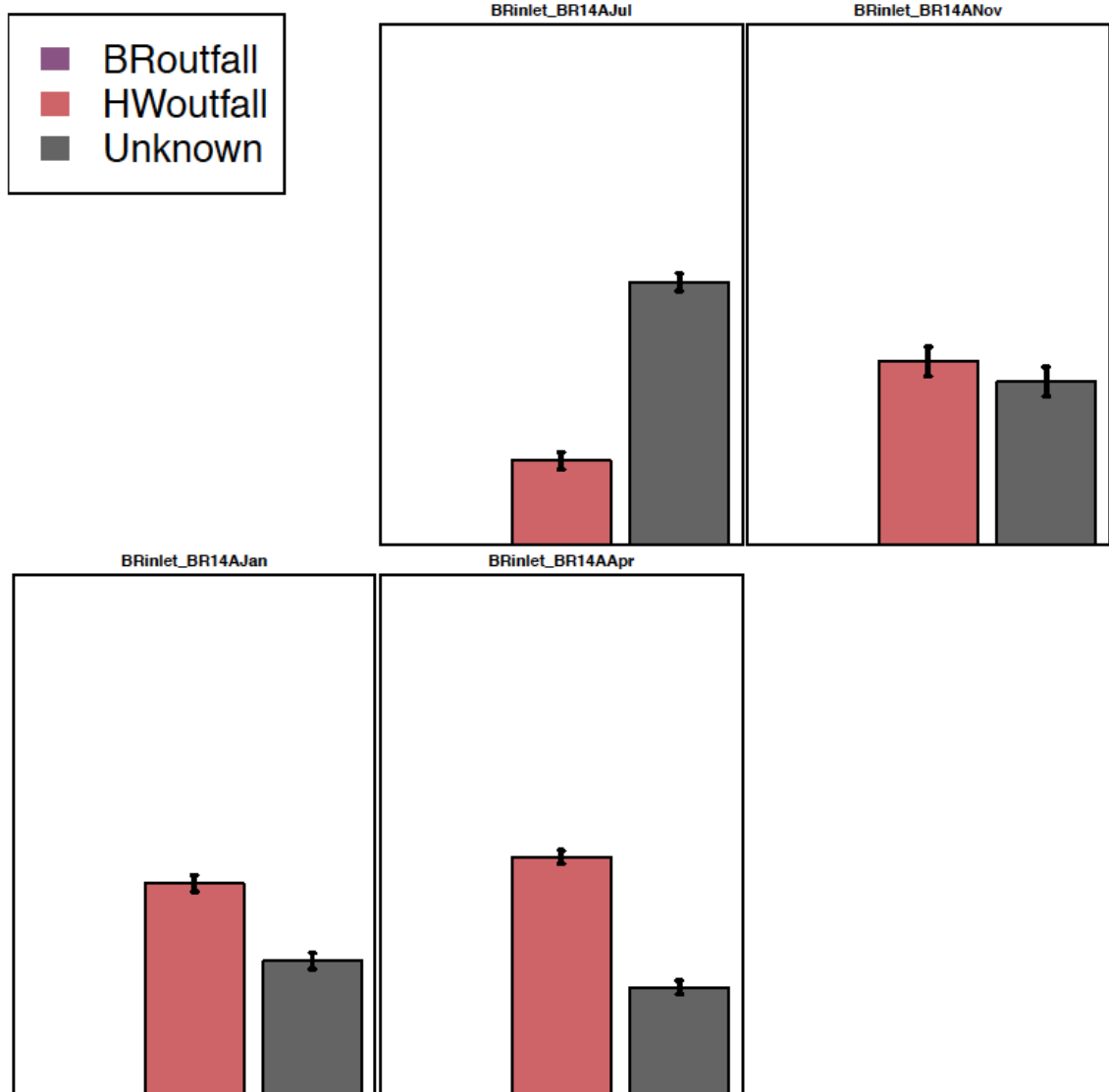
APPENDIX 9

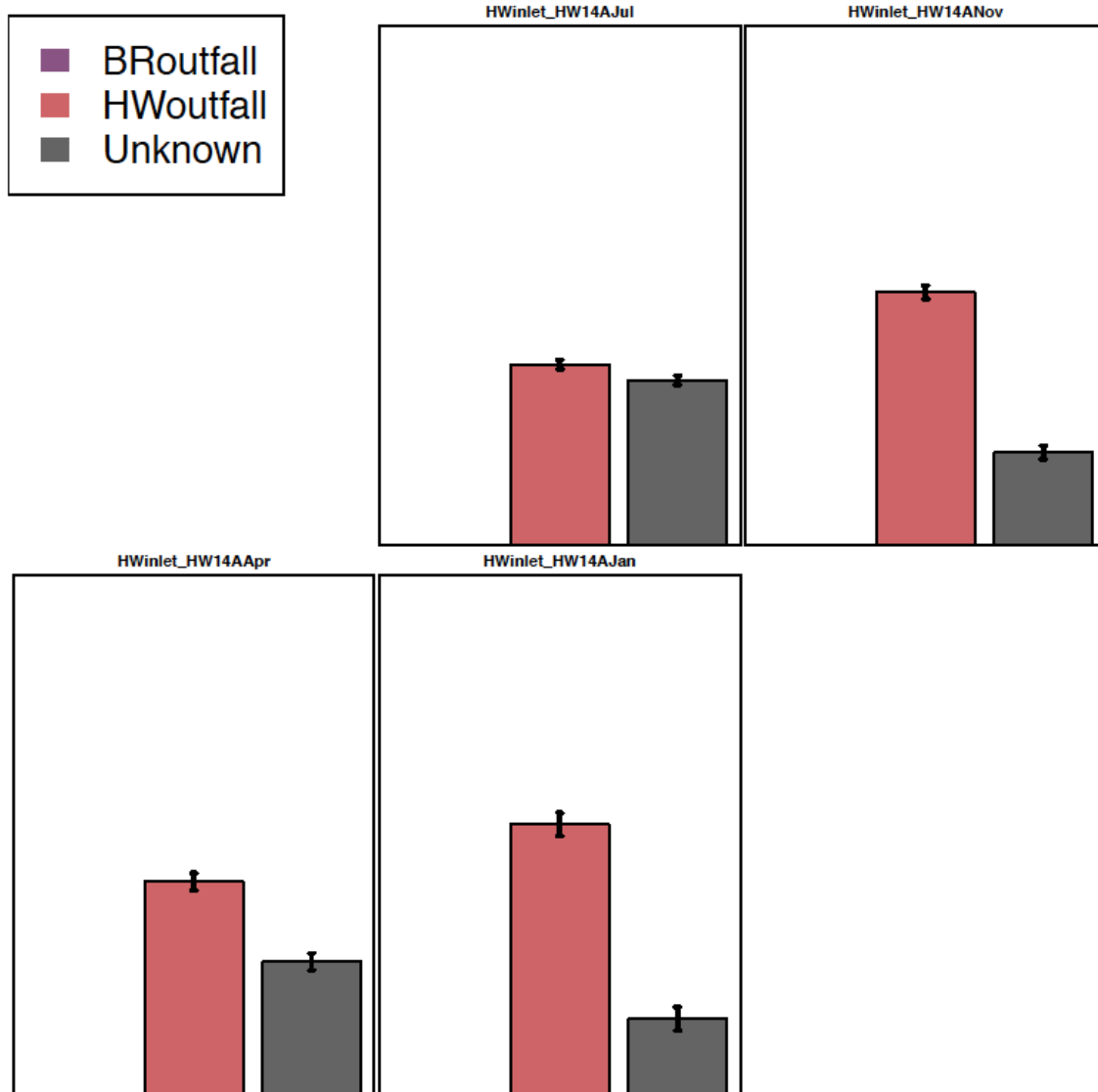
Name	mean(group1)	variance(group1)	std.err(group1)	mean(group2)	variance(group2)	std.err(group2)	pvalue	qvalue
Acidimicrobiales	0.000107	2.3E-08	1.1E-04	1.5E-03	3.4E-06	0.0003080	0.0423	1
f_[Balneolaceae]	0.002688	2.9E-07	3.8E-04	9.1E-04	8.7E-07	0.0001555	0.0386	1
Cryptophyta	0.002010	1.8E-07	3.0E-04	1.1E-02	1.5E-04	0.0020357	0.0335	1
Alphaproteobacteria	0.111895	5.4E-06	1.6E-03	6.5E-02	8.1E-04	0.0047429	0.0018	1
Rhodospirillaceae	0.002116	1.8E-08	9.5E-05	4.1E-03	7.2E-06	0.0004472	0.0375	1
Sva0853	0.000394	2.2E-08	1.1E-04	2.6E-03	9.8E-06	0.0005205	0.0439	1
Thiohalorhabdaceae	0.000000	0.0E+00	0.0E+00	3.0E-04	1.9E-07	0.0000725	0.0466	1
c_Verruco-5;o_R76-B128	0.001902	2.8E-08	1.2E-04	8.6E-04	1.5E-06	0.0002056	0.0375	1

APPENDIX 10-SourceTracker graphic output-outfalls as source









APPENDIX 12-Graphical output of SourceTracker-inlets as source

