12-2-2019

Characterization of Bacterial Communities in Biscayne Bay Through Genomic Analysis

Eric Fortman

Thesis of
Eric Fortman

Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Science
M.S. Biological Sciences

Nova Southeastern University
Halmos College of Natural Sciences and Oceanography

December 2019

Approved:
Thesis Committee

Major Professor: Jose V. Lopez PhD.

Committee Member: J. Matthew Hoch PhD.

Committee Member: Robert P. Smith PhD.

HALMOS COLLEGE OF NATURAL SCIENCES AND OCEANOGRAPHY

Characterization of Bacterial Communities in Biscayne Bay Through Genomic Analysis

By

P. Eric Fortman

Submitted to the Faculty of Halmos College of Natural Sciences and Oceanography in partial fulfillment of the requirements for the degree of Master of Science with a specialty in

Biology

Nova Southeastern University

December 2019

Table of Contents

Acknowledgements

I would like to recognize the members of my committee: Jose V. Lopez, J. Matt Hoch, and Robert Smith. I would especially like to thank J. Matt Hoch for being a friend and mentor for over a decade. Chris Avila and his team at DERM for providing logistic support, and a deep knowledge of Biscayne Bay. My lab mates for providing feedback and support —especially Colleen McManken, for help processing samples; Chase Donnelly for training me; and Denise Swack for the company and help during lab work. I would like to thank my parents —Pam and Phil, for all their support. My undergraduate mentor Denise Flaherty, for giving me a solid foundation in genetics. And my climbing partners, Grant Albert and Paige Taft for listening to me complain.

Abstract

Biscayne Bay is a shallow oligotrophic estuary in Southeast Florida. Channelization of rivers, and dredging of canals has greatly altered the historical flow of fresh water into the bay. This, coupled with the rise of a sprawling urban & suburban development, has greatly increased the nutrient load in the bay. This study examined the bacterial community at 14 stations throughout Biscayne Bay —6 stations were located at the mouths of canals; 1 upstream-canal station; 6 stations in the center of the bay; and one ocean influenced station, located near the entrance to the bay. One liter, surface water samples were taken monthly for one year. The 16S rRNA gene was used to identify bacterial community composition. There were 19,680 Amplicon Sequence Variants (ASVs) identified across all 146 samples. Salinity and total phosphorous were the primary factors explaining bacterial biodiversity. Biodiversity in microbial communities in the Miami River and the ocean influenced site, were unique compared to other sites in the study. Alpha and β-diversity were generally homogeneous over most of the study area. Looking at α-diversity, the two stations on the Miami River were statistically identical and had higher diversity. The ocean influenced station, located near the Safety Valve, was statistically unique, and had lower α-diversity. The remaining 11 stations had moderate diversity and were statistically identical, appearing to be a combination of the previously mentioned Miami River sites and the ocean influenced site. Beta diversity showed a similar pattern; with the exception that the site located at the mouth of Black Creek could now be grouped with the Miami River sites.

Key words: Biscayne Bay, Florida, Microbiome, 16S, 16S rRNA, Bacterial Ecology

Introduction

Biscayne Bay

Biscayne Bay is a shallow oligotrophic estuary on the southeast coast of Florida. Excluding dredged areas, the maximum depth is 4 m, with an average of 1.8 m (Caccia & Boyer, 2005). The Bay was formed about 4,000 years ago, when a rising sea filled a freshwater marsh (Leynes & Cullison, 1998). Historically there were several free-flowing rivers into the bay. Landscape level human impacts began with efforts to drain the Everglades starting in 1903 (Cantillo et al., 2000). Rivers were channelized and new canals dredged, to increase water flow out of western Dade County. By 1913 the rapids on the Miami River had been removed, and the Snapper Creek Canal, Cutler Canal, and the Coral Gables Waterway were dredged (Cantillo et al., 2000). The historic pattern of seasonal freshwater flow from rivers, creeks and sloughs in to the bay, has been replaced by discrete releases through flood gates along canals. Water flow is tightly controlled by the South Florida Water Management District and Army Corps of Engineers. There are 19 canals that drain into Biscayne Bay (Cantillo et al., 2000). The primary drainages for urban and suburban Dade County are the Little River, Miami River, Coral Gables Waterway, Snapper Creek, Cutler Drain, and Black Creek Canal. Further south in the bay the Princeton and Mowry canals drain agricultural and some suburban areas, but these are beyond the scope of this study.

Seasonality in South Florida is principally delineated by rainfall, with the wet season running form May–October and the dry season running form November–April (Dame et al., 2000). Because of the bay's large surface area, precipitation is the dominant source of freshwater to the bay; followed by canal input and ground water discharge (Stalker et. al, 2009). Biscayne Bay is periodically exposed to naturally occurring disturbances such as tropical cyclones. In August 2005 Hurricane Katerina hit the Bay dumping up to 14" of rainfall within the watershed (Zhang et al., 2009). While the storm event caused many short-term changes to water quality, Zhang et al. (2009) observed that water quality returned to pre-storm conditions within three months of the event. More recently in September 2017 Hurricane Irma hit Biscayne Bay. The hurricane significantly increased freshwater inflow to the bay. In the first week after the storm freshwater inflow increased by 148% –compared to a week before (Wachnicka et al., 2019).

Similar to Hurricane Katrina water quality in the bay returned to "normal" in less than three months after Hurricane Irma (Wachnicka et al., 2019).

The northern end of the Bay (Oleta River through Key Biscayne) is more sheltered and receives less water exchange with the Atlantic Ocean. Dredging of Government Cut began in 1902. The spoils were used to construct Lummus, Dodge, and Fisher Islands –the first man made islands in the bay. The North Bay is now heavily modified, with very little natural shoreline remaining. This area is also home to the most urban and industrial land use. Turbidity, industrial pollution, nutrient loading, and sewage pollution are the primary problems facing the Northern Bay (Caccia & Boyer, 2005). The portion of the bay south of Key Biscayne, through the Safety Valve, and Ragged Keys sees more exchange with oceanic water (see figure 3). Development becomes less dense as you move south along the cost. The central bay (the area south of Cape Florida through Black Point) is characterized by suburban development and more remaining mangrove tracts along the coast. Pollution sources here come from localized problems such as marinas (Caccia & Boyer, 2005). The mainland of the southern Bay (Black Point to Card Sound) is a mix of suburban development, agriculture, and mangrove habitat. One anthropogenic feature of note in this area is the South Dade land fill, near Black Point.

In general, nutrient loads are higher near the coast (Caccia & Boyer, 2005). Nutrients from septic tanks, leaky sewage lines, and fertilizer have led to eutrophication in the Bay. Caccia & Boyer (2005) identified several geographic patterns in water quality in the bay, noting that land use is the major factor affecting water quality in the bay. Eutrophication from nitrate/nitrite–nitrogen seems to be more of a problem in the southern part of the watershed. Whereas total ammonia-nitrogen and total phosphorus are the major pollutants in the northern part of the watershed (Caccia & Boyer, 2007; Carey et al., 2007). Canals are responsible for the bulk of nitrogenous inputs into the bay (Caccia & Boyer, 2007; any more recent refs from other areas?). Stalker et. al (2009) cautions even though ground water is the lowest constituent of freshwater input, it should not be ignored because it generally contains higher levels of nutrients, notably nitrogen and phosphorous. As a result of the increased nutrient load, persistent algal blooms and reduced seagrass coverage have been reported. Collado-Vides et al. (2013) described a persistent

bloom of *Anadyomene spp*., which was first noted in 2006. The geographic range of the bloom extended from the Rickenbacker Causeway south to Chicken Key, with some sites experiencing algal coverage > 75% (Collado-Vides et al. 2013).

Macrobiomes

The typical habitats present in Biscayne bay include mangrove shoreline; seagrass, sand, and mud flats; patch reefs; hardbottom communities: consisting of sponge and soft corals (Cantillo et al., 2000). Three species of mangrove are found in the bay: red (*Rhizophora mangle*), black (*Avicennia germinans*) and white (*Laguncularia racemosa*). Mangroves provide many ecosystem services, perhaps most importantly is shoreline stabilization. Replacement of mangroves with seawalls or unstabilized shoreline, in the northern bay, is one of the main causes of the turbidity in the area (Caccia & Boyer, 2005). Seagrass flats provide sediment stabilization

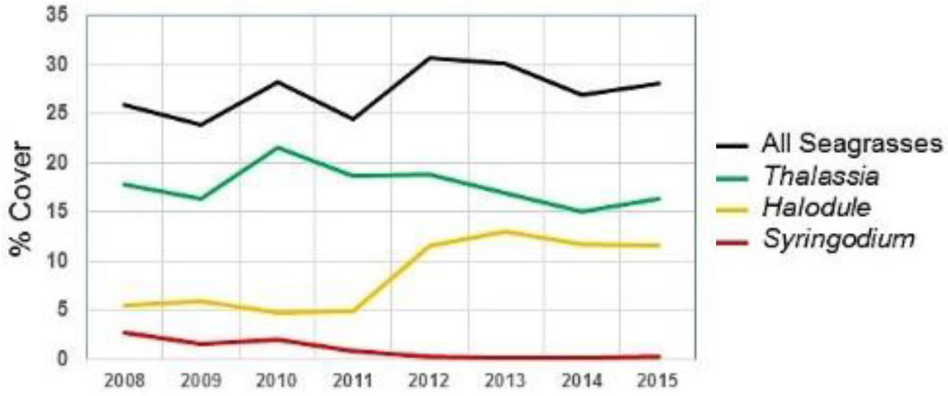


Figure 1: Percent cover of seagrasses in nearshore (<500 m from shore) habitats of western Biscayne Bay from Matheson Hammock to Turkey Point, 2008–2015. (From: Lirman et al., 2016)

within the basin. Seagrass meadows are primarily composed of three species: turtle grass (*Thalassia testudinum*), manatee grass (*Syringodium filiforme*), and shoal grass (*Halodule wrightii*). While *T. testudinum* is the dominate species in the bay, *H. wrightii* is tolerant of the widest range of salinities and is often found near the mouths of canals (Lirman & Cropper, 2003). Because of turbidity and nutrient loading *T. testudinum* is not reported north of the Port of Miami (Lirman et al., 2016). From 2008-2015 average seagrass coverage in nearshore waters has oscillated between 24-31% (Figure 1; Lirman et al., 2016). Between 2011 and 2015 the percent

coverage of *Halodule wrightii*, in the nearshore environment, has more than doubled (Lirman et al., 2016). This may be due to the tolerance of *H. wrightii* to a wide range of salinities. A loss in overall seagrass density would cause an increase in phytoplankton abundance, which would present as an increase of chlorophyll *a* concentration (Millette et al., 2017).

There are at least 400 species of fish in the bay, ~93 have been identified as economically important to either the food, bait or aquarium trade (Ault et al., 2007; Idyll et al., 1999). The bay hosts several federally listed endangered/threatened species including American crocodile, West Indian manatee, and several species of sea turtle (Cantillo et al., 2000).

Microbiomes

Microbes in natural habitats generally exist as microbial communities (or "microbiomes") instead of in isolation. Marine bacterioplankton microbiomes play an important role in many biogeochemical processes (Bunse & Pinhassi, 2017). In marine ecosystems heterotrophic bacteria are the only organisms that fix dissolved organic material for use by primary producers (Bunse & Pinhassi, 2017). Bacterial diversity is often higher in eutrophic waters because of the high abundance of organic material (Rösel et al. 2012; Tang et al. 2015). Seasonal variability in the microbial community is more pronounced in temperate and polar habitats, but it is still observed in subtropical and tropical regions (Figure 2; Bunse & Pinhassi, 2017). In Port Everglades inlet, an estuary just north of Biscayne Bay, seasonal variation in the bacterioplankton community was noted by O'Connell et al. (2018). The wet (May – October) season was characterized by higher species richness, and lower species evenness. Changes in community composition were most closely tied to changes in salinity and temperature (O'Connell et al., 2018).

Population dynamics of bacteria and phytoplankton influence each other (Bunse & Pinhassi, 2017; Smith et al., 1999). Further, phytoplankton blooms can decrease light penetration and shade seagrasses, causing reduced seagrass coverage. In turn this causes the release of nutrients tied up in seagrass biomass and sediments, exacerbating the bloom (Boyer et al., 2009). A better understanding of how bacteria and phytoplankton affect each other can have applications in predicting and preventing hazardous algae blooms. Most time series data for microbiome studies are sampled in monthly intervals. However, the generation time of

bacterioplankton can be hours or days. Therefore smaller-scale population fluctuations may serve as a precursor for more prolonged ecological shifts (Bunse & Pinhassi, 2017).

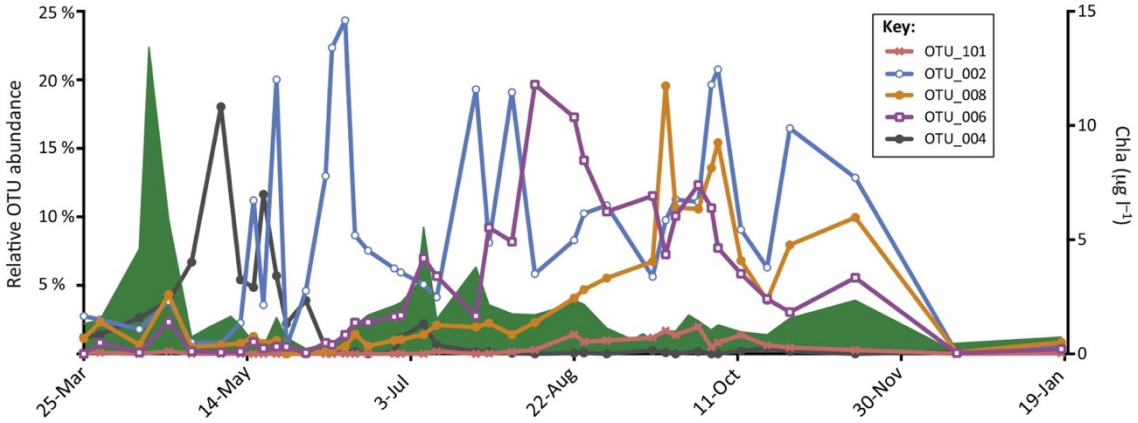


Figure 2: Seasonal Succession of Marine Bacterioplankton. Changes in relative abundances of bacterial populations operational taxonomic units (OTUs) in the temperate Baltic Sea during 2011; redrawn from Lindh et al. Note differences in timing, duration, and amplitude of changes in abundances over time (From: Bunse & Pinhassi, 2017).

16S RNA

Traditionally, bacterial communities were studied by plating environmental samples on a petri dish and culturing them in the lab. A major drawback to this technique is that many –if not most– species of bacteria do not grow well in the laboratory (Pace, 1997). Advancements in genetic techniques, now allow environmental samples to be tested directly. High throughput sequencing allows researchers to sequence genes, relatively quickly and cheaply (Mardis, 2008). These technologies have also made it possible to obtain sequences from many organisms simultaneously. The resulting data can then be analyzed to identify the number of amplicon sequence variants (ASVs) or be BLASTed to search for known sequences from specific species. An ASV is an genetic sequence that is used as a proxy to represent a discrete taxa. An ASV has no Linnaean rank. But it can be cross-referenced to a database of known sequences to link the ASV to a specific Linnaean taxa (e.g. Kingom, Phyla, Class, etc.). The number of ASVs present can be used a proxy to measure diversity and identify community structure. A Basic Local Alignment Search Tool (BLAST) may be used to identify the prescience of specific organisms. In order to do

this there must be a preexisting sequence that has been identified to use a template for the search. This tool can be especial useful for identify the presence of bacteria belonging to certain guilds; e.g. oil degraders, nitrogen fixers. This has been demonstrated by Mustafa et al. (2016) identified microbial communities dominated by hydrocarbon digesting bacteria at contaminated ports in the Red Sea.

The 16S rRNA gene was first used to study phylogeny in 1977 by Woese & Fox. The 16S gene has become the standard for bacterial phylogeny for three reasons: it is present in nearly all bacteria; the function of the gene has not changed over time, suggesting randomly occurring mutations are a good measure of evolution; the gene is suitably large (1,500bp) for informatics analysis (Janda & Abbott, 2007). Despite advances in whole-genome sequencing techniques, amplicon sequencing of the 16S rRNA gene is still a viable method for comparing bacterial communities (Thompson et al., 2017).

In 2010 the Earth Microbiome Project (EMP) was founded to survey bacterial, archaeal, and eukaryotic microbial diversity (Thompson et al., 2017). The EMP suggests using exact sequences of 16s rRNA and a standardized, but decentralized approach for compiling a catalog of microbiological life on Earth. Thompson et al. (2017) suggests using the software package *Deblur* to denoise and assemble sequences into ASVs (amplicon sequence variants). However, *DADA2* may be a better alternative (Callahan et al., 2016). *DADA2* leverages finer-scale resolution to groups sequences into ASVs –which may reveal more information about ecological niches, temporal dynamics, and population structure (Callahan et al., 2016). Thompson et al. (2017) found a weak but significant increase in environmental microbiome diversity at lower latitudes. The Earth Microbiome project emphasizes the importance on collecting physicochemical parameters (e.g. salinity, temperature, nutrient data) for each genetic sample. These meta data are key for revealing global patterns of microbial diversity (Thompson et al. 2017).

Hypotheses

1. Microbial community will correlate closely with water quality. More oligotrophic areas will have lower diversity and eutrophic areas will display higher diversity.

2. Stations located at canal mouths will have higher diversity.

3. The ocean influenced site will have the lowest diversity.

4. Sites located in the middle of the bay would have moderate diversity.

Methods

Sample Collection

Water samples were collected in partnership with Miami-Dade County's Division of Environmental Resource Management (DERM). There were 14 fixed-stations throughout Biscayne Bay that were irregularly sampled between September 2017 and January 2019 (Figure 3 & table 1). The samples used for genetic analysis consisted of 1.0L surface water grab-samples. Several more liters of water were collected by DERM for chemo-physical analysis that included: salinity, temperature, dissolved oxygen, ammonia-nitrogen, nitrate-nitrite, and total phosphate. The sample locations range from Little River down through Black Point (Figure 3). These chemo-physical data are key for providing context for microbiome data (Knight et al. 2012).

Sample Preparation & Sequencing

The samples bound for genetic analysis were filtered through a 0.45µm nylon filter. DNA extraction conducted using a *Qiagen DNeasy PowerSoil Kit*. The extracted DNA then went through a series of Quality control steps. To confirm successful extraction each sample was run on an agarose gel. A test PCR using *Platnum MasterMix,* 515 forward and 806 reverse primers was conducted to confirm the DNA could be successfully amplified. Another gel electrophoresis was done to check for the successful amplification of the 16s region, which is ~300bp in length. Then another PCR was run, this time using a barcoded 515F primer and an 806R primer with a barcode unique to each sample. Magnetic *AMPure XP* beads were used to purify the 16S V3 and V4 amplicon away from free primers and primer dimer species. The DNA was then quantified using *Qubit* high sensitivity fluorometry, and diluted to 4.0 n*M* for sequencing. The samples were pooled and then, as a final quality control step, automated electrophoresis was conducted using an *Agilent* TapeStation. Sequencing was performed on the *Illumina MiSeq* platform. Proof of

theory establishing that sequencing on the MiSeq platform accurately reflects a known bacterial community was established by Caporaso et al. (2012).

Microbiome analyses

The mapping file, which matches the sample names to their respective nucleotide barcodes was validated using the *Keemei* plug in in Google Sheets. The MiSeq output, containing the DNA sequences, was post processed using QIIME2 –an open source, Unix based command line program specifically designed for microbial community analysis (Bolyen et al., 2018). The forward and reverse reads along with the index file were imported to QIIME2 as a *QIIME artifact* (.qza file) using the *emp-import* command. Because the samples are pooled when they are loaded into the sequencer the output emerges as a tangle of data, which needs to be teased apart in software. The mapping file is used to tell the software which barcodes belong to which samples. In QIIME2 the *demux* command was used to untangle (demultiplex) the samples. Within QIIME2 the DADA2 algorithm was used to remove chimeras (artifact sequences that don't represent a real organism) and reads with a quality score <25, this was done using the *dada2 denoise-paired* command. The quality score is prediction of the probability of an error the sequencer misidentifying a nucleotide base (Illumina, 2016) .The advantage of DADA2 over other denoising techniques is that it infers sample sequences exactly, without coarse-graining into OTUs, and has high resolution –resolving differences of as little as one nucleotide (Callahan et al., 2016). Using exact sequences offers more flexibility than ASVs. By nature, exact sequences are "stable identifiers" and can be compared to any 16s rRNA database (Thompson et al., 2017). An alpha rarefication plot was generated in QIIME2 and used to determine if adequate sampling depth was achieved. Phylogenetic trees were constructed using the *mafft* alignment and *fasttree* commands. Taxonomy was determined for each unique sequence, by comparing the sequence to the Silva 132 learned classifier. The feature table, taxonomy file, and phylogenetic tree was exported from QIIME2 for downstream analysis in *R Studio* with the *PhyloSeq* and *Vegan* packages.

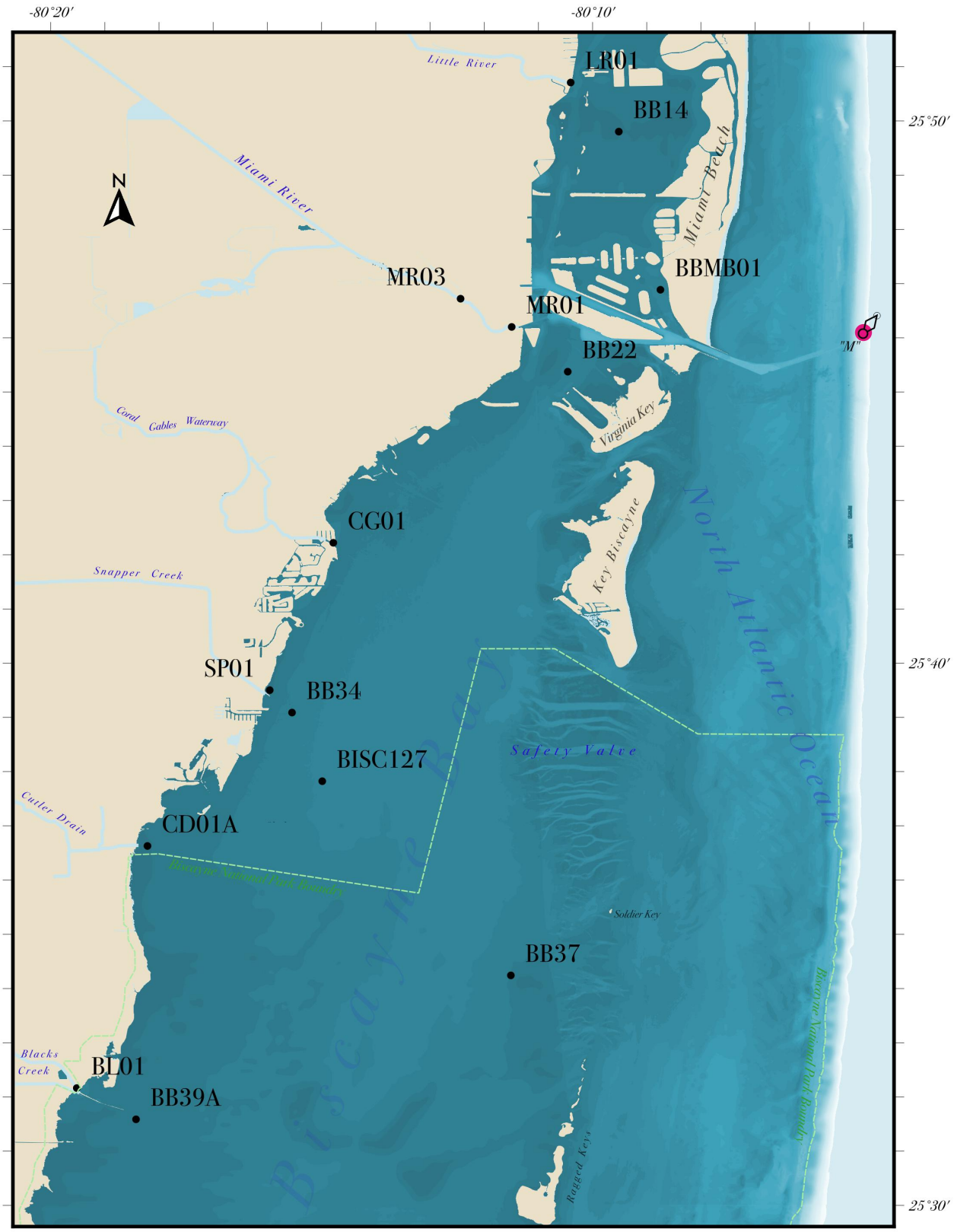


Figure 3: geographic locations of the 14 sampling sites. This map was generated using *QGIS v2.18*. Bathometry data was derived from LandSat data.

Table 1: List of field sites, showing their absolute location (latitude & longitude, in decimal degrees) along with a description of their relative location. The sites can be grouped into broad categories based on their geographic location: bay, ocean influenced, canal mouth, & canal. Note: MR03 is the upstream Miami River site, sometimes referred to as the canal site. BB37 is the most seaward site and is subject to the most influence form oceanic water.

Site	Site Type	Absolute Location		Relative Location
BB14	Bay	25.83008	-80.15857	Biscayne Bay North of Julia Tuttle Cswy, 2km east of green Mrk "31"
BB22	Bay	25.75628	-80.17427	Midway between Marine Stadium and NOAA slip at Dodge Island, 1.4 km east of ICW, green Mrk "65"
BB34	Bay	25.65148	-80.25907	Biscayne Bay 2000m east of the mouth of Snapper Creek (C-2)
BB37	Ocean influenced	25.57068	-80.19177	West of Ragged Keys at green Mrk "1B"
BB39A	Bay	25.52643	-80.30706	Southeast of Black Point
BBMB01	Bay	25.78146	-80.14577	Biscayne Bay 260m west of the Bay Side Seawall and 11th Street (Miami Beach)
BISC127	Bay	25.63038	-80.24977	Approx. 1.8 Miles East of the Bay Side Seawall of Chapman Field Park at SW 152nd Street
BL01	Canal mouth	25.53604	-80.32527	Confluence of Goulds Channel and Black Creek Channel
CD01A	Canal mouth	25.61047	-80.30354	~1000m from mouth of canal, adjacent to the manatee sign.
CG01	Canal mouth	25.70368	-80.24637	SW 32nd Ave/SW 72nd St. Mouth of Coral Gables Waterway
LR01	Canal mouth	25.84517	-80.17337	Bayshore Ct/Belle Meade Blvd. Northern mouth of Little River
MR01	Canal mouth	25.77004	-80.19151	Biscayne Blvd/SW 3 St. Mouth of Miami River at green Mrk "3"
MR03	Canal	25.77871	-80.20723	NW 7 Ave/NW 6 St. Miami River between Wagner Creek and 5th St. bridge
SP01	Canal mouth	25.65837	-80.26593	SW 47 Ave/SW 124 St. mouth of Snapper Creek

The PhyloSeq package was used to analyze α-diversity. Alpha-diversity is the diversity of taxa within each site or sample (Whitaker, 1972) and was assessed using Shannon and Inverse Simpson indices. The Shannon and the Inverse Simpson indices are both measures of biodiversity. While there are many methods of measuring biodiversity, these were chosen because they are two of the most widely used. A non-parametric, Kruskal–Wallis test was used to see if α-diversity differs between each site and site type. The Vegan package was used to analyze β-diversity. Beta diversity is comparative diversity between sites, this assesses the similarity/dissimilarity of diversity between different sites. The Bray-Curtis Distances for β-diversity were calculated using Vegan. To assess relatedness between populations, Principle component analysis (PCoA) was be done using Vegan, which incorporates phylogenetic signals in the 16S rRNA data. A Kruskal-Wallis test was used to determine if the diversity metrics differed across sites, and group sites with similar diversity measures together. Several, Multiple Least Square Regression analyses were run to look for a possible correlations between microbial community and chemo-physical water quality data (Campbell et al, 2015; O'Connell et al, 2018). Canonical Correspondence Analysis (CCA) was used to identify possible correlations between species abundance and chemo-physical water quality data. To compare diversity between site types and identify taxa leading to significant differences, a SIMPER similarity percentage table was generated using *Vegan*. A relative abundance table was generated through Vegan, stacked bar graphs for relative abundance were generated using the R package *ggplot2* and stacked pie charts were generated using *Excel*.

Results

There were 19,680 bacterial taxa identified across all 146 samples. The alpha rarefication plot illustrates the plateau in α-diversity reached for each site type (Figure 4). The plateau signifies that an asymptote was reached during sequencing, and therefore adequate sampling depth was attained. This result indicates that within the sequencing run, no new taxa were being sequenced.

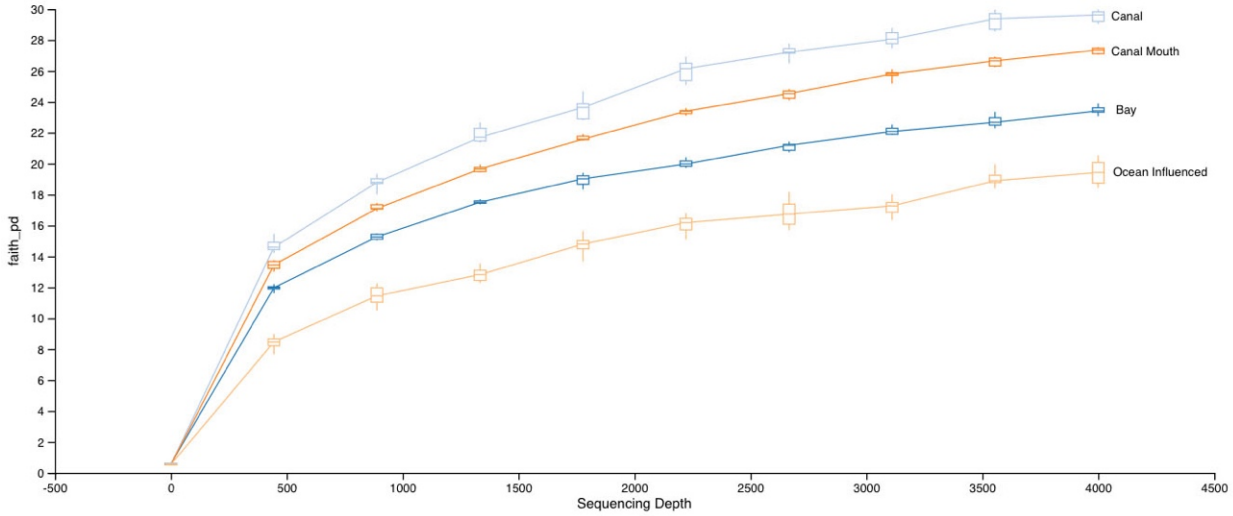


Figure 4: Alpha rarefaction plot for each site type. All four curves reach a plateau showing that adequate sampling depth was reached. This means that the microbial community at each site type was adequately sampled. Error bars represent standard deviation. For canal n=8; for canal mouth n= 58; for bay n= 71 for ocean influenced n= 9.

When looking at site type, the canal and canal mouth sites were statistically identical; and the bay and ocean influenced sites were statistically identical (p=6.406e-05) (Figure 5). A similar pattern is apparent Alpha diversity was also visualized with an NMDS plot (Figure 7). Looking at the sites individually, average α-diversity at MR01 (Miami River mouth), MR03 (the canal site) and BL01 (at the mouth of Black Creek) are statistically the same (Figure 6). These three sites also had some of the highest α-diversity observed in the study. Alpha diversity at BB37 (the ocean influenced site) is statistically distinct from all the other sites (p= 3.728e-03) (Figure 6). Site BB37 had the lowest α-diversity observed in the study. The remaining sites statistically fall in between these two extremes, sharing some combination of the "ocean influenced type" and the "Miami River type" sites. The ocean influenced site (BB37) had the least variability in α-diversity; while the Little River site (LR01) had the widest range of α-diversity, recorded in the study (Figure 6).

A PCoA (Principal Coordinates Analysis –used to asses dissimalarity) comparing β-diversity determined that MR01 and MR03 were statistically identical (Figure 8). Site BB37 (the ocean influenced site) is distinct from the all other sites. The remaning sites posessed characteristics of both the Miami River sites and the ocean influenced site (p= 7.649e-03).

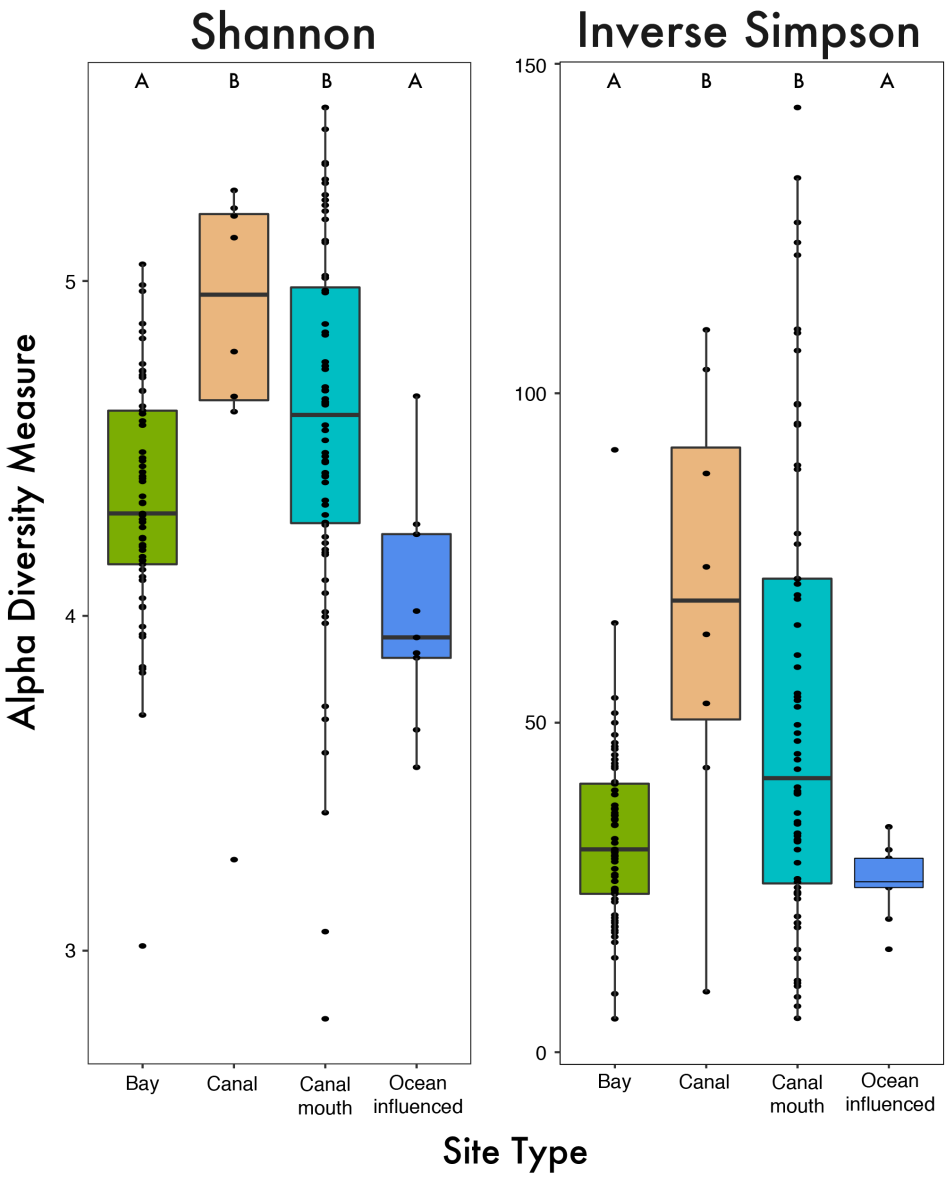


Figure 5: shows the α-diversity by site type. Alpha diversity at the canal and canal mouth sites are statistically the same. Alpha diversity at the Ocean influenced and Bay sites are statistically the same (p=6.406e-05). Alpha diversity was plotted on a graph and a boxplot was overlaid. Each point on the plot represents one sampling event. The thick black line with in the box represents the average α-diversity for the site. The higher the line, the higher the average α-diversity.

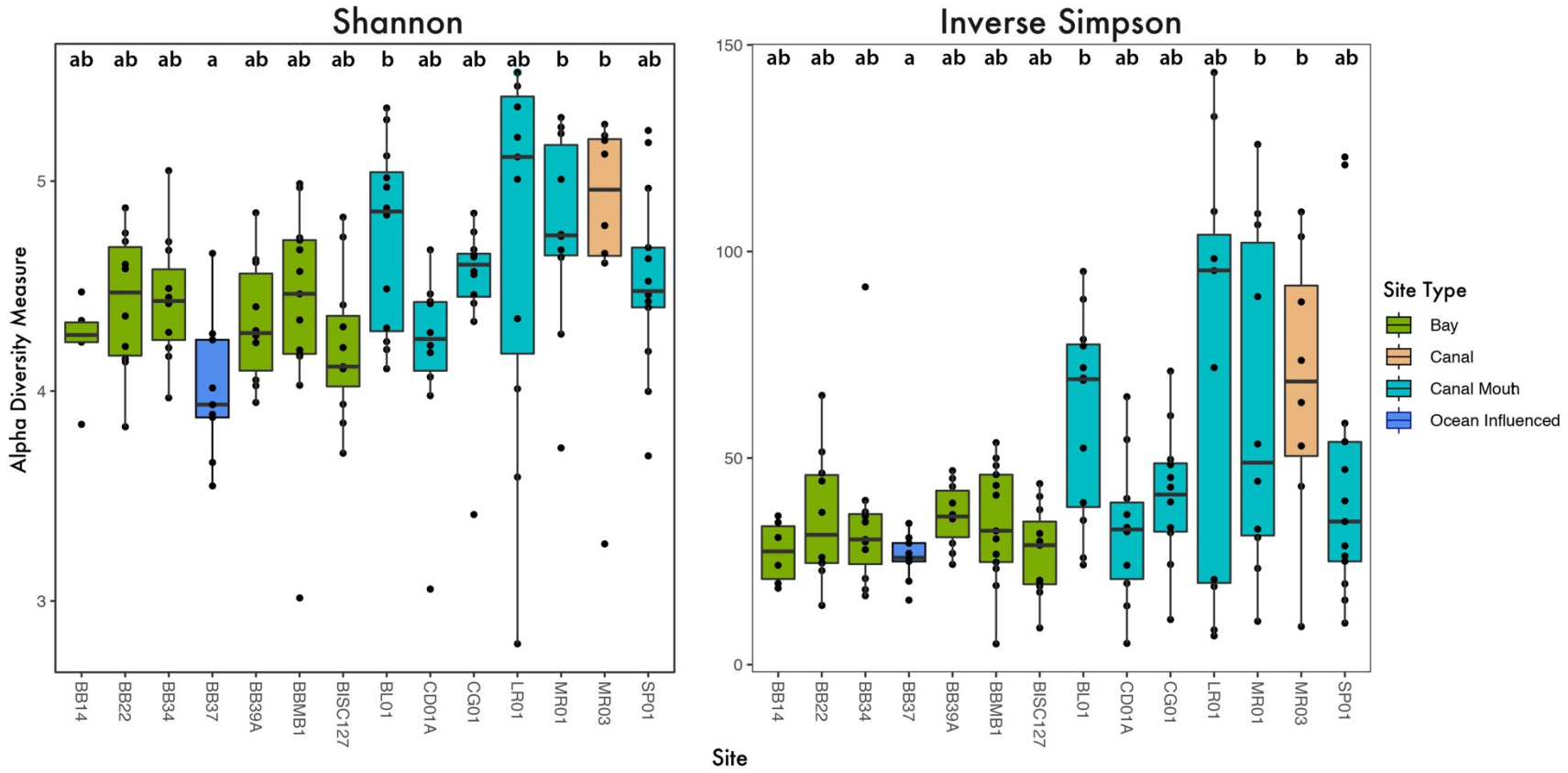


Figure 6: Box plots illustrating alpha diversity at each site. The sites could be grouped into two statistical groups: **A)** the oceanic site (BB37) **B)** the Miami River sites (MR01 & MR03) and the Mouth of Black Creek (BL01). **AB)** the rest of the sights share properties of both groups (p= 3.728e-03). Color indicates the site type. The Shannon and Inverse Simpson indices were used to measure alpha diversity. Alpha diversity was plotted on a graph and a boxplot was overlaid. Each point on the plot represents one sampling event. The thick black line with in the box represents the average α-diversity for the site. The higher the line, the higher the average α-diversity. These plots were generated using the Phyloseq package in R studio.

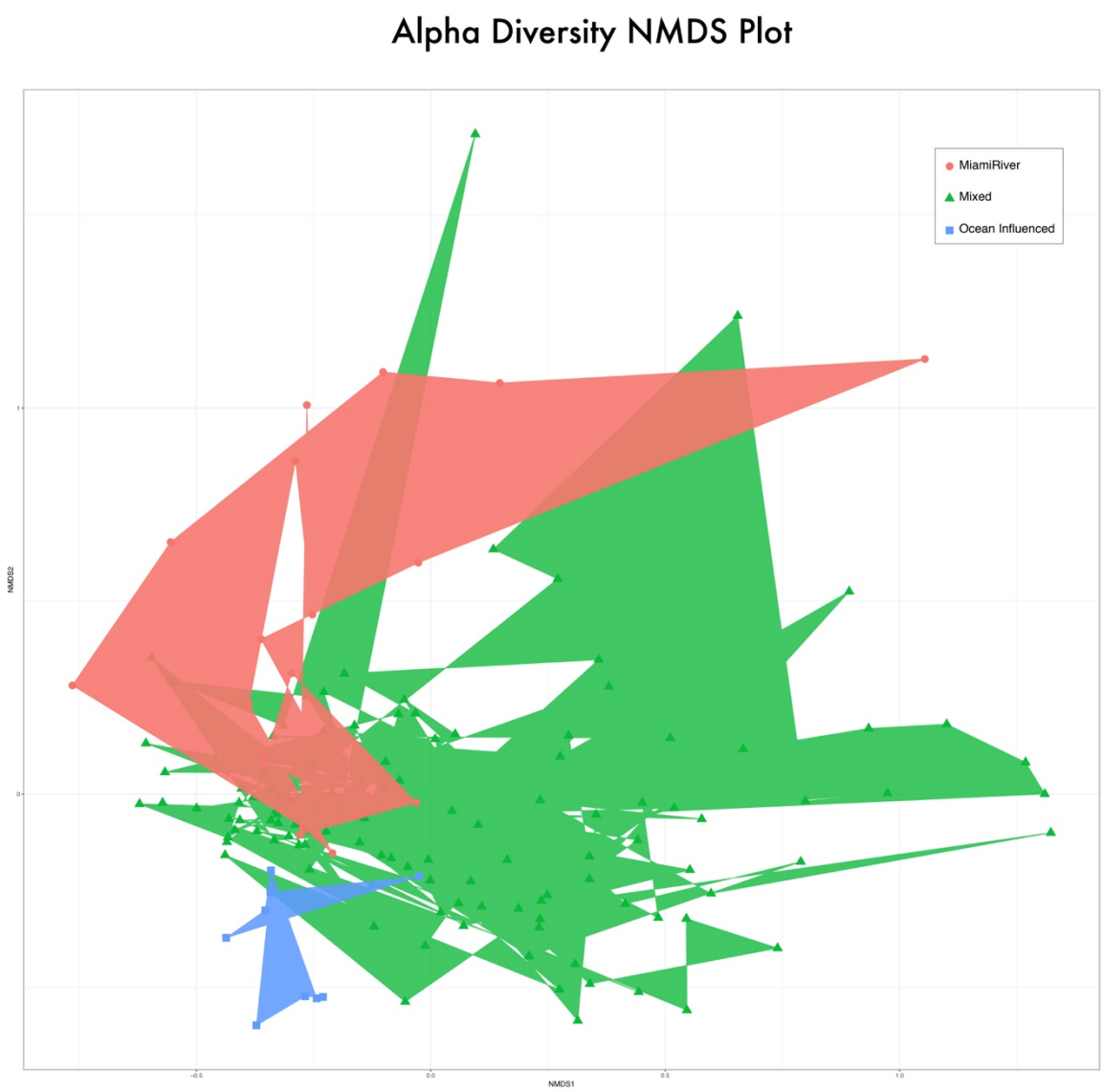


Figure 7: A non-metric multidimensional scaling plot showing the similarities of α-diversity for each sampling event. The ocean influenced site (BB37) forms one group. The Miami River sites form another group. The remaining sites (mixed) share aspects of both the Oceanic group and the Miami River group.

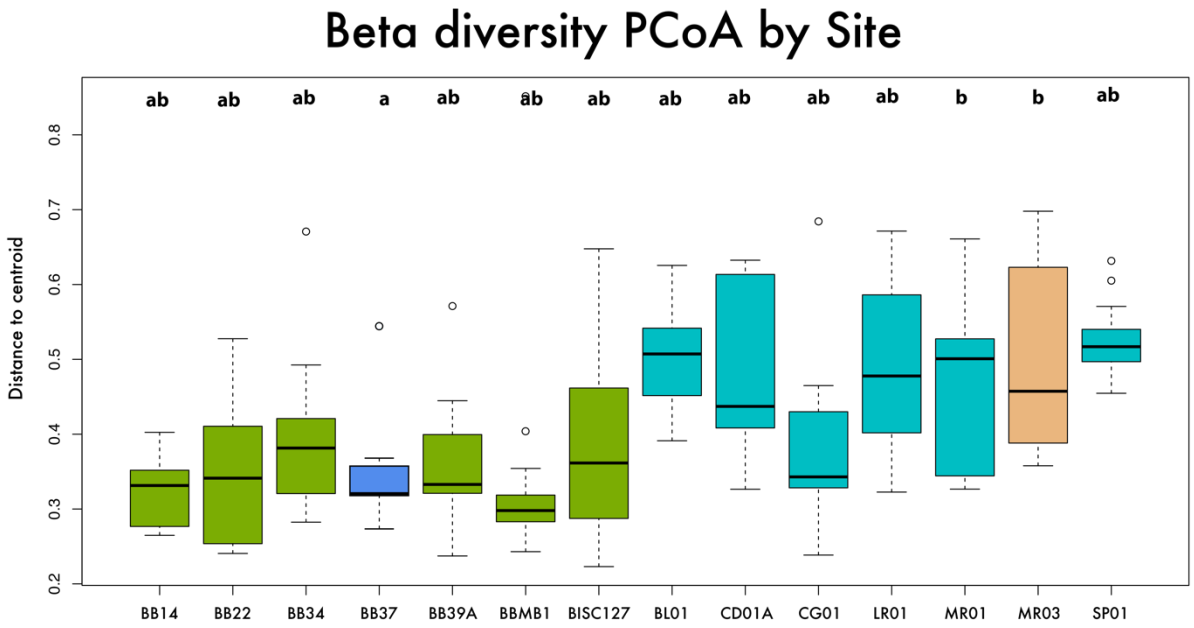


Figure 8: Boxplot generated from a Principal Coordinates Analysis using Bray-Curtis dissimilarity metric for β-diversity. The letters at the top of the chart mark the statistical group each site belongs too. Bacterial β-diversity in the bay followes a similar pattern to the one observed in α-diversity.

A canonical correspondence analysis returned R^2 values for salinity, temperature, percent dissolved oxygen, nitrate/nitrite, and total phosphorus as 0.050, 0.063, 0.073, 0.082, and 0.086 –respectively. For the same test, the Akaike information criterion (AIC) –to measure goodness of fit– returned as 323.69, 322.73, 322.14, 321.67, and 322.00 (Fgure 9). Salinity had the highest AIC score, and total phosphorius had the highest R^2 value. The salinity curves for all site types followed the same general pattern (Figure 10). Salinity at the ocean influenced site was the most stable, averaging around 34. The more confined body of water generally indicated higher variability in salinity. For example, at the canal site, the salinity was the most varable –ranging from 2-25. Looking at total phosphorus, the same patern of open waters being more stable and more confined waters being more varable is seen (Figure 11). A sharp spike in total phosphorus, acompinied by a decline in salinity, was observed in December.

Taxonomic data were transormed for rank abundance. The top 20 most abundent taxa for each site type were visualized with stacked bar grahs and hirearcheral charts. The stacked bar graphs were generated for Order (Figure 12) and Family level (Appendix 3) taxonomy. The

lowest identifyable taxa was used on the hirearcheral chart. In most caseses this was Family level (Appendix 4).

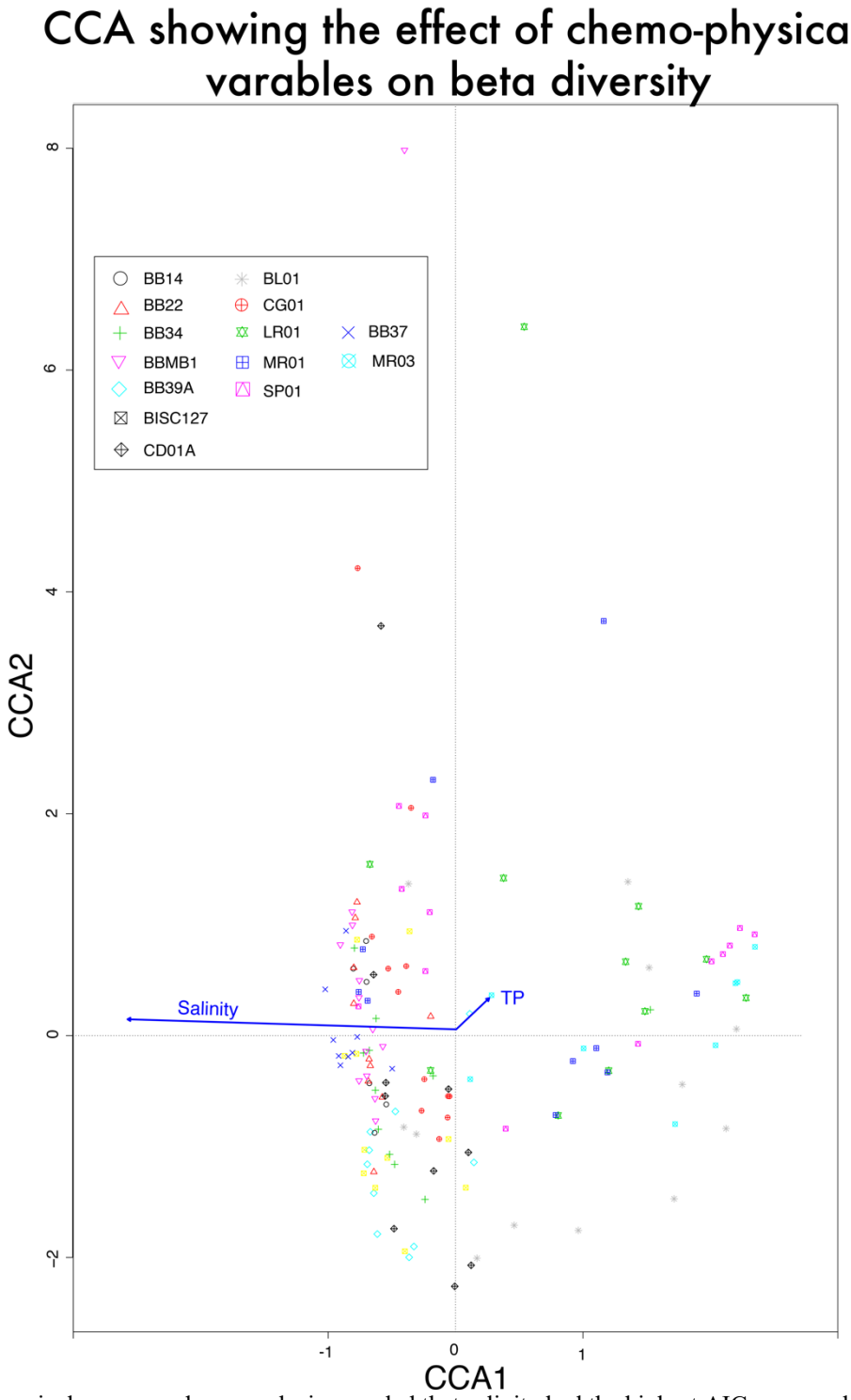


Figure 9: A canonical correspondence analysis revealed that salinity had the highest AIC score and Total Phosphate (TP) had the highest R^2 value. Salinity likely drives the separation between points on the x-axis and TP likely drives the separation a between points on the y-axis.

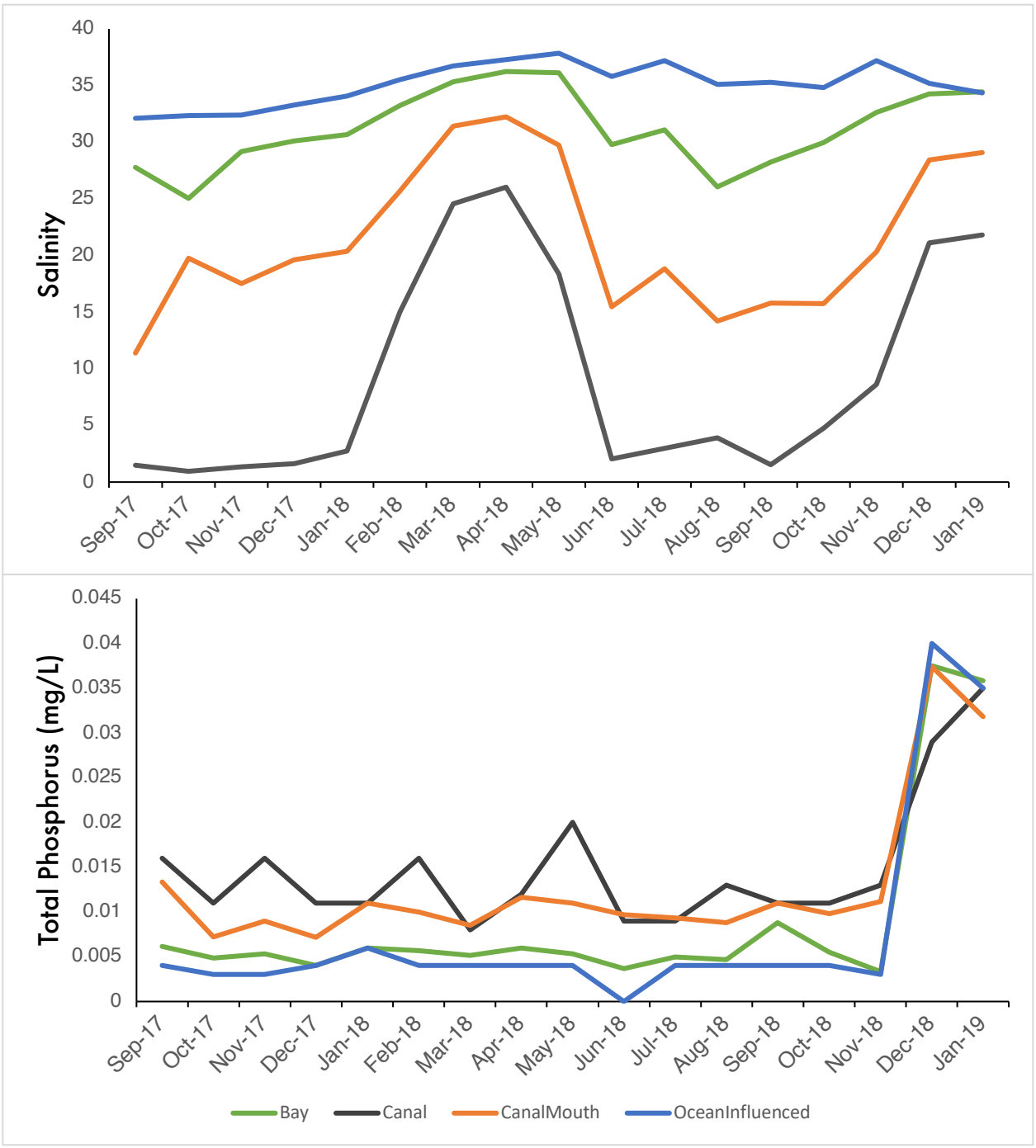


Figure 10 (top): Line graph showing the average salinity for each site type. In a CCA looking at the effect of various chemo-physical parameters on β-diversity, salinity had the highest AIC score (323.69). Salinity was most stable for the ocean influenced site (BB37). While the upstream canal site (MR03) was the most variable. The Bay and Canal Mouth values fell in between those two extremes with the canal mouths having slightly less salinity and slightly more variability than the Bay sites. **Figure 11 (bottom):** Line graph showing the average Total Phosphate (TP) for each site type. In a CCA looking at the effect of various chemo-physical parameters on β-diversity, TP had the highest R^2 value. Note TP followed the same pattern as described above: with more confined waters having more variability and more open waters having less variability.

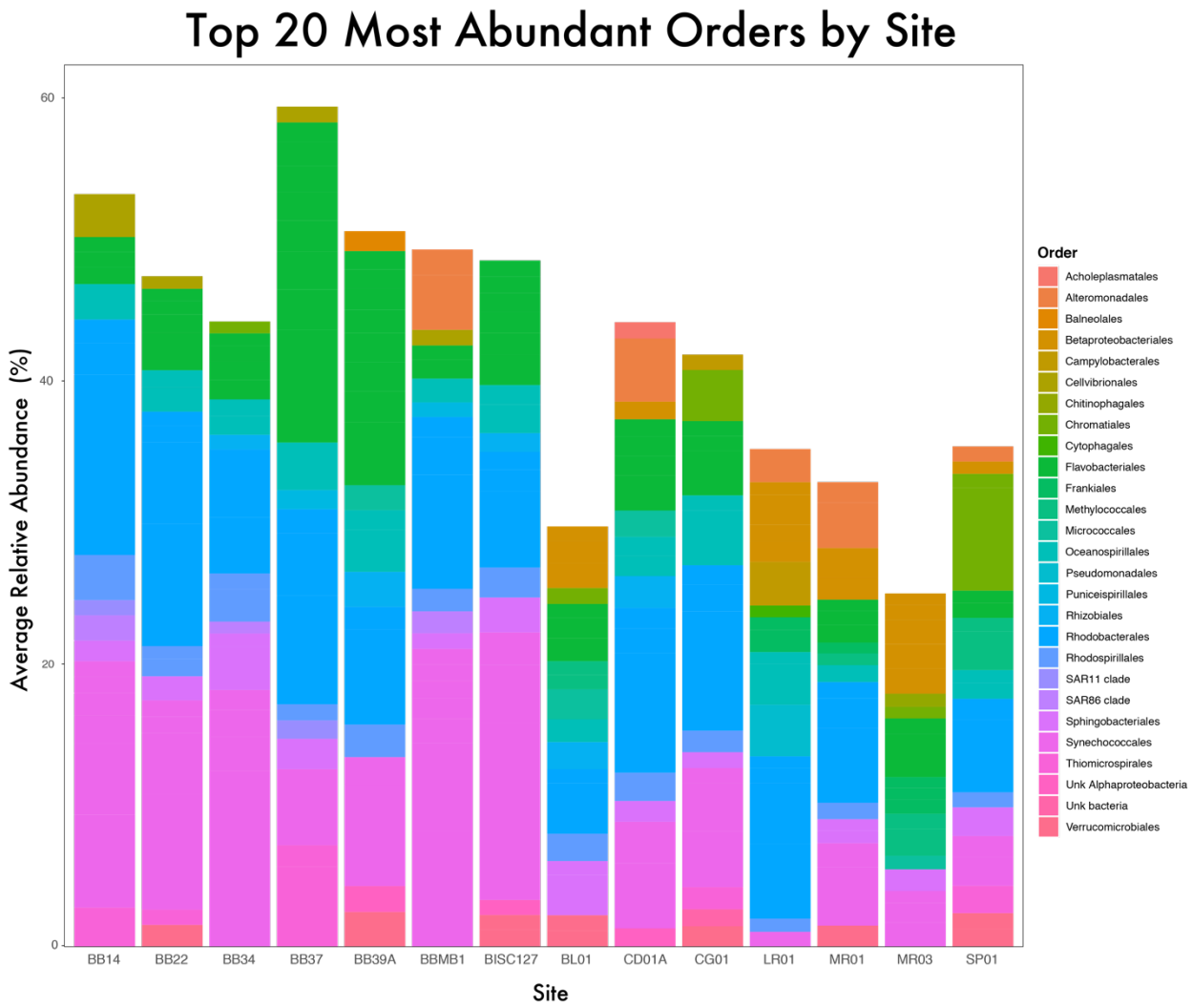


Figure 12: Stacked bar chart showing the average relative abundance of the top 20 most abundant taxonomic Orders at each station.

Discussion

Overall, α and β-diversity microbiomes were fairly homogeneous across a majority of the study area. In regard to α-diversity, we found three groups of sites. The first group consists of the canal site and two of the canal mouth sites: MR01, MR03 and BL01, which have statistically the same α-diversity. These three sites are distinct from BB37, the apparently most oceanic influenced site in its own group with relating to α-diversity. The remaining 11 sites have statistically identical α-diversity. Regarding β-diversity the sites once again could be organized into three groups, based on statistical significance. The first group consisted of MR01 and MR03.

The second group was the oceanic influenced site BB37. The remining 11 sites were statistically identical to each other, in regard to β-diversity, meaning they have proportionally the same amount of unique taxa present. As predicted, the ocean influenced site (BB37) had the lowest α- and β-diversity (Figure 5 & 8). This is likely because of the relatively stable conditions at the site, as it is regularly flushed with oceanic water. The mouth of the Little River had the most variability in α and β-diversity (Figure 5 & 8).

Beta diversity (Figure 8) at the two Miami River sites (MR01 & MR02) were identical. The Miami River is the most urban and industrialized river in the study; therefore it stands to reason it would be highly influenced by these land uses. Site BB37 is the most seaward site and it is regularly flushed with oceanic water. Therefore, its reasonable for this site to be an outlier because it would be less influenced by land. The β-diversity at the remaining sites possess traits of a combination of the Miami River and oceanic site. Biscayne bay is regularly flushed with semi-diurnal mixed tides. This mixing combined with the less urbanized land use, outside of Miami's urban core, probably accounts for the patterns observed in this study.

A canonical correspondence analysis revealed that salinity and total phosphorous had the greatest impact on β-diversity. Salinity drove most of the horizontal separation between the samples, and total phosphorous drove more of the vertical separation between the samples – along the axes. The oceanic influenced site (BB37) had the most stable α and β-diversity. Salinity and total phosphorous were also most stable at this site. The increased variability in the diversity metrics at the other stations is attributed to the increased variability of these abiotic factors as well. It should be noted that while salinity and phosphorus significantly affected bacterial community, the strength of the effect was not particularly strong. O'Connell et al. (2018) determined that salinity and temperature were the main factors driving bacterial community. This study supports the finding that salinity significantly affects bacterial community, but temperature did not seem to play as important role in determining bacterial community.

Abundant & distinguishing taxa

An analysis of similarity (SIMPER) was used to determine which taxa were responsible for distinguishing the sites from each other. There were no Archaea in the top 20 most abundant taxa for each site. Likewise, no Archaea appeared in the SIMPER analysis either. Simper analysis

revealed the main taxa responsible for the difference between the oceanic and canal mouth sites were Cryomorphaceae, Rhodobacteraceae, SUP05 cluster, and the NS5 marine group . All these taxa were more abundant at the oceanic site, indicating they are marine taxa. Rocca et al. (2019) suggests that marine taxa may be more resilient in brackish conditions. The preponderance of marine taxa in estuarine conditions in this study supports that finding. The family Cryomophaceae is non-monophyletic (Bowman, 2014). Its members are generally secondary producers and inhabit locations relatively rich in organic carbon (Bowman, 2014). The family Rhodobacteraceae are a common family of bacteria in marine environments (Simon et al, 2017). All species in the family are obligate aerobic, chemoheterotrophs (Rosenberg, 2014). Many marine members of the family use aerobic anoxygenic photosynthesis –meaning they use light to produce ATP, but the process does not result in the release of O_2 (Simon et al, 2017). The five species within the family Rubritaleaceae are not distinguishable based on 16S analysis alone (Rosenberg, 2014). Bacteria from the SUP05 cluster seem to play an important role in the nitrogen and sulphur cycles (Shah et al, 2017). Members of the NS5 marine group are heterotrophs associated with phytoplankton blooms (Seo et al, 2017). NS5 marine group members possess enzymes for catalyzing many phytoplankton-derived macromolecules (Seo et al, 2017). Overall, the most abundant taxa fit in niches responsible for carrying basic nutrient cycling processes you would expect to find in a marine habitat.

Cyanobacteria are a major, diverse group of photosynthetic bacteria that can inhabit freshwater and a wide range of salinities (Cohen & Gurevitz, 2006). Cyanobacteria were most abundant at the bay sites and least abundant at the oceanic influenced site. Cyanobacteria species can function as aerobic photoautotrophs; anaerobic photo-autotrophs; photoheterotrophs; or chemoheterotrophs (Cohen & Gurevitz, 2006). Many cyanobacteria are known to be N_2 fixers (Arrigo, 2005). In some primarily oligotrophic waters, their contributions to available nitrogen is significant; while in other areas their contribution to N_2 fixation is quite low (Arrigo, 2005). Because the resolution of taxa identified in this study is largely limited to Family level, it is difficult to identify the implications of the presence of various cyanobacteria in the samples. Through SIMPER analysis Cyanobiaceae ASV40 and Cyanobiaceae ASV4 were identified as being a distinguishing taxa between bay sites and the canal site. Cyanobiaceae ASV40 and Cyanobiaceae ASV4 were found predominantly at more saline sites, so presumably they

represent saltwater tolerant taxa. *Fluviicola spp.* (in the Family Cryomorphaceae) was also identified through SIMPER analysis as being a distinguishing taxa between The canal (MR03) and oceanic (BB37) sites. The name *Fluviicola* translates as "river dweller" (Woyke et al., 2011). So perhaps unsurprisingly *Fluviicola*it was completely absent from the oceanic site, and present in relatively large numbers at the canal site. Other members of the genus are known to be predominantly fresh water bacteria (O'Sullivan et al., 2005 & Yang et al., 2014). *Fluviicola spp.* was found in moderate abundance at the bay and canal mouth sites, supporting the idea that those sites are influenced by a combination of oceanic and fresh water factors (Appendix 7).

Several members of the family Flavobacteriaceae were key in distinguishing the sites from each other. Flavobacteria (family Flavobacteriaceae) are one of the most abundant organisms in aquatic habitats (McBride, 2014). Unsurprisingly flavobacteria were one of the most abundant bacteria observed in this study. While they were still present at fresher sites flavobacteria were much more abundant at more saline sites. No species of flavobacteria are known to be photosynthetic; nearly all species are aerobic chemoorganotrophs (McBride, 2014). Some aquatic flavobacteria are typically not free floating, they rather grow on a surface —i.e. floating organic matter (McBride, 2014). Some are known pathogens for fish (Chen et al, 2017) and possibly sponges (Mulheron, 2014). Typically, flavobacteria are associated with flocculent —as such they are important decomposers in aquatic habitats (McBride, 2014).

Most taxa were not identifiable to species or genus level, however one relatively abundant taxon was *Shewanella frigidimarina* (Family Shewanellaceae), which was the 8th most common (relative abundance= 1.5%) bacteria at the mouth of the Miami River (MR01). *S. frigidimarina* is capable of using a wide verity of molecules as an electron acceptor in the electron transport chain of cellular respiration, including: oxygen, iron, manganese, uranium, nitrate, nitrite and fumarate (Copeland et al., 2006). Therefore, it is frequently used in bioremediation (Copeland et al., 2006). At the upstream Miami River site (MR03) the most abundant bacteria (1.9% relative abundance) belong to the family Methylococcaceae. Bacteria in this family are chemoautotrophs that metabolize methane (Bowman, 2014). Methylococcaceae are obligate methane and methanol metabolizers. These molecules are their only carbon and energy source as they are unable to use other substrates containing carbon-carbon bonds (Bowman, 2014). These methane loving bacteria play a critical roll in carbon cycling and Earth's homeostatic processes (Bowman, 2014).

Methylococcaceae have also been used in bioremediation applications, because of their ability to sequester large amounts of methane (Bowman, 2014). Donnelly (2018) observed Methylococcaceae in high abundance in urban canals in urban Ft. Lauderdale, FL. While both *S. frigidimarina* and Methylococcaceae are beneficial, their presence in high abundance suggests the location is highly polluted. The Miami River is the most urbanized river in the study, therefor finding bacteria which exploit heavy metals and methane is not surprising.

Bacteria in the family Enterobacteriaceae can be used as an indicator of anthropogenic pollution (Leite et al., 2018). Of the 116 samples that had detectible *Enterococci* (through traditional culture methods) only 11 of those samples had detectable Enterobacteriaceae through 16S analysis. Of those 11 samples 3 were under the EPA limit of 20 MPN per 100mL for *Enterococci* (US EPA, 2012). The discrepancy between culture methods and 16S analysis can likely be attributed to holding time. The holding time for the 16S samples ranged from 24hrs – 120hrs, and likely exceeded the EPA's maximum holding time of 30hrs (US EPA, 1982). *Enterococci* blooms from rain events are typically short lived and sampling strategies should have high temporal resolution, to adequately detect presence of the bacteria (Aranda et al., 2016).

Currents & Hydrology

Water transport in the bay is principally tidal influenced (Wang, et al., 2003). However small subtidal currents that are not easily measured, strongly influence residence time of water (Wang, et al., 2003). Wind over the shallow bay follow two distinct seasonal patterns (Wang, et al., 2003). Prevailing winds in the summer are gentle Southeasterlies. In the winter winds are generally Southeasterly, but stronger, and they are occasionally interrupted by clockwise rotating winds associated with passing cold fronts (Wang, et al., 2003). Tides in the bay are mixed-semi-diurnal; having two high tides and two low tides each day —with the two highs being of unequal zenith and the two lows of unequal nadir (Smith, 2001). The tidal range in Biscayne Bay is well below 1.0m (Smith, 2001). As one moves south in the bay, tidal range decreases (Wang, et al., 2003). The region of the bay north of Key Biscayne sees less exchange with the ocean. Region of the bay between Key Biscayne and the Ragged keys is, for the most part, unencumbered by islands and therefore is well flushed with oceanic water.

Precipitation is the dominant source of freshwater to the bay; followed by canal input and ground water discharge (Stalker et. al, 2009). The overall volume of water introduced to the bay varies from the wet to dry seasons, but the ratio of water introduced by these three sources remains constant (Stalker et al., 2009). Historically, the volume of groundwater discharged into the bay was much higher than it is today (Stalker et al., 2009; Cantillo et al., 2000). This is mainly due to anthropogenic alteration of the water table (Stalker et al., 2009; Cantillo et al., 2000). Over the study period Black Creek, Snapper Creek, Miami River, and Little River were each responsible for delivering hundreds of millions of cubic meters of fresh water into the bay. The Cutler Drain and Coral Gables Waterway conducted much less water –on the order of tens of millions of cubic meters (Appendix 11).

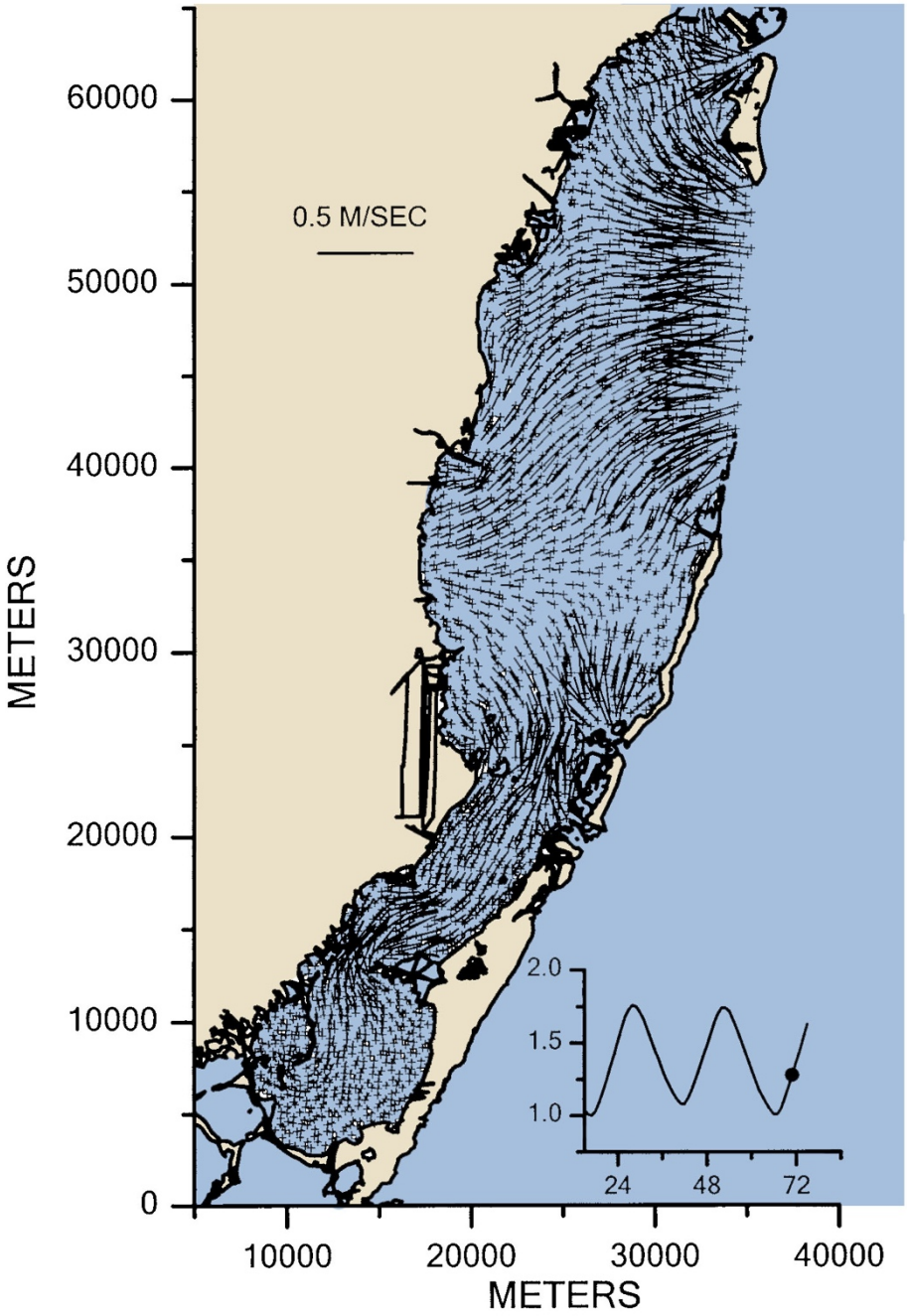


Figure 13: Hydrodynamic model output velocity field for flood conditions. Only one third of the velocity vectors are shown to avoid overcrowding the graph. Each velocity vector is plotted as a stick indicating magnitude and direction. + marks the location of the vector and the velocity scale is indicated in the graph. The inset graph in the lower right shows the depth variation at one point in the model and • indicates the time of the velocity field. Colorized graphic from Wang et al, 2003.

Weather Events

Rain fall data (Appendix 1) was initially included in analysis, but it was removed, so as not to over fit the model. Rain fall was not a better explanatory variable than any of the others. Further, it was assumed that salinity is a suitable proxy for rain fall. Canal flow (Appendix 2) was also considered, but again it would directly affect salinity, meaning salinity should be a suitable proxy. The overall trends in flow were consistent across water control structures, i.e. when flow was high at one, flow was high at all.

Hurricane Irma passed through Biscayne Bay September 10th through the 11th, 2017 (NWS, 2017). While we have microbial community data from September 2017, just after the storm, unfortunately we do not have any pre-storm data, nor data from October and November 2017. Water quality seems to return to "normal" with-in three months of a hurricane (Zhang et al., 2009). This held true for hurricane Irma with the impact lasting less than months (Wachnicka et al., 2019). This makes the effects of Hurricane Irma on the microbial communities in Biscayne Bay, difficult to discern in our data set. In 2017 Biscayne Bay received a record setting inflow of fresh water, the highest in a decade and 26% more than in 2016 (Wachnicka et al., 2019). Because salinity is such an important factor determining microbial community, it is likely that such an influx of fresh water would greatly affect bacterial community assemblage. Outflow through the downstream most water control structure on each canal, was considered in the CCA analysis. But it was later excluded to avoid over fitting the model, as it was not a better predictor variable than any of the other variables considered.

Significance

To date, this is the largest scale microbiome project conducted in Biscayne Bay. Other microbiome research projects in the bay have focused on relatively small regions with in the bay. There is a large gap in our understanding of bacterial community structure and biogeography. The Earth Microbiome Project was founded in 2010 by Knight et al., with the lofty goal of sequencing all microbial life on Earth (Thompson et al., 2017). These kind of base line data are just as important to Ecology –as the five vital signs are to a physician. Building a database of microbial communities will allow us to better understand what a "normal" or "healthy" community looks like. Eventually microbial biodiversity data will help guide management

decisions as much as macro flora/fauna biodiversity do today. Ongoing technological developments are making genetic sequencing increasingly cost efficient (Mardis, 2008). Therefore, genetic analysis of microbial communities may soon become part of the typical suite of water quality parameters resource managers use to make informed decisions.

This study describes patterns of microbial diversity and relative abundance in Biscayne Bay, and is the first of its kind in this area. The interaction of saline oceanic water with freshwater appears to be a major controlling factor of bacterial community. Freshwater bacterial communities exposed to brackish salinities suffer a 96% taxa loss (Rocca et al., 2019). Marine bacterial communities exposed to brackish salinities suffer a 66% taxa loss (Rocca et al., 2019). Biotic interactions between fresh water and marine communities result in another 29% loss from freshwater communities and a 49% loss from marine communities (Rocca et al., 2019).

Because α and β-diversity of planktonic bacteria are so homogeneous across the bay, planktonic bacteria may not be the best metric for making site specific management decisions. Wickes (2018), as well, found α-diversity to be homogeneous across their study sites in northern Biscayne Bay. However, Wickets (2018) did find significant differences in β-diversity across their study sites. It is worth investigating if the microbiome of sediments is more indicative of conditions at a specific site. Mustafa et al. (2016) used interstitial bacteria to describe the impact of pollution a several sites in the Red Sea. Leite et al. (2018) and O'Connell et al. (2017) describe seasonal variation in bacterial community between the wet and dry season. This study found that rain had a minor effect on microbial community, but salinity was a better predictor.

Works Cited

Aranda D, Lopez JV, Solo-Gabriele HM, and Fleisher JM. (2016) Using probabilities of enterococci exceedance and logistic regression to evaluate long term weekly beach monitoring data. *Journal of Water and Health* 14(1) 81-89

Arrigo KR. (2005). Marine microorganisms and global nutrient cycles. *Nature*. 437(15) 349-355. doi.org/10.1038/nature04159

Ault JS, Smith SG, Tilmant JT (2007) Fishery Management Analysis for Reef Fish in Biscayne National Park: Bag and Size Limit Alternative. Natural Resource Technical Report NPS/NRPC/WRD/NRTR—2007/064. National Park Service.

Boyer JN, Kelble CR, Ortner PB, Rudnick DT (2009) Phytoplankton bloom status: Chlorophyll a biomass as an indicator of water quality condition in the southern estuaries of Florida, USA *Ecological Indicators* 9(6): S56-S67. doi.org/10.1016/j.ecolind.2008.11.013

Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, et al. (2018) QIIME 2: Reproducible, interactive, scalable,and extensible microbiome data science. *PeerJ* doi.org/10.7287/peerj.preprints.27295v2

Bowman JP (2014) The Family *Cryomorphaceae*. In: Rosenberg E, DeLong EF, Lory S, Stackebrandt E., Thompson F. (eds). Springer, Berlin, Heidelberg. 539-550 doi.org/10.1007/978-3-642-38954-2_135

Bowman JP (2014) The Family *Methylococcaceae*. In: Rosenberg E, DeLong EF, Lory S, Stackebrandt E., Thompson F. (eds). Springer, Berlin, Heidelberg. 539-550 doi.org/10.1007/978-3-642-38922-1_226

Bunse C, Pinhassi J (2017) Marine Bacterioplankton Seasonal Succession Dynamics. *Trends in Microbiology*. 25(6) doi.org/10.1016/j.tim.2016.12.013

Caccia VG, Boyer JN (2005) Spatial patterning of water quality in Biscayne Bay, Florida as a function of land use and water management. *Marine Pollution Bulletin* Vol. 50 doi.org/10.1016/j.marpolbul.2005.08.002

Caccia VG, Boyer JN (2007) A Nutrient Loading Budget for Biscayne Bay, Florida. *Marine Pollution Bulletin* Vol. 54 doi.org/10.1016/j.marpolbul.2007.02.009

Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ. (2016) DADA2: High resolution sample inference from Illumina amplicon data. *Nature Methods*. 13(7) doi.org/10.1038/nmeth.3869.

Callahan1 BJ, McMurdie PJ, Rosen DADA2: High resolution sample inference from Illumina amplicon data. MJ, Han AW, Johnson AJ. (2016) *Nature Methods*. 13(7) doi.org/10.1038/nmeth.3869.

Cantillo AY, Hale K, Collins E, Pikula L, Caballero R (2000) Biscayne Bay: Environmental history and annotated bibliography. NOAA Technical Memorandum. National Status and Trends Program for Marine Environmental Quality.

Campbell AM, Fleisher J, Sinigalliano C, White JR, Lopez JV. 2015. Dynamics of marine bacterial community diversity of the coastal waters of the reefs, inlets, and wastewater outfalls of Southeast Florida. *Microbiol-Open*. 4(3); 390-408. doi.org/10.1002/mbo3.245.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley, GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R, (2010) QIIME allows analysis of high-throughput community sequence data. *Nature Methods*. 7 (5) doi.org/10.1038/nmeth.f.303

Caporaso JG, Lauber CL, Walters, WA, Berg-Lyons D, Huntley J, Fierer N (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* Vol. 6

Carey RO, Migliaccio KW, Brown MT (2011) Nutrient discharges to Biscayne Bay, Florida: Trends, loads, and a pollutant index. *Science of the Total Environment* Vol. 409 doi.org/10.1016/j.scitotenv.2010.10.029

Chen S, Blom J, Loch TP, Faisal M, & Walker ED. (2017). The emerging fish pathogen Flavobacterium spartansii isolated from chinook salmon: comparative genome analysis and molecular manipulation. *Frontiers in microbiology*. 8(2339). doi.org/10.3389/fmicb.2017.02339

Cohen Y., Gurevitz M. (2006) The Cyanobacteria—Ecology, Physiology and Molecular Genetics. In: Dworkin M., Falkow S., Rosenberg E., Schleifer KH., Stackebrandt E. (eds) The Prokaryotes. Springer, New York, NY. doi.org/10.1007/0-387-30744-3_39

Collado-Vides, Avila, Blair, Leliaert, Rodriguez, Thybergf, Schneider, Rojas, Sweeney, Drury, & Lirman (2013) A persistent bloom of Anadyomene J.V. Lamouroux (Anadyomenaceae, Chlorophyta) in Biscayne Bay, Florida. *Aquatic Botany*.

Copeland A, Lucas S, Lapidus A, Barry K, Detter JC, Glavina del Rio T, Hammon N, Israni S, Dalin E, Tice H, Pitluck S, Fredrickson JK, Kolker E, McCuel LA, DiChristina T, Nealson KH, Newman D, Tiedje JM, Zhou J, Romine MF, Culley DE, Serres M, Chertkov O, Brettin T, Bruce D, Han C, Tapia R, Gilna P, Schmutz J, Larimer F, Land M, Hauser L, Kyrpides N, Mikhailova N, Richardson P *Complete sequence of Shewanella frigidimarina NCIMB 400*. Submitted (AUG-2006) to the EMBL/GenBank/DDBJ databases.

Dame R, Alber M, Allen D, Mallin M, Montague5 C, Lewitus A, Chalmers A, Gardner R, Gilman C, Kjerfve B, Pinckney J, Smith N. (2000) Estuaries of the South Atlantic Coast of North America: Their Geographical Signatures. *Estuaries and Coasts*. 26(6) 793–819. doi.org/10.2307/1352999

Donnelly C. (2018) Microbiota of South Florida Canal and Water Systems: Comparing Pristine and Anthropomorphic Environments using 16S rRNA sequencing. (Masters Thesis)

Idyll CP, et al. (1999) Economically important Marine Organisms in Biscayne Bay. University of Miami Division of Fisherey Sciences internal report.

Illumina. (2016) Understanding Illumina Quality Scores. [Technical publication]

Janda JM, Abbot SL. (2007) 16S rRNA Gene Sequencing for Bacterial Identification in the Diagnostic Laboratory: Pluses, Perils, and Pitfalls *Journal of Clinical Microbiology*. 45(9). doi.org/10.1128/JCM.01228-07

Kembel SW, Cowan PD, Helmus MR, Cornwell WK, Morlon H, Ackerly DD, Blomberg SP, Webb CO. (2010) Picante: R tools for integrating phylogenies and ecology. Bioinformatics 26(11). doi.org/10.1093/bioinformatics/btq166

Kim M, Chun J (2014) 16S rRNA Gene-Based Identification of Bacteria and Archaea using the EzTaxon Server. *Methods in Microbiology* Vol. 41 doi.org/10.1016/bs.mim.2014.08.001

Knight R., Jansson J., Field D., Fierer N., Desai N. et al. 2012. Designing better metagenomic surveys: The role of experimental design and metadata capture in making useful metagenomic datasets for ecology and biotechnology. *Nature Biotech.* 30(6): 513-520. doi.org.10.1038/nbt.2235

Leite DCA, Salles JF, Calderon EN, Castro CB, Bianchini A, Marques JA, vanElsas JD, Peixoto RS. (2018) Coral Bacterial-Core Abundance and Network Complexity as Proxies for Anthropogenic Pollution. *Frontiers in Microbiology* 2018(9). doi.org/10.3389/fmicb.2018.00833

Leynes JB, Cullison D. (1998) Biscayne National Park Historic Resource Study

Lirman D, Cropper W. (2003) The Influence of Salinity on Seagrass Growth, Survivorship, and Distribution within Biscayne Bay, Florida: Field, Experimental, and Modeling Studies. *Estuaries*. 26(1)

Lirman D, Monty J, Avila C, Buck E, Hall MO, Bellmund S, Carlson P, Collado-Vides L. (2016) Seagrass Integrated Mapping and Monitoring Program Mapping and Monitoring. *FWRI Technical Report TR-17.*

Mardis ER. (2008) Next-Generation DNASequencing Methods. *Annu. Rev. Genom. Human Genet*. 2008(9) 387-402 doi.org/10.1146/annurev.genom.9.081307.164359

McBride MJ. (2014) The Family Flavobacteriaceae. In: Rosenberg E., DeLong E.F., Lory S., Stackebrandt E., Thompson F. (eds) The Prokaryotes. Springer, Berlin, Heidelberg. doi.org/10.1007/978-3-642-38954-2_130

McMurdie PJ, Holmes S. (2013) phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLOS one*. 8(4) doi.org/10.1371/journal.pone.0061217

Millette NC, Kelble C, Linhoss A, Ashby S, Visser L. (2017) Shift in baseline chlorophyll a concentration following a three-year Synechococcus bloom in southeastern Florida. *Bulletin of Marine Science*. 94(1). doi.org/10.5343/bms.2017.1046

Mulheron RM. (2014) Microbial Community Assembly found with Sponge Orange Band Disease in *Xestospongia muta* (giant barrel sponge). (Masters Thesis).

Mustafa GA, Abd-Elgawad A, Ouf A, Siam R. (2016) The Egyptian Red Sea coastal microbiome: A study revealing differential microbial responses to diverse anthropogenic pollutants. *Environmental Pollution* 214: 892-902. doi.org/10.1016/j.envpol.2016.04.009

National Weather Service. (2017-2019) Meteorological Data [Electronic Data Set] Weather.gov

O'Connell L, Gao S, Fleischer, Lopez JV. (2018) Fine Grained Compositional Analysis of Port Everglades Inlet Microbiome Using High Throughput DNA Sequencing. *PeerJ* 6:e4671. doi.org/10.7717/peerj.4671

O'Sullivan LA, Rinna J, Humphreys G, Weightman AJ, Fry JC. (2005) Fluviicola taffensisgen. nov., sp. nov., a novelfreshwater bacterium of the family Cryomorphaceaein the phylum 'Bacteroidetes'. *International Journal of Systematic and Evolutionary Microbiology*. 55(5), 2189–2194. doi.org/10.1099/ijs.0.63736-0

Pace NR. (1997) A molecular view of microbial diversity and the biosphere. *Science*. 276(5313) 734-40. doi.org/ 10.1126/science.276.5313.734

Rocca JD., Simonin M, Wright J, Washburne A, Bernhardt ES. (2019). Rare Microbial Taxa Emerge When Communities Collide: Freshwater and Marine Microbiome Responses to Experimental Seawater Intrusion. *BioRxiv* [Preprint]. dx.doi.org/10.1101/550756

Rösel, S., and H. P. Grossart. 2012. Contrasting dynamics in activity and community composition of free-living and particle-associated bacteria in spring. *Aquat. Microb. Ecol*. 66: 169–181. doi.org/10.3354/ame01568

Rosenberg E. (2014) The Family *Rubritaleaceae*. In: Rosenberg E., DeLong E.F., Lory S., Stackebrandt E., Thompson F. (eds) The Prokaryotes. Springer, Berlin, Heidelberg. doi.org/10.1007/978-3-642-38954-2_146

Seo J-H, KangI, YangS-J, Cho J-C. (2017) Characterization of spatial distribution of the bacterial community in the South Sea of Korea. *PLoS ONE* 12(3). doi.org/10.1371/journal.pone.0174159

Shah V, Chang BX, Morris RM (2017) Cultivation of a chemoautotroph from the SUP05 clade of marine bacteria that produces nitrite and consumes ammonium. *The ISME Journal* 11: 263-271. doi:10.1038/ismej.2016.87

Simon M, Scheuner C, Meier-Kolthoff JP, Brinkhoff T, Wagner-Döbler I, Ulbrich M, Klenk HP, Schomburg D, Petersen J, Göker M. (2017) Phylogenomics of Rhodobacteraceae reveals evolutionary adaptation to marine and non-marine habitats. *The ISME Journal* 2017(11): 1483–1499. doi.org/10.1038/ismej.2016.198

Smith DC, Steward GF, Long RA, Azam F. (1995) Bacterial mediation of carbon fluxes during a diatom bloom in a mesocosm. *Deep-Sea Research*. 2(42), 75–97. doi.org/10.1016/0967-0645(95)00005-B

Smith NP. (2001) Tides of Biscayne Bay, Card Sound, Barnes Sound, and Manatee Bay, Florida. *Florida Scientist*. 64(3) 224-236. jstor.org/stable/24321024

Tamura T. (2014) The Family *Sporichthyaceae*. In: Rosenberg E., DeLong E.F., Lory S., Stackebrandt E., Thompson F. (eds) The Prokaryotes. Springer, Berlin, Heidelberg. doi.org/10.1007/978-3-642-30138-4_182

Tang, X., L. Li, K. Shao, B. Wang, X. Cai, L. Zhang, J. Chao, and G. Gao. 2015. Pyrosequencing analysis of free-living and attached bacterial communities in Meiliang Bay, Lake Taihu, a large eutrophic shallow lake in China. *Can. J. Microbiol*. 61: 22–31. doi.org/10.1139/cjm-2014-0503

Thompson LR, Sanders JG, McDonald D, Amir A, et al. (2017) A communal catalogue reveals Earth's multiscale microbial diversity. *Nature*. 551(7681): 457-463. doi.org/10.1038/nature24621

US Environmental Protection Agency, Environmental Monitoring and Support Laboratory. (1982) *Handbook for Sampling and Sample Preservation of Water and Wastewater*. NEPIS.EPA.gov/Exe/ZyPURL.cgi?Dockey=30000QSA.txt

US Environmental Protection Agency. (2012) *Recreational Water Quality Criteria*. EPA.gov/sites/production/files/2015-10/documents/rwqc2012.pdf

Wachnicka A, Browder J, Jackson T, Louda W, Kelble K, Abdelrahman O, Stabenau E, Avila C. (2019) Hurricane Irma's Impact on Water Quality and Phytoplankton Communities in Biscayne Bay. *Estuaries and Coasts*. doi.org/10.1007/s12237-019-00592-4

Wang JD, Luo J, Ault JS. (2003) Flows, Salinity, and Some Implications for Larval Transport in South Biscayne Bay, Florida. *Bulletin of Marine Science*. 72(3) 695–723.

Whittaker, R. H. (1972). Evolution and Measurement of Species Diversity. *Taxon*. 21, 213-251

Wickes M. (2018) Microbiome Diversity of Coastal Tidal Floodwater in Southeastern Florida. (Masters Thesis)

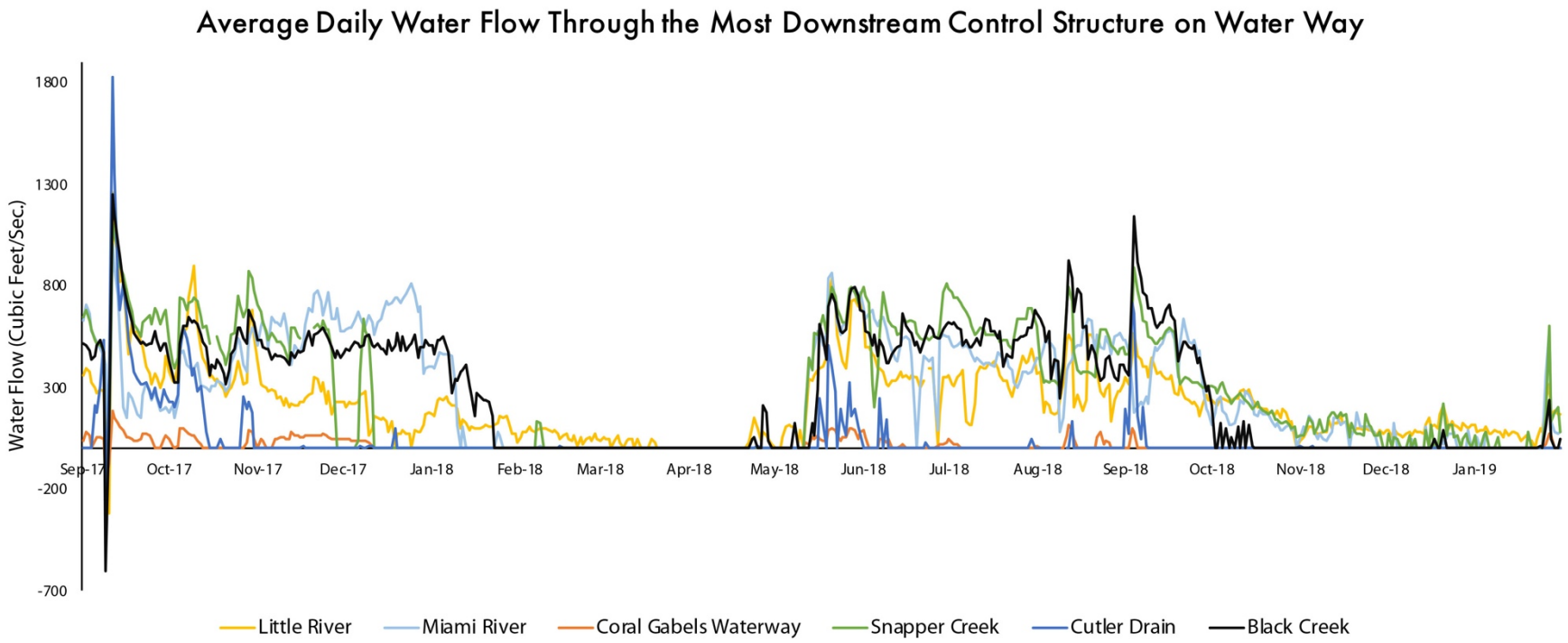
Woese CR, Fox GE. (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *PNAS*. 74(11) 5088-90. doi.org/10.1073/pnas.74.11.5088

Woyke T, Chertkov O, Lapidus A, Nolan M, Lucas S, Del Rio TG, Tice H, Cheng JF, Tapia R, Han C, Goodwin L, Pitluck S, Liolios K, Pagani I, Ivanova N, Huntemann M, Mavromatis K, Mikhailova N, Pati A, Chen A, Palaniappan K, Land M, Hauser L, Brambilla EM, Rohde M, Mwirichia R, Sikorski J, Tindall BJ, Göker M, Bristow J, Eisen JA, Markowitz V, Hugenholtz P, Klenk HP, Kyrpides NC. (2011) Complete genome sequence of the gliding freshwater bacterium *Fluviicola taffensis* type strain (RW262). *Standards in Genomic Sciences*. 5(1) 21-9. doi.org/10.4056/sigs.2124912

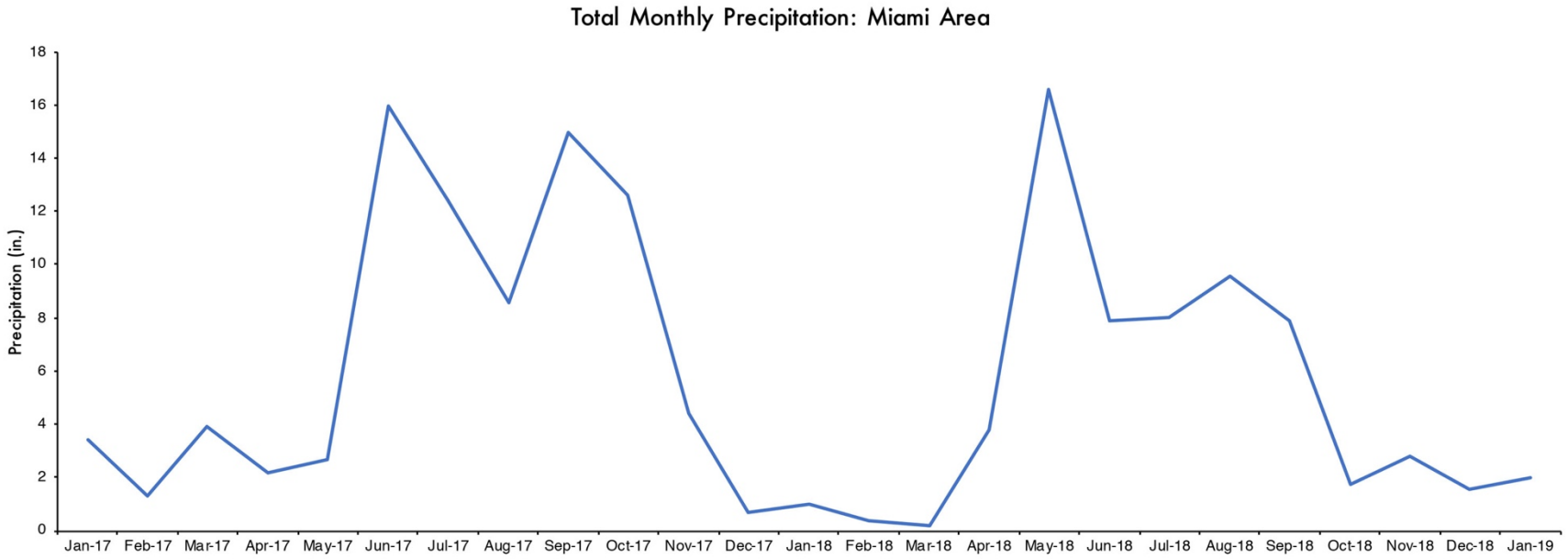
Yang HX, Wang X, Liu XW, Zhang J, Yang GQ, Lau KWK, Li SP, Jiang JD. (2014) Fluviicola hefeinensis sp. nov., isolated from the waste water of a chemical factory. *International Journal of Systematic and Evolutionary Microbiology*. 64(3), 700-704. doi.org/10.1099/ijs.0.056705-0

Zhang J, Kelble CR, Fisher CJ, Moore L (2009) Hurricane Katrina induced nutrient runoff from an agricultural area to coastal waters in Biscayne Bay, Florida. *Estuarine, Coastal and Shelf Science*. 84(2): 209-218. doi.org/10.1016/j.ecss.2009.06.026

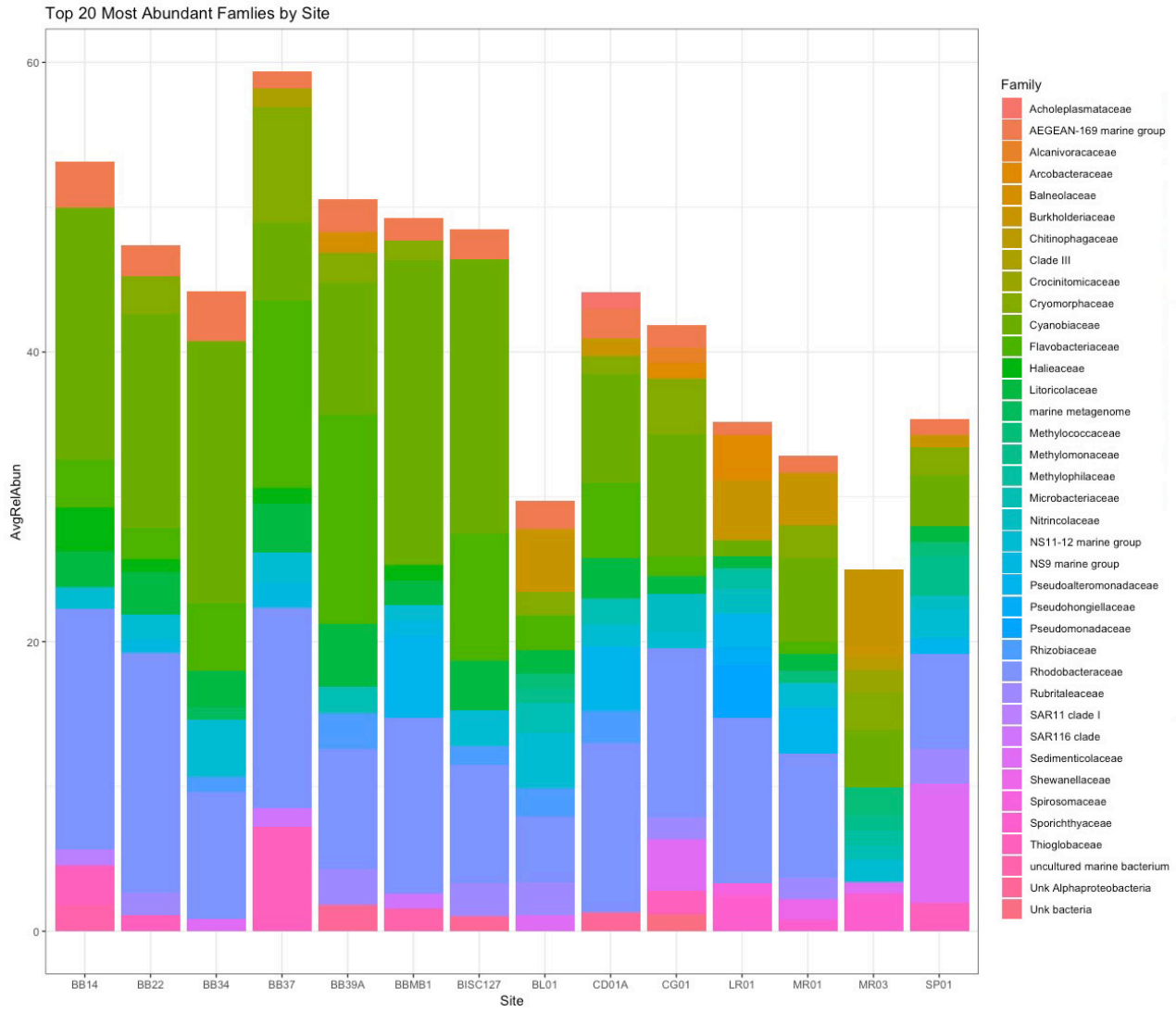
Appendix



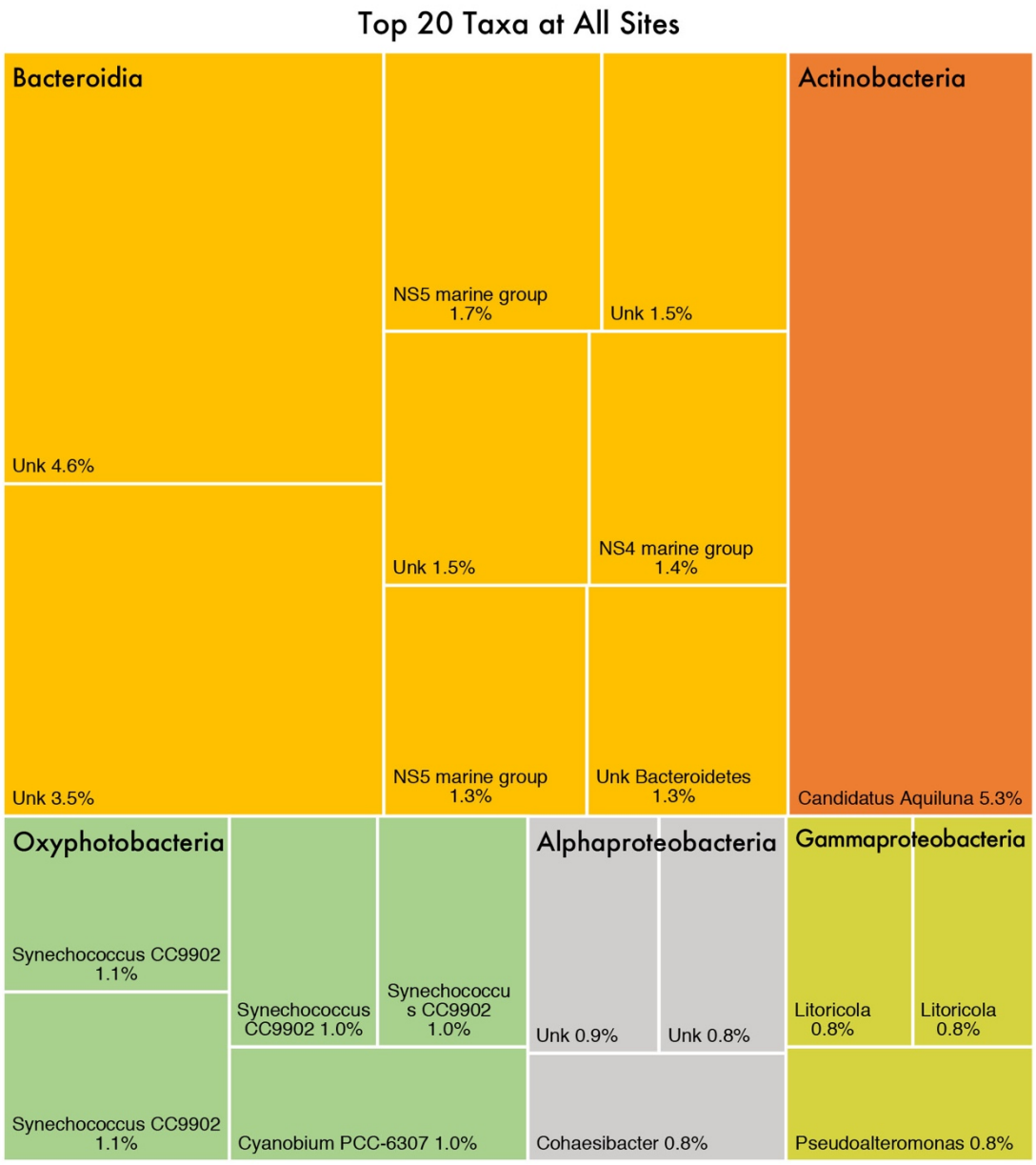
Appendix 1: Line graph showing the average daily flow rate of the most downstream water control structure on each waterway. The negative dip in September 2017 is inundation form Hurricane Irma. Data from: South Florida Water Management District 2017-2019.



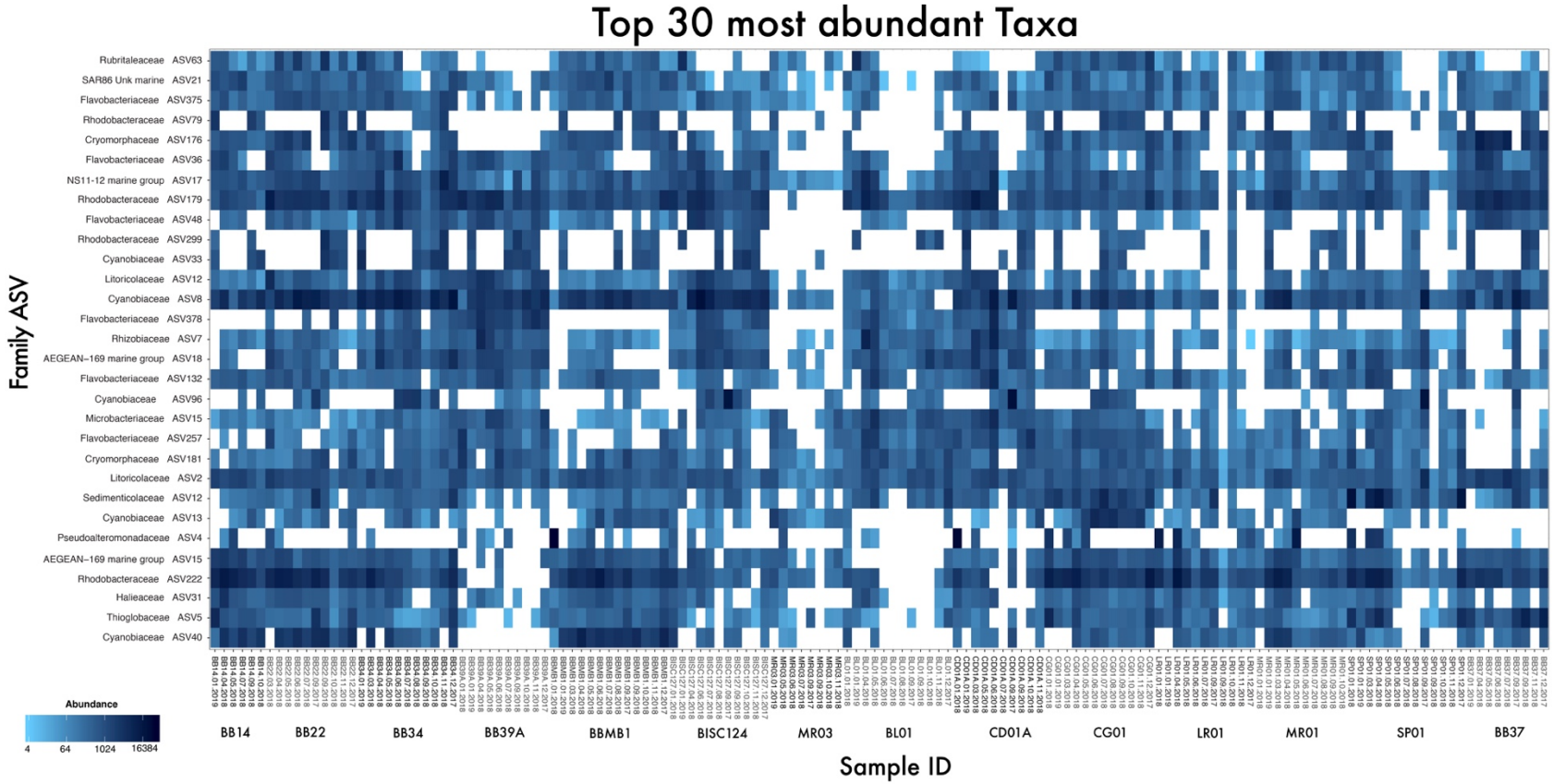
Appendix 2: Shows the total precipitation for each month of the study period. Note the typical Wet, Dry seasons typical of South Florida. Data from the National Weather Service, 2017-2019.



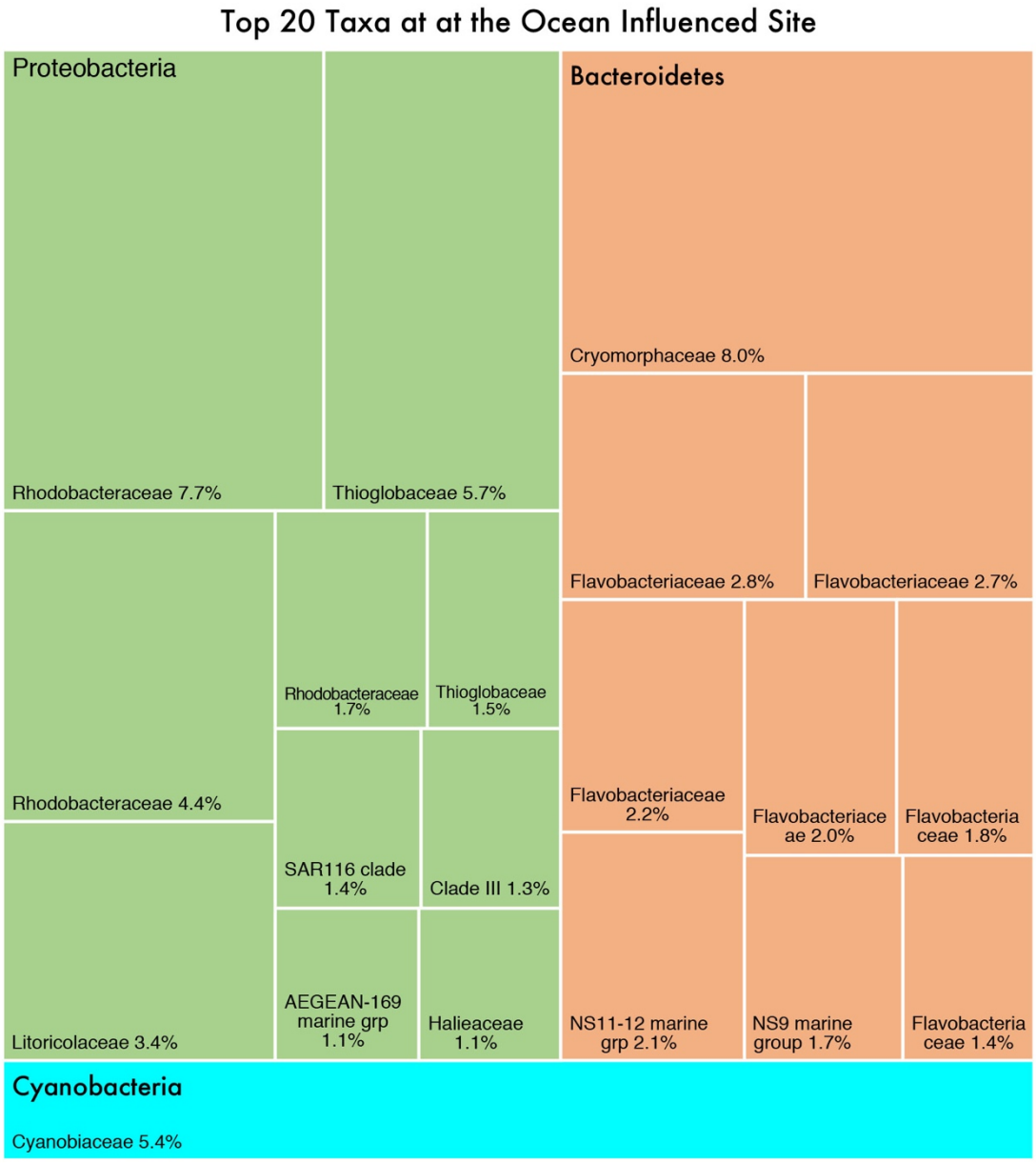
Appendix 3: Stacked bar graph showing the top 20 most abundant Families at each station.



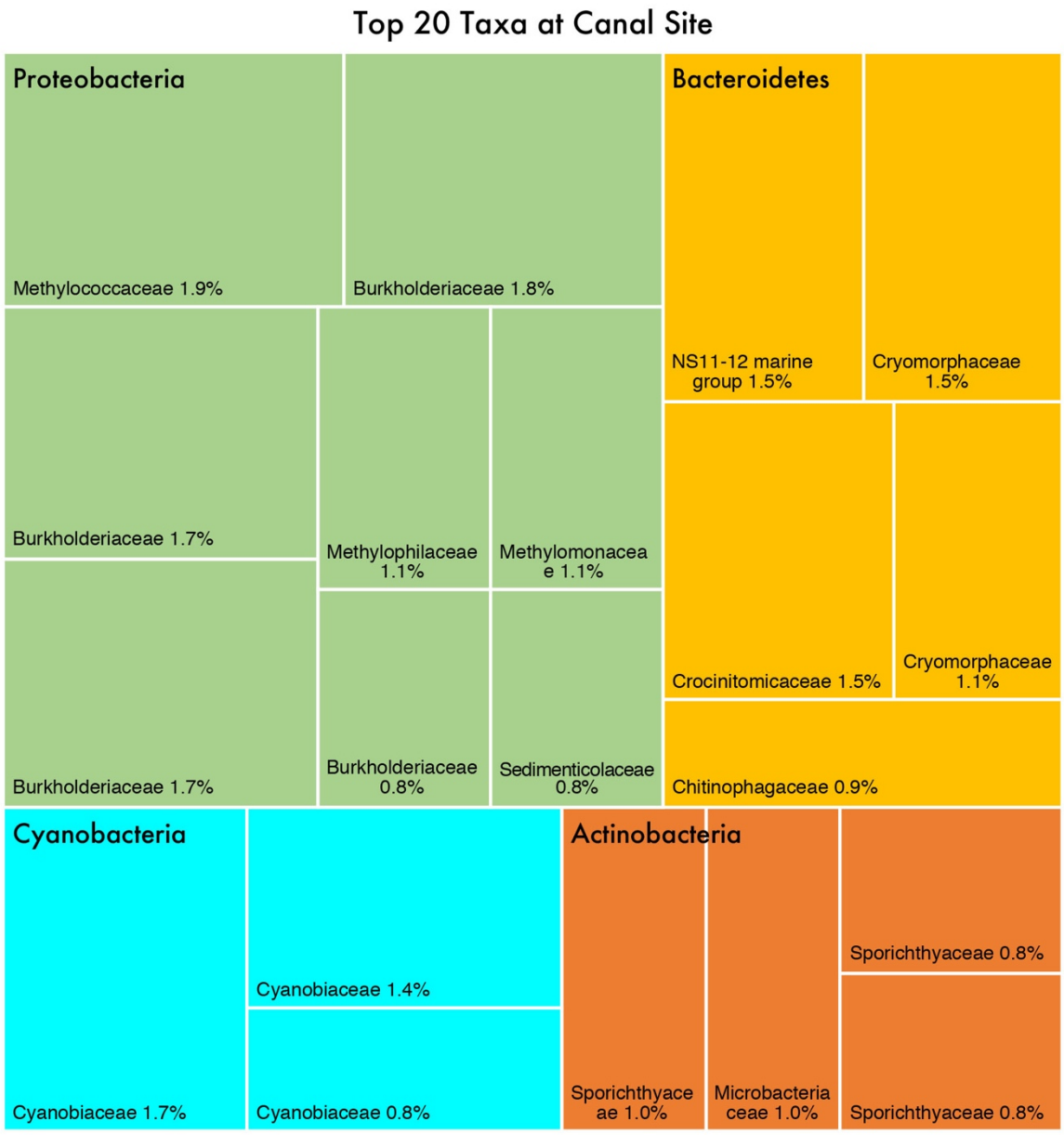
Appendix 4: Stacked pie chart showing the relative abundance for the top 20 most abundant taxa, at Phylum (large text) and Family (small text) level for all stations.



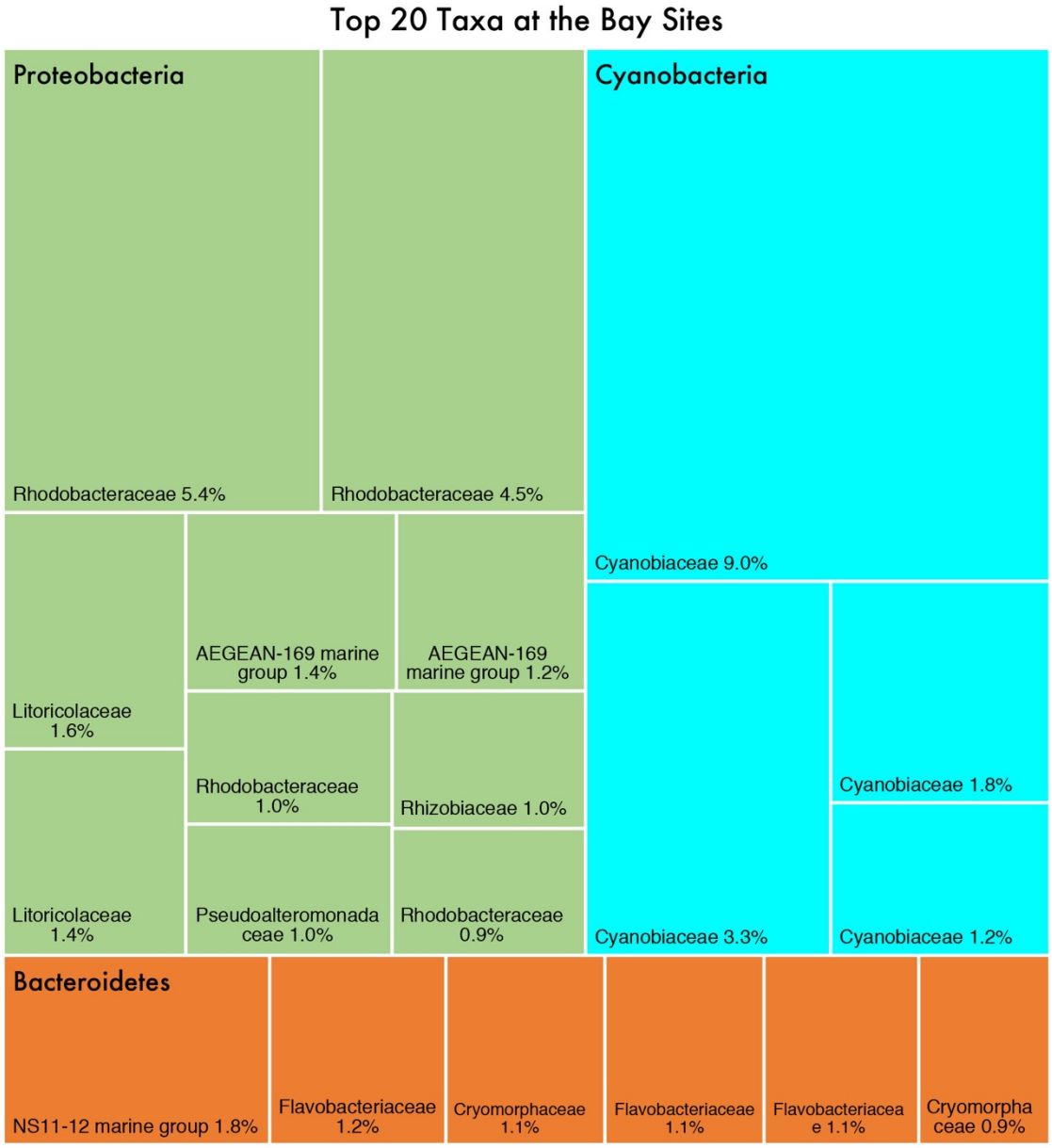
Appendix 5: Heatmap showing the 30 most abundant Families for each sampling event. Family level taxonomy is shown on the y-axis. The ASV number is an arbitrary serial number for distinguishing the taxa from other members of the Family. Lighter colors indicate lower abundance darker colors indicate higher abundance.



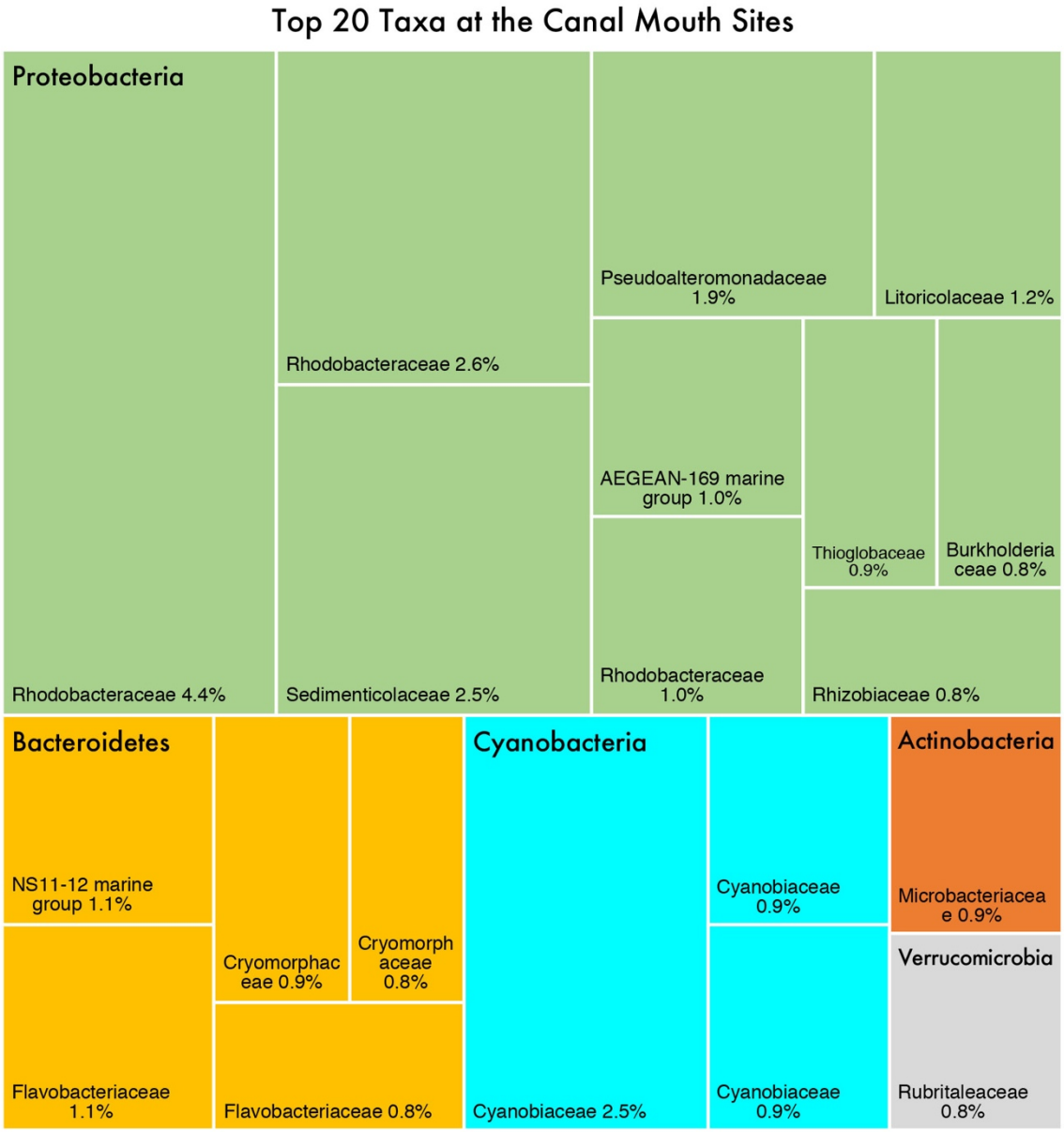
Appendix 6: Stacked pie chart showing the relative abundance for the top 20 most abundant taxa, at Phylum (large text) and Family (small text) level at the ocean influenced site (BB37).



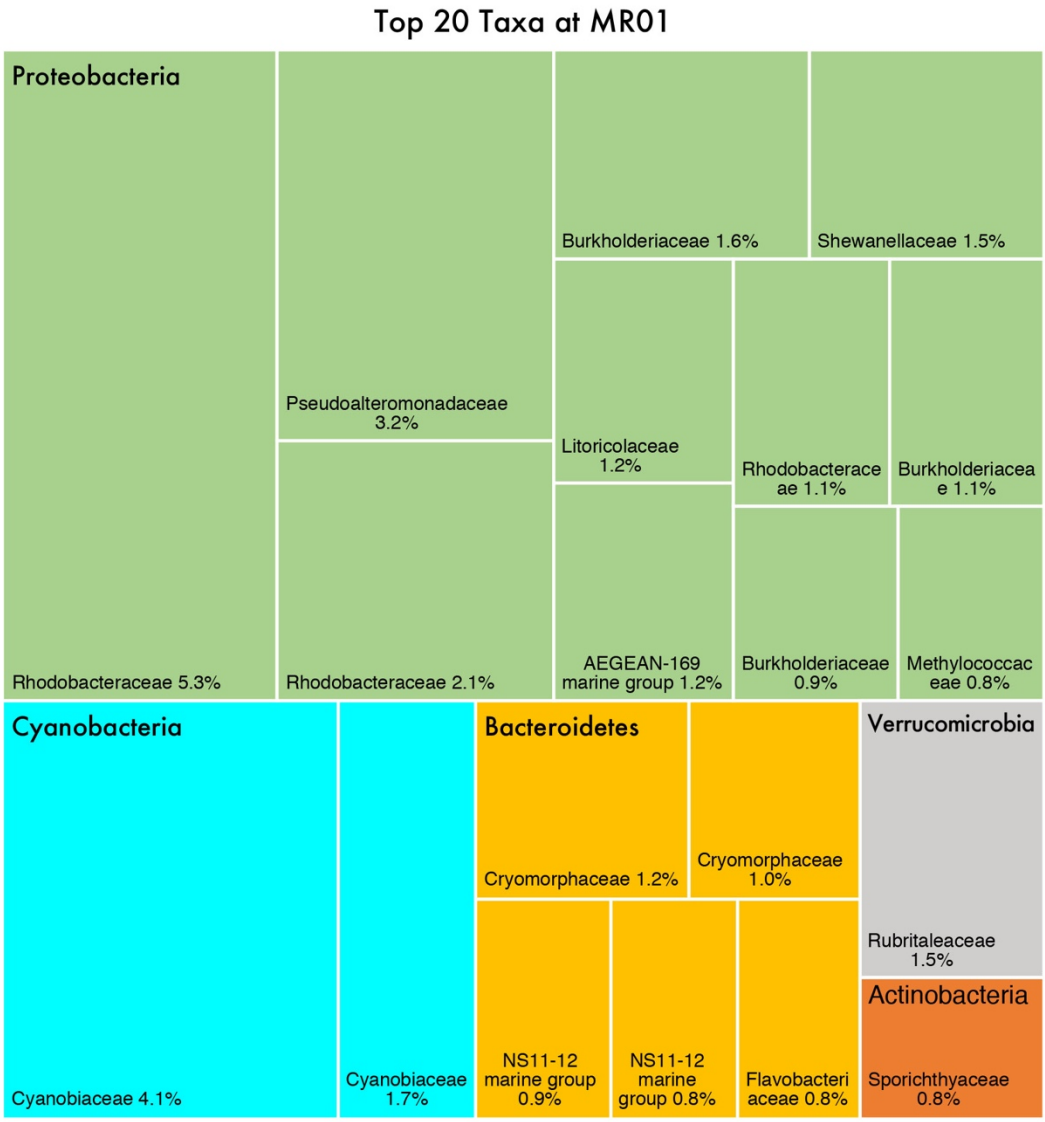
Appendix 7: Stacked pie chart showing the relative abundance for the top 20 most abundant taxa, at Phylum (large text) and Family (small text) level at the canal site (MR03).



Appendix 8: Stacked pie chart showing the relative abundance for the top 20 most abundant taxa, at Phylum (large text) and Family (small text) level at the bay sites.



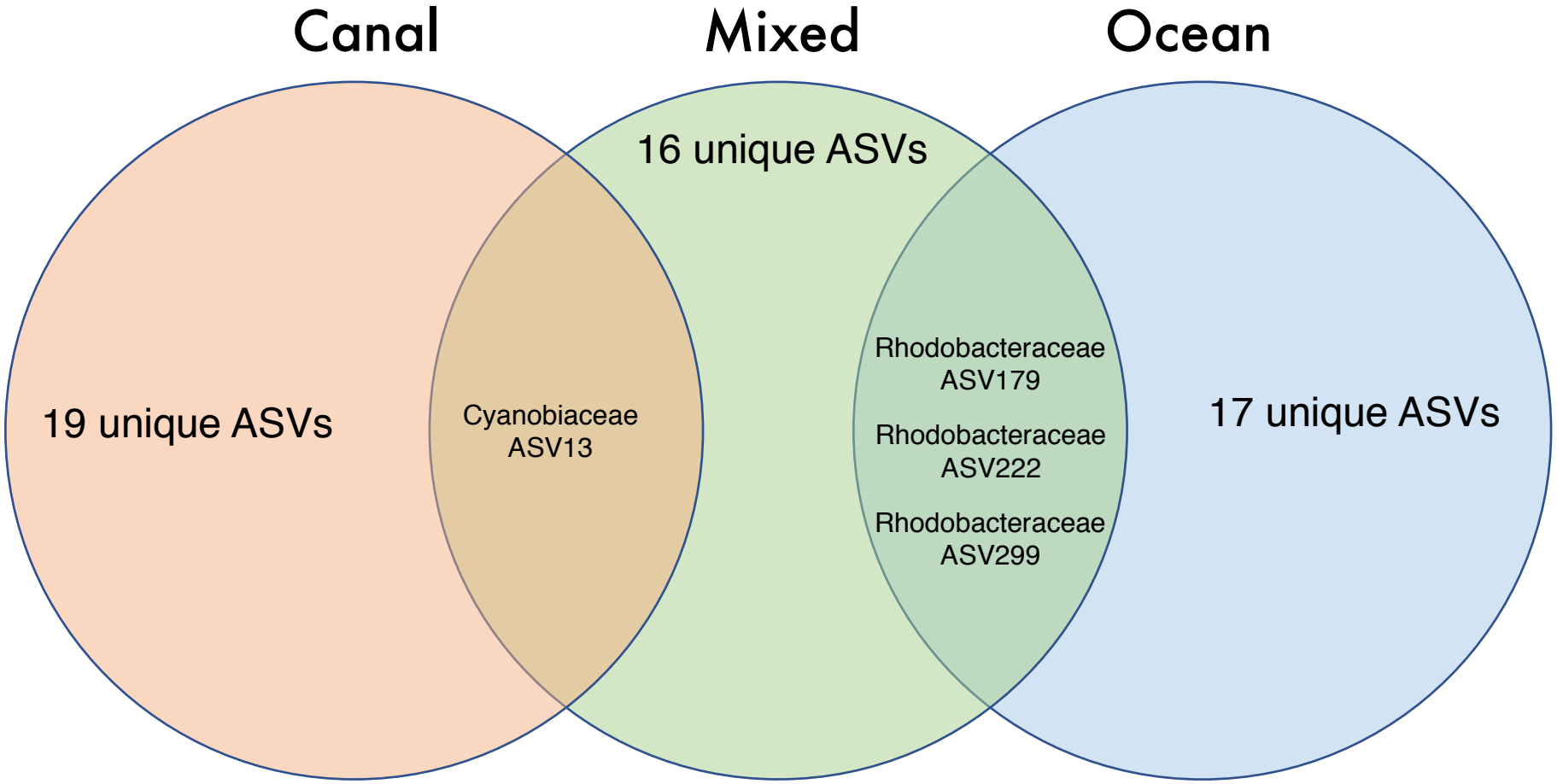
Appendix 9: Stacked pie chart showing the relative abundance for the top 20 most abundant taxa, at Phylum (large text) and Family (small text) level at the canal mouth sites.



Appendix 10: Stacked pie chart showing the relative abundance for the top 20 most abundant taxa, at Phylum (large text) and Family (small text) level at the mouth of the Miami River (MR01).

Appendix 11: Shows the total volume of water conducted through each waterway, over the course of the study period (Sep. 2017 – Jan. 2019). The volume of the outflow is expressed in millions of cubic meters of water. Data from: South Florida Water Management District.

Water Way	**Volume of outflow (Millions of m^3)**
Black Creek	367
Snapper Creek	352
Miami River	350
Little River	280
Cutler Drain	57
Coral Gables Waterway	17



Appendix 12: ven diagram showing the top 20 most abundant taxa at three locations. Although there are families in common between sites, they typically belong to different ASVs. The canal and mixed sites only share one ASV in common. The ocean influenced and mixed sites only share three ASVs in common. The canal and ocean influenced sites have no ASVs in common.

Laboratory Protocol for the Biscayne Bay Microbiome Project

Filtering water samples

1. Autoclave the Büchner funnel(s), and a 1.0L bottle used for filter sterile water.

2. Set up vacuum pump, and the two aspirator flasks.

3. Sterilize forceps with ETOH, and flame. Wait a moment to let the forceps cool.

4. Open the envelope containing the sterile 0.45µm filter paper. Using the forceps, carefully remove the paper backing and place on the filter stand checker-side up.

   Note: Never touch filters with bare hands. Please always use gloves or forceps. Filter must only have microbes found on sample.

5. Turn on vacuum, and pour water into the funnel. Make sure the vacuum is ≤10 PSI.

   Note: The volume you can put through one filter depends on the amount of suspended particulate in your sample. You can typically filter ~0.5L through each filter, before the process becomes painfully slow.

6. When filtering in completed, sterilize the forceps again and use them to fold the filter paper like a taco, and then like a pizza. Carefully place into a 1.5mL centrifuge tube for storage.

   Note: For this project sample tubes are labeled with the site name, month and year collected. If one, two or three filters are produced from the same sample. Also label the tube with with an A, B or C, respectively.

7. Samples can now be stored indefinitely in a freezer at -20 °C or -80 °C.

8. Between samples thoroughly flush funnel with filter sterile water (e.g. Millipour).

   Note: If you produced 2 filters from the same sample store in separate freezers (if possible), so if one freezer crashes you have a backup.

DNA extraction

Follow the protocol provided with the DNeasy Powerlizer Power Soil Kit. For each sample set up a rack with the tube containing the sample to be extracted, the power bead tube, 4 – 2.0 mL collection tubes and the MB spin column.

1. Add 0.25 g of soil sample to the PowerBead Tube provided.

2. Add 750µL of PowerBead Solution to the PowerBeadTube.

3. Add 60µL of Solution C1 and invert several times or vortex briefly.

4. Secure tubes in the homogenizer and run at 4,000 RPM for 45 s.

5. Centrifuge tubes at 10,000 x *g* for 30 s.

6. Transfer the supernatant to a clean 2 mL collection tube.

7. Add 250µL of Solution C2 and vortex for 5 s. Incubate at 2–8°C for 5 min.

8. Centrifuge the tubes for 1 min at 10,000 x *g*.

9. Avoiding the pellet, transfer up to 600µL of supernatant to a clean 2 ml collection tube.

10. Add 200 µL of Solution C3 and vortex briefly. Incubate at 2–8°C for 5 min.

11. Centrifuge the tubes for 1 min at 10,000 x *g*.

12. Avoiding the pellet, transfer up to 750µL of supernatant to a clean 2 ml collection tube.

13. Shake to mix Solution C4 and add 1200µL to the supernatant. Vortex for 5 s.

14. Load 675µL onto an MB Spin Column and centrifuge at 10,000 x *g* for 1 min. Discard flow through.

15. Repeat step 14 twice, until all of the sample has been processed.

16. Add 500µL of Solution C5. Centrifuge for 30 s at 10,000 x *g*.

17. Discard the flow through. Centrifuge again for 1 min at 10,000 x *g*.

18. Carefully place the MB Spin Column into a clean 2 ml collection tube. Avoid splashing any Solution C5 onto the column.

19. Add 100µL of Solution C6 to the center of the white filter membrane. Alternatively, you can use sterile DNA-Free PCR Grade Water for this step (cat. no. 17000–10).

20.Centrifuge at room temperature for 30 s at 10,000 x *g*. Discard the MB Spin Column. The DNA is now ready for downstream applications. Store at –20° C.

Check Extraction

To confirm you have successfully extracted DNA from your water sample, run a quick gel electrophoresis. Select an appropriately sized gel box.

	TBE (new)	**Agar**
Small	50 mL	0.5g
Medium	100 mL	1.0g
Large	350 mL	3.5g

Warm the TBE and Agar in a beaker using the microwave (30sec - 1min). Make sure the combs are in place and let the gel cool and set (~15min). Mix 2µL of the extracted DNA with 2µL of the loading buffer & GelRed solution. In one well is each row load 2-5µL of the 100bp ladder.

Run gel for 45 min at 75V. Image the gel using transmission UV lighting.

Test PCR

If the gel shows you have successfully extracted DNA, now run a test PCR. The following steps should be done in a sterile environment (i.e. PCR hood). To prepare your samples for PCR, create the following PCR solution:

Component	**25 µL rxn**
Water, Nuclease-free	9.5 µL
Platinum 2x MasterMix	12.5 µL
10µ*M* forward primer	0.5 µL
10µ*M* reverse primer	0.5 µL
Template DNA	2.0 µL

The above recipe is per sample. If you have 10 samples multiply the volume of the first 4 components by 12 (10 samples + 2 extras to account for error). Pipet 24µL of the solution into 10 PCR tubes, then add 1.0 µL of the template DNA to each tube.
If you want each sample to have 25 µL of product you will need to mix 10.5 µL
Note: if you have a low concentration of DNA in your sample you may need to use more than 1.0 of your template DNA. For every additional micro-liter of DNA subtract an equal amount of water from your solution. (e.g. if you use 2.0µL of template DNA, then you will only add 9.5 µL of water.)

In addition to your template DNA prepare two identical vials as positive and negative controls. For the positive control replace the template DNA with 1.0 µL of extracted DNA form E. coli. For the negative control replace the template DNA with an extra 1.0 µL of water.

Load the vials into the thermocycler, making sure all the caps are securely closed, and carefully tighten the lid of the machine. In the "saved files" run the protocol for *16s Platnium*. This should take about 2.5hrs. and the end of the process the thermocycler will hold the samples at 4 °C, indefinitely.

Afterward run the samples, positive and negative controls, and a ladder on a gel for 45 min at 75V. See previous directions for preparing gel electrophoresis. A thick bright band of ~300bp indicates successful amplification of the targeted 16S region.

If the gel shows significant levels of contaminants, try diluting the sample at a ratio of 1:10 or 1:20. Then run the PCR and gel electrophoresis again.

Barcoding PCR

If the test PCR successfully amplified your DNA. Run another PCR that will be used for sequencing.

Component	**50µL rxn**
Water, Nuclease-free	21µL
Platinum 2x MasterMix	25µL
10µ*M* forward primer w/barcode	1.0µL
10µ*M* reverse primer w/unique barcode	1.0µL
Template DNA	2.0µL

Afterward run the samples, positive and negative controls, and a ladder on a gel for 45 min at 75V, 200mA. See previous directions for preparing gel electrophoresis.

PCR Clean-up

In this step *AMPure XP* beads are used to purify the 16S V3 and V4 amplicon away from free primers and primer dimer species.

Bring the *AMPure XP* beads to room temperature.

1. Centrifuge the PCR plate at 1,000 x g at 20C for 1 minute to collect condensation. Then carefully remove the seal.
2. Vortex the *AMPure XP* beads for 30 sec, and pour in a trough for the multichannel pipet.
3. With the multichannel pipet add 56uL of *AMPure XP* beads to each well of the plate and triturate 10 times.
4. Incubate at room temperature without shaking for 5 minutes.

Note: steps 5-9 are performed on the magnetic stand.

5. Place the plate on a magnetic stand for 2 minutes or until the supernatant has cleared.
6. With the Index PCR plate on the magnetic stand, use a multichannel pipette set to 200µL, to remove and discard the supernatant. Change tips between samples.

7. With the Index PCR plate on the magnetic stand, wash the beads with freshly prepared 80% ethanol as follows:
    a. Using a multichannel pipette, add 200µL of freshly prepared 80% ethanol to each sample well.
    b. Incubate the plate on the magnetic stand for 30 seconds.
    c. Carefully remove and discard the supernatant.
8. With the Index PCR plate on the magnetic stand, perform a second ethanol wash as follows:
    a. Using a multichannel pipette, add 180µL of freshly prepared 80% ethanol to each sample well.
    b. Incubate the plate on the magnetic stand for 30 seconds.
    c. Carefully remove and discard the supernatant (200µL).
    d. Use a P20 multichannel pipette with fine pipette tips to remove excess ethanol.
9. With the Index PCR plate still on the magnetic stand, allow the beads to air-dry for 10 minutes.
10. Remove the Index PCR plate from the magnetic stand. Using a multichannel pipette, add 27.5µL of 10 mM Tris pH 8.5 to each well of the Index PCR plate.
11. Vortex and briefly centrifuge the PCR plate, until beads are fully resuspended.
12. Incubate at room temperature for 2 minutes.
13. Place the plate on the magnetic stand so that only the tips of the wells are touching the magnets. Incubate for 2 minutes or until the supernatant has cleared. Slowly slide the plate deeper into the stand so the magnets collect on the sides of the wells.
14. Using a multichannel pipette, carefully transfer 25µL of the supernatant from the Index PCR plate to a new 96-well PCR plate. Change tips between samples to avoid cross-contamination.

Determine DNA concentration using *Qubit*

Note: Do not operate the instrument in direct sunlight. All reagents and steps should be at room temperature (22–28ºC).

1. Set up the required number of 0.5-mL tubes for the two standards and your samples. Label the tube lids.

    Note: Use only thin-wall, clear, 0.5-mL PCR tubes. Do not label the side of the tube as this could interfere with the sample read.

2. Prepare the Qubit working solution by diluting the Qubit dsDNA HS Reagent 1:200

in Qubit dsDNA HS Buffer. Use a clean plastic tube each time you prepare Qubit working solution. Do not mix the working solution in a glass container.

Note: The final volume in each tube must be 200 µL. Prepare sufficient Qubit working solution to accommodate your samples and both standards. E.g. 8 samples + 2 standards = 10 tubes: ~200 µL per tube in 10 tubes yields 2 mL of working solution.

3. Add 190 µL of Qubit working solution to both of the tubes used for standards.

4. Add 10 µL of each Qubit standard to the appropriate tube, then mix by vortexing 2–3 seconds. Be careful not to create bubbles.

   Note: Careful pipetting is critical to ensure that exactly 10 µL of each Qubit standard is added to 190 µL of Qubit working solution.

5. Add 199 µL of the Qubit working solution to each individual assay tube

6. Add 1.0µL of your sample to its corresponding assay tube. Then vortex for 2–3 seconds.

7. Allow all tubes to incubate at room temperature for 2 minutes.

   Note: after incubation, the fluorescence signal is stable for 3 hours, at room temperature.

Calibration

For each assay, you have the choice to calibrate the fluorometer using new standard solutions **or** to use the values from the previous calibration.

8. On the home screen, choose the High Sensitivity DNA assay.

9. Press Yes to read new standards. A prompt to insert Standard #1 appears on the screen.

10. Insert Standard #1 into the sample chamber, close the lid, and press "Read".

    Note: Take care to not get fingerprints or other marks on the side of the Qubit tube. Giving the tube a quick wipe with a Kimwipe is not a bad idea.

11. Repeat step 10 using standard #2.

    Note: Make sure you insert the standards in the correct order (i.e #1 then #2)

Reading Samples

12. Insert your first sample into the sample chamber, close the lid, and press "Read".

13. Record the value in the Dilutions Excel Sheet. Repeat steps 12 and 13 until all your samples have been processed.

Dilutions

Enter the values obtained from the Qubit assay into the "dilutions" spreadsheet to calculate the dilution factor necessary.

Pooling

After the individual samples are diluted to the prescribed amount, combine 5μL of each sample into one microfuge tube. Check the DNA concentration again, using the Qubit protocol (repeat this on 3 different sub-samples form the pool). The DNA concentration should be between 4-6 ng/μL. If the pool passes the Qubit assay run the pool on the tapestation to determine the quality of DNA (see tapestation protocol).

MiSeq Loading

Follow the directions provided by Illumina for loading the pool into the MiSeq.

R code

Sample of the code used for statistical analysis in RStudio. This code represents the collective knowledge of the Lopez Lab.

Phyloseq Package

```
#PhyloSeq
source('http://bioconductor.org/biocLite.R') #if you don't have Phyloseq installed.
biocLite('phyloseq') #if you don't have Phyloseq installed.

#load packages
library(phyloseq)
library(ggplot2)
#set default theme for graphics
theme_set(theme_bw())

##load library
library(ggplot2)
library(phyloseq)
library(ape)
###now to import to phyloseq
#read in otu table
otu_table=read.table(file= "feature-table.tsv", header=TRUE, sep ="\t", row.names = 1)
otu_table=as.matrix(otu_table)

##Read in taxonomy. Make sure your taxonomy file is separated columns for Kingdom,
Phylum, Class, etc...
taxonomy=read.table(file = "TaxonomyClean.tsv", sep = "\t", header = T, row.names = 1)
head(taxonomy)
taxonomy=as.matrix(taxonomy)
##add metadata
metadata=read.table("WQ_Data2.tsv", header=T, sep = "\t", row.names = 1)

##load tree
phy_tree=read_tree("tree-unrooted.nwk")
##import as phyloseq objects
OTU= otu_table(otu_table,taxa_are_rows=TRUE)
TAX=tax_table(taxonomy)
META=sample_data(metadata)

##check that you OTU names are consistent across objects
taxa_names(TAX)
taxa_names(OTU)
taxa_names(phy_tree)

##merge into one phyloseq object
physeq = phyloseq(OTU,TAX,META,phy_tree)
physeq
##check rank names of taxonomy
rank_names(physeq)
##now continue analysis in phyloseq
## check reads of samples
sample_sums(physeq)[1:10]
```

```
## basic stats for read of samples
mean(sample_sums(physeq))
min(sample_sums(physeq))
max(sample_sums(physeq))
sd(sample_sums(physeq))

##prune taxa from the OTU table that are in zero samples (these are in other samples
on the run)
merge=prune_taxa(taxa_sums(physeq)>0,physeq)
merge

##create for taxa above relative abundance of 1%
merge99 = transform_sample_counts(merge, function(x){x/sum(x)})
otu_table(merge99)[otu_table(merge99)<.01] <- 0
merge99 = prune_taxa(taxa_sums(merge99)>0,merge99)
merge99 = transform_sample_counts(merge99, function(x){x*100})
otu_table(merge99) = floor(otu_table(merge99))
merge99


#create a normalized data set for lowest reads
# Normalize to 24381 reads per sample (proportions) rounding down
mnorm = transform_sample_counts(physeq, function(x) {24381*x/sum(x)})
otu_table(mnorm) = floor(otu_table(mnorm))
mnorm = prune_taxa(taxa_sums(mnorm)>0,mnorm)
mnorm

##look at the rank abundance plots for the top 100 OTUs
sampleprop = transform_sample_counts(physeq, function(x) {x/sum(x)})
barplot(sort(taxa_sums(sampleprop),TRUE)[1:100]/nsamples(sampleprop),las=2,names.arg="
",cex.axis=.7)
title(main="Rank abundance plots for the top 100 OTUs") #Places a title on the graph

##Alpha diversity
plot_richness(merge, color = "Site")
plot_richness(merge, color = "SiteType")

plot_richness(merge, x="SiteType", color = "SiteType")
plot_richness(merge, x="SiteType", color = "Site")

plot_richness(merge, x= "Site", color = "Site")
plot_richness(merge, x= "Site", color = "SiteType")

##obersved vs choa1
p = plot_richness(merge, x="Site", color = "SiteType", measures =
c("Observed","Chao1"))
p + geom_boxplot(data = p$data, aes(x=Site, y=value, color=NULL, fill=NULL
),apha=0.1)##+ geom_point(size =3, alpha=0.7)

##Shannon vs Inv Simpson by Site
q = plot_richness(merge, x="Site", color = "SiteType", measures =
c("Shannon","InvSimpson"))
```

```
q + geom_boxplot(data = q$data, aes(x=Site, y=value, color=NULL, fill=NULL
),alpha=0.1) ##+ geom_point(size =3, alpha=0.7)
##Shannon vs Inv Simpson by SiteType
q = plot_richness(merge, x="SiteType", color = "SiteType", measures =
c("Shannon","InvSimpson"))
q + geom_boxplot(data = q$data, aes(x=SiteType, y=value, color=NULL, fill=NULL
),alpha=0.1) ##+ geom_point(size =3, alpha=0.7)
estimate_richness(merge, measures=c("InvSimpson", "Shannon")) Reterns the Shannon and
InvSimpson index for each sample ID


##NDMS Charts
library(ggplot2)
library(plyr)
#set theme
theme_set(theme_bw())
##prune
GP = merge
wh0 = genefilter_sample(GP, filterfun_sample(function(x) x > 5), A=0.5*nsamples(GP))
GP1 = prune_taxa(wh0, GP)
##transform
GP1 = transform_sample_counts(GP1, function(x) 1E6 * x/sum(x))
##keep only the most abundant phyla
phylum.sum = tapply(taxa_sums(GP1), tax_table(GP1)[, "Phylum"], sum, na.rm=TRUE)
top20phyla = names(sort(phylum.sum, TRUE))[1:20]
GP1 = prune_taxa((tax_table(GP1)[, "Phylum"] %in% top20phyla), GP1)
GP1
#look at plots
GP.ord <- ordinate(GP1, "NMDS", "bray")
p1 = plot_ordination(GP1, GP.ord, type="taxa", color="Phylum", title="taxa")
print(p1)
#justsamples
p2 = plot_ordination(GP1, GP.ord, type="samples", color="SiteType", shape="SiteType")
p2 + geom_polygon(aes(fill=AlphaType)) + geom_point(size=5) + ggtitle("samples")

#biplot graphic
p3 = plot_ordination(GP1, GP.ord, type="biplot", color="SiteType", shape="Phylum",
title="biplot")
# Some stuff to modify the automatic shape scale
GP1.shape.names = get_taxa_unique(GP1, "Phylum")
GP1.shape <- 15:(15 + length(GP1.shape.names) - 1)
names(GP1.shape) <- GP1.shape.names
GP1.shape["samples"] <- 16
p3 + scale_shape_manual(values=GP1.shape)
p4 = plot_ordination(GP1, GP.ord, type="split", color="Phylum", shape="SiteType",
label="SiteType", title="split")
p4

##
ordu = ordinate(GP1, "PCoA", "unifrac", weighted=TRUE)
plot_ordination(GP1, ordu, color="SiteType", shape="SiteType")

p = plot_ordination(GP1, ordu, color="SiteType", shape="SiteType", label="SiteType")
```

```
p = p + geom_point(size=7, alpha=0.75)
p = p + scale_colour_brewer(type="qual", palette="Set1")
p + ggtitle("MDS/PCoA on weighted-UniFrac distance, GlobalPatterns")


##looking at alpha diversity
# Initialize matrices to store richness and evenness estimates
richness = matrix(nrow=137,ncol=100)
row.names(richness) <- sample_names(physeq)
evenness = matrix(nrow=137,ncol=100)
row.names(evenness) <- sample_names(physeq)

# It is always important to set a seed when you subsample so your result is replicable
set.seed(3)
# For 100 replications, rarefy the OTU table to 1000 reads and store the richness and
evenn es estimates. The default for the rarefy_even_depth command is to pick with
replacement so I set it to false. Picking without replacement is more computationally
intensive
for (i in 1:100) {
  r=rarefy_even_depth(physeq,sample.size=1000,verbose=FALSE,replace = FALSE)
  rich= as.numeric(as.matrix(estimate_richness(r,measures="Observed")))
  richness[,i]=rich
  even=as.numeric(as.matrix(estimate_richness(r,measures="Shannon")))
  evenness[,i]=even
}
# Create a new matrix to hold the means and standard deviations of all the richness
estimates
rich.stats = matrix(nrow=137,ncol=2)
rich.stats[,1] = apply(richness,1,mean)
rich.stats[,2] = apply(richness,1,sd)
rich.stats = data.frame(row.names(richness),rich.stats)
colnames(rich.stats) = c("samples","mean","sd")
# Create a new matrix to hold the means and standard deviations of the evenness
estimates
even.stats = matrix(nrow=137,ncol=2)
even.stats[,1] = apply(evenness,1,mean)
even.stats[,2] = apply(evenness,1,sd)
even.stats = data.frame(row.names(evenness),even.stats)
colnames(even.stats) =c("samples","mean","sd")

##create a boxplot
# A data frame of all sample names and associated butterfly species
Sp = data.frame(X.SampleID=sample_data(physeq)$id,Site=sample_data(physeq)$Site)
head(Sp)
#
# Rename the headers
colnames(rich.stats)[1] <- "X.SampleID"
rich.stats2 = merge(rich.stats, Sp,by="X.SampleID")
# Make a boxplot of community richness
boxplot(mean~SampleLocation,data=rich.stats2, ylab="Richness (500
reads)",xlab="",xaxt="n",main="Microbial community richness of butterfly species")
text(1:33, par('usr')[3]-.25, labels = levels(Sp$SampleLocation), srt = 45, adj = 1.2,
xpd = TRUE, cex=.9)
```

```
##calculate alpha diversity based on core microbiome
coreRichness = (estimate_richness(merge99,measures = "Observed"))
coreevenness = (estimate_richness(merge99,measures = "Shannon"))
#combine data frame
# Reformat data frames for core and noncore richness so they can be combined
coreRich = data.frame(richness = coreRichness$Observed)
coreRich$type = "core"
Rich = data.frame(richness = rich.stats$mean)
Rich$type = "full"
combinedRich = rbind (Rich,coreRich)
# Make a histogram of richness estimates colored by type (core or full)
ggplot(combinedRich,aes(richness,fill=type)) + geom_histogram(alpha = 0.5, position =
'identity')

# Reformat data frames for core and noncore evenness so they can be combined
coreEven = data.frame(evenness = coreevenness$Shannon)
coreEven$type = "core"
Even = data.frame(evenness = even.stats$mean)
Even$type = "full"
combinedEven = rbind (Even,coreEven)
# Make a histogram of evenness estimates colored by type (core or full)
ggplot(combinedEven,aes(evenness,fill=type)) + geom_histogram(alpha = 0.5, position =
'identity')
#Now we will do a kruskal-wallis test to look for differences in community alpha
diversity between sites
kruskal.test(mean~Site, data = rich.stats2)

library(pgirmess)
kruskalmc(rich.stats2$mean, rich.stats2$Site)
kr.out = read.csv("/Users/ericfortman/Nova/Thesis/Analysis/Phyloseq results/")
head(kr.out)

##Heatmap
gpt <- subset_taxa(physeq, Kingdom=="Bacteria")
gpt <- prune_taxa(names(sort(taxa_sums(gpt),TRUE)[1:30]), gpt)#top 30 taxa

#Creates the same plot, but with a different look.
plot_heatmap(gpt, "NMDS", "bray", "Sample_ID2", "Family_ASV", low="#66CCFF",
high="#000033", na.value="white", sample.order= "SiteType.Sample.ID")
```

Vegan Package

```
#start with setting your working directory
setwd("C:\\Users\\your_file_path_here") #you can also manually set your WD by going to
"Session" in the menu bar above

#now we need to load our data
dat <- read.table(file= "feature-table.tsv", header = TRUE, sep ="\t", row.names = 1)
```

```
#lets look at imported file
View(dat)

#as you can see the samples are in columns and need to be in the rows, so we need to
flip or transpose the file
#transpose the data to rows
t.dat <- as.data.frame(t(dat))

# lets look at the first rows of the new file to see if our code worked
t.dat[1:5,1:5]

#as you can see the samples are now row names and we can set this new file to be our
data file
dat <-t.dat

#Now we need to import the metadata file into our R image, we will do this with the
file choose command as another example of how to load a data file
metadata <- read.table(file.choose(), header=T, sep ="\t", row.names=1)

#view to check the file
View(metadata)

#now we need to make it, so we only have the data for the specific rows we are looking
at, aka all the samples are the same for both files
#first we are creating a new object for common row names from both files using the
intersect command
common.rownames <- intersect(rownames(dat),rownames(metadata))
View(common.rownames)
#next we will set the data file and metadata file to have only the data that includes
these common names
dat <- dat[common.rownames,]
metadata <- metadata[common.rownames,]

#now to make sure all the row names are the same (equal) following our code, if they
are not this will return a False
all.equal(rownames(dat),rownames(metadata))

#reduce noise (get rid of single and doubletons), this removers OTUs that only show up
once or twice
otu.abund<-which(colSums(dat)>2)
dat.dom<-dat[,otu.abund]

#reduce OTUs that occur in small amount of samples, this will get rid of taxa that are
non-dominant and is your choice on whether to include in your final code
#need to load required packages using the library command, these can be downloaded in
the packages tab in the lower right screen.
library(vegan)
library(base)

#all this will get rid of OTUs that are below 0.05 percent in the data, aka probably
not important
dat.pa<-decostand(dat.dom, method ="pa")
dat.otus.05per<-which(colSums(dat.pa) > (0.05*nrow(dat.pa)))
```

```
dat.05per<-dat.dom[,dat.otus.05per]

#now our data is ready to start answering some questions
#transform (Standardization not transformation?) data for relative abundance (this is
an important tool for answering many questions)
dat.ra<-decostand(dat.05per, method = "total")

shann<- diversity(dat.ra, "shannon") #returns Shannon index of beta diversity for each
site
betainvsimp <- as.data.frame(t(shann))
View(betashann)
invsimp<- diversity(dat.ra, "invsimpson") #Returns Inverse Simpson index of beta
diversity for each site
betainvsimp <- as.data.frame(t(invsimp))
View(betainvsimp)

#print to Excel sheet, this allows you to view your relative abundance data and is
needed to make charts such as Kronos
dat.rat <- as.data.frame(t(dat.ra))
View(dat.rat) #double check it worked before making a txt file
write.table(dat.rat, "/Users/ericfortman/Nova/Thesis/Analysis/relative_abundance.txt",
sep="\t",row.names = T)

#lets look into beta diversity with Bray Curtis index
#look at bray curtis dissimilarity
dat.bc.dist<-vegdist(dat.ra, method = "bray")


#adonis - Permutational Multivariate Analysis of Variance Using Distance Matrices
adonis(dat.bc.dist~Site*Date, data = metadata)


#run a pcoa for adonis results based on sample location
dat.betadisp<-betadisper(dat.bc.dist,metadata$Site)

#view in boxplot
boxplot(dat.betadisp)


#view in pcoa graphic form
plot(dat.betadisp)
title(main="Vegan PCoA") #Places a title on the graph

#now to run a pairwise adonis (Performs pairwise comparisons between group levels with
corrections for multiple testing)
library(RVAideMemoire)
pairwise.perm.manova(dat.bc.dist,metadata$Site)

#Now lets see what the significance of the environmental factors is for our diversity
with a CCA
#we need to choose a set seed or our numbers will be different each time
```

```
set.seed(42);env.cca<-
cca(dat.ra~DO_percent+Salinity+Temperature+NH3_N+NOX+TP+DepthSounding+Rain3DayTotal,
data =metadata)#CCA for envronmental data
env.cca
vif.cca(env.cca)
#make sure they add up to more than ten or you may need to remove if its over 20 def
remove

#step 2, zero the variables
set.seed(42);lwr<- cca(dat.ra~1, data=metadata)
lwr

#unsing a forward selecting model, must keep our set seed
set.seed(42);mods.all<- ordiR2step(lwr, scope = formula(env.cca))
mods.all
vif.cca(mods.all)

R2.adj.all<-RsquareAdj(mods.all)
R2.adj.all

mods.all$anova
#repeat this for different sites to see if the variance is different for each site (to
do this just change the metadata file)

## try plotting this CCA
cca.p <- plot(mods.all,type = "none")
points(cca.p, "sites", col= as.numeric(metadata$Site), pch =
as.numeric(metadata$Site))

ef.all<-
envfit(cca.p,metadata[,c("Salinity","DO_percent","Temperature","NOX","TP","DepthSoundi
ng")])
plot(ef.all)
title(main="mods.all$anova CCA plot")

#To place a title and legend
title(main="mods.all$anova CCA plot")
legend("center",legend = as.character(paste(" ", unique(metadata$Site))), pch=
as.numeric(unique(metadata$Site)))

#now lets look into ndms chart
comm.bc.mds<-metaMDS(dat.ra, distance="bray")
mds.fig<-ordiplot(comm.bc.mds, display="sites")
ordiellipse(mds.fig, metadata$Site, label = T, conf = 0.95)#adds circles and lables

##this is how you can adjust the x and y axis
mds.fig<-ordiplot(comm.bc.mds, display="sites", xlim=c(-1.5,3), ylim = c(-1,2))  ####
adjust x-limit and y-limit

##adjust colors: 15=square,16=circle, 17=triangle 18=diamond
points(mds.fig,"sites", pch = 15, col = "grey", select = metadata$Site == "BB14")
points(mds.fig,"sites", pch = 16, col = "grey", select = metadata$Site == "BB22")
points(mds.fig,"sites", pch = 15, col = "green3", select = metadata$Site == "BB34")
```

```
points(mds.fig,"sites", pch = 15, col = "blue", select = metadata$Site == "BB37")
points(mds.fig,"sites", pch = 15, col = "red", select = metadata$Site == "BB39A")
points(mds.fig,"sites", pch = 16, col = "green3", select = metadata$Site == "BISC127")
points(mds.fig,"sites", pch = 16, col = "red3", select = metadata$Site == "BL01")
points(mds.fig,"sites", pch = 15, col = "yellow", select = metadata$Site == "CD01A")
points(mds.fig,"sites", pch = 17, col = "green3", select = metadata$Site == "CG01")
points(mds.fig,"sites", pch = 17, col = "gray", select = metadata$Site == "LR01")
points(mds.fig,"sites", pch = 13, col = "black", select = metadata$Site == "MR01")
points(mds.fig,"sites", pch = 18, col = "grey", select = metadata$Site == "MR03")
points(mds.fig,"sites", pch = 18, col = "green3", select = metadata$Site == "SP01")

#legend
legend("topright",legend=as.character(paste(" ",unique(metadata$Site))), cex =
0.99,pch=19,col=1:length(unique(metadata$Site)))
ordiellipse(mds.fig, metadata$Site, label = F, conf = 0.95, lty = 2) #adds circles

title(main="Vegan NMDS plot") #Places a title on the graph

###Simper Test
dat.simp<-simper(dat.ra, metadata$Site, permutations = 99)##change to 999 after intial
run
sink("Simper_by_site.csv")
summary(dat.simp)
sink()
##look at the file and you can see what OTUs are causeing the difference between the
sites, look up the OTU and see if that is interesting

#Simper by Site Type
dat.simp<-simper(dat.ra, metadata$SiteType, permutations = 999)##change to 999 after
intial run
sink("Simper_by_sitetype999.csv")
summary(dat.simp)
sink()
```

Bar Plots Code

```
library(ggplot2)
library(ggthemes)
library(plyr)
library(scales)

charts.data <- read.csv("BarPlot_SiteType.csv") #specify source data
#Assign x-axis, y-axis, labels
BarPlotBySiteType <- ggplot()+ theme_bw() + geom_bar(aes(y= AvgRelAbun, x=
SiteType, fill= Order), data= charts.data,stat="identity")+ ggtitle("Top 20
Most Abundant Taxa by Site Type")
BarPlotBySiteType #Renders the graph

charts.data <- read.csv("BarPlot_Site.csv") #specify source data
```

```
#Assign x-axis, y-axis, labels
BarPlotBySite <- ggplot()+ theme_bw() + geom_bar(aes(y= AvgRelAbun, x= Site,
fill= Order), data= charts.data,stat="identity") + ggtitle("Top 20 Most
Abundant Orders by Site")
BarPlotBySite #Renders the graph

charts.data <- read.csv("BarPlot_Site.csv") #specify source data
#Assign x-axis, y-axis, labels
BarPlotBySite <- ggplot()+ theme_bw() + geom_bar(aes(y= AvgRelAbun, x= Site,
fill= Family), data= charts.data,stat="identity") + ggtitle("Top 20 Most
Abundant Famlies by Site")
BarPlotBySite #Renders the graph
```

Draft Manuscript

To be submitted to the Bulletin of Marine Science

Characterization of Bacterial Communities in Biscayne Bay Through Genomic Analysis

P. Eric Fortman, Christian Avila, Jose V. Lopez

Abstract

Biscayne Bay is a shallow oligotrophic estuary in Southeast Florida. Dredging of rivers and canals has greatly altered the flow of freshwater into the bay. This, coupled with the rise of a sprawling urban & suburban development, has greatly increased the nutrient load in the bay. This study examined the bacterial community at 14 stations throughout Biscayne Bay —6 stations were located at the mouths of canals; 1 upstream-canal station; 6 stations in the center of the bay; and one ocean influenced station, located near the entrance to the bay. Surface water samples were taken monthly for one year. The 16S rRNA gene was used to identify bacterial community composition. There were 19,680 Amplicon Sequence Variants (ASVs) identified across all 146 samples. Salinity and total phosphorous were the primary factors explaining bacterial biodiversity. Biodiversity in bacterial communities in the Miami River and the ocean influenced site, were unique compared to other sites in the study. Alpha and β-diversity were generally homogeneous over most of the study area. Looking at α-diversity, the two stations on the Miami River were statistically identical and had higher diversity. The ocean influenced station, was statistically unique and had lower α-diversity. The remaining 11 stations had moderate diversity and were statistically identical, appearing to be a combination of the previously mentioned Miami River sites and the ocean influenced site. Beta diversity showed a similar pattern; with the exception that the site located at the mouth of Black Creek could now be grouped with the Miami River sites.

Introduction

Biscayne Bay is a shallow oligotrophic estuary on the southeast coast of Florida. Excluding dredged areas, the maximum depth is 4 m, with an average of 1.8 m (Caccia & Boyer, 2005). Historically there were several free-flowing rivers into the bay. Landscape level human impacts began with efforts to drain the Everglades starting in 1903 (Cantillo et al., 2000). Rivers were channelized and new canals dredged, to increase water flow out of western Dade County. The historic pattern of seasonal freshwater flow from rivers, creeks and sloughs in to the bay, has been replaced by discrete releases through flood gates along canals. Water flow is tightly controlled by the South Florida Water Management District and Army Corps of Engineers. There are 19 canals that drain into Biscayne Bay (Cantillo et al., 2000). The primary drainages for urban and suburban Dade County are the Little River, Miami River, Coral Gables Waterway, Snapper Creek, Cutler Drain, and Black Creek Canal. Further south in the bay the Princeton and Mowry canals drain agricultural and some suburban areas, but these are beyond the scope of this study.

Seasonality in South Florida is principally delineated by rainfall, with the wet season running form May–October and the dry season running form November–April (Dame et al., 2000). Because of the bay's large surface area, precipitation is the dominant source of freshwater to the bay; followed by canal input and ground water discharge (Stalker et. al, 2009). Biscayne Bay is periodically exposed to naturally occurring disturbances such as tropical cyclones. In August 2005 Hurricane Katerina hit the Bay dumping up to 14" of rainfall within the watershed (Zhang et al., 2009). While the storm event caused many short-term changes to water quality, Zhang et al. (2009) observed that water quality returned to pre-storm conditions within three

months of the event. More recently in September 2017 Hurricane Irma hit Biscayne Bay. The hurricane significantly increased freshwater inflow to the bay. In the first week after the storm freshwater inflow increased by 148% –compared to a week before (Wachnicka et al., 2019). Similar to Hurricane Katrina water quality in the bay returned to "normal" in less than three months after Hurricane Irma (Wachnicka et al., 2019).

The northern end of the Bay (Oleta River through Key Biscayne) is more protected and receives less water exchange with the Atlantic Ocean. Dredging of Government Cut began in 1902. The spoils were used to construct Lummus, Dodge, and Fisher Islands –the first man made islands in the bay. The North Bay is now heavily modified, with very little natural shoreline remaining. This area is also home to the most urban and industrial land use. Turbidity, industrial pollution, nutrient loading, and sewage pollution are the primary problems facing the Northern Bay (Caccia & Boyer, 2005). The portion of the bay south of Key Biscayne, through the Safety Valve, and Ragged Keys sees more exchange with oceanic water (see figure 3). Development becomes less dense as you move south along the cost. The central bay (the area south of Cape Florida through Black Point) is characterized by suburban development and more remaining mangrove tracts along the coast. Pollution sources here come from localized problems such as marinas (Caccia & Boyer, 2005). The mainland of the southern Bay (Black Point to Card Sound) is a mix of suburban development, agriculture, and mangrove habitat. One anthropogenic feature of note in this area is the South Dade land fill, near Black Point.

Nutrients from septic tanks, leaky sewage lines, and fertilizer have led to eutrophication in the Bay. Caccia & Boyer (2005) identified several geographic patterns in water quality in the

bay, noting that land use is the major factor affecting water quality in the bay. Eutrophication from nitrate/nitrite–nitrogen seems to be more of a problem in the southern part of the watershed. Whereas total ammonia-nitrogen and total phosphorus are the major pollutants in the northern part of the watershed (Caccia & Boyer, 2007; Carey et al., 2007). Canals are responsible for the bulk of nitrogenous inputs into the bay (Caccia & Boyer, 2007; any more recent refs from other areas?). Precipitation directly into the bay is responsible for the bulk of fresh water inputs into the bay, followed by canals and then ground water (Stalker et. Al, 2009). Stalker et. al (2009) cautions even though ground water is the lowest constituent of freshwater input, it should not be ignored because it generally contains higher levels of nutrients, notably nitrogen and phosphorous. As a result of the increased nutrient load, persistent algal blooms and reduced seagrass coverage have been reported. Collado-Vides et al. (2013) described a persistent bloom of *Anadyomene spp*., which was first noted in 2006. The geographic range of the bloom extended from the Rickenbacker Causeway south to Chicken Key, with some sites experiencing algal coverage > 75% (Collado-Vides et al. 2013).

Microbiomes

Microbes in natural habitats generally exist as microbial communities (or "microbiomes") instead of in isolation. Marine bacterioplankton microbiomes play an important role in many biogeochemical processes (Bunse & Pinhassi, 2017). In marine ecosystems heterotrophic bacteria are the only organisms that fix dissolved organic material for use by primary producers (Bunse & Pinhassi, 2017). Seasonal variability in the microbial community is more pronounced in temperate and polar habitats, but it is still observed in subtropical and tropical regions (Figure

2; Bunse & Pinhassi, 2017). In Port Everglades inlet, an estuary just north of Biscayne Bay, seasonal variation in the bacterioplankton community was noted by O'Connell et al. (2018). The wet (May – October) season was characterized by higher species richness, and lower species evenness. Changes in community composition were most closely tied to changes in salinity and temperature (O'Connell et al., 2018).

Population dynamics of bacteria and phytoplankton reciprocally influence each other (Bunse & Pinhassi, 2017; Smith et al., 1999). Further, phytoplankton blooms can decrease light penetration and shade seagrasses, causing reduced seagrass coverage. In turn this cause the release of nutrients tied up in seagrass biomass and sediments, exacerbating the bloom (Boyer et al., 2009). A better understanding of how bacteria and phytoplankton affect each other can have applications in predicting and preventing hazardous algae blooms. Most time series data for microbiome studies are sampled in monthly intervals. However, the generation time of bacterioplankton can be hours or days. Therefore smaller-scale population fluctuations may serve as a precursor for more prolonged ecological shifts (Bunse & Pinhassi, 2017).

16S RNA

Traditionally, bacterial communities were studied by plating environmental samples on a petri dish and culturing them in the lab. A major drawback to this technique is that many –if not most– species of bacteria do not grow well in the laboratory (Pace, 1997). Advancements in genetic techniques, now allow environmental samples to be tested directly. The 16S rRNA gene was first used to study phylogeny in 1977 by Woese & Fox. The 16S gene has become the standard for bacterial phylogeny for three reasons: it is present in nearly all bacteria; the function

of the gene has not changed over time, suggesting randomly occurring mutations are a good measure of evolution; the gene is suitably large (1,500bp) for informatics analysis (Janda & Abbott, 2007). Despite advances in whole-genome sequencing techniques, amplicon sequencing of the 16S rRNA gene is still a viable method for comparing bacterial communities (Thompson et al., 2017). High throughput sequencing allows researchers to sequence genes, relatively quickly and cheaply (Mardis, 2008). These technologies have also made it possible to obtain sequences from many organisms simultaneously. The resulting data can then be analyzed to identify the number of amplicon sequence variants (ASVs). The number of ASVs present can be used a proxy to measure diversity and identify community structure.

Methods

Sample Collection

Water samples were collected in partnership with Miami-Dade County's Division of Environmental Resource Management (DERM). There were 14 fixed-stations throughout Biscayne Bay that were irregularly sampled between September 2017 and January 2019 (Figure 3 & table 1). The samples used for genetic analysis consisted of 1.0L surface water grab-samples. Several more liters of water were collected by DERM for chemo-physical analysis that included: salinity, temperature, dissolved oxygen, ammonia-nitrogen, nitrate-nitrite, and total phosphate. The sample locations range from Little River down through Black Point (Figure 3). These chemo-physical data are key for providing context for microbiome data (Knight et al. 2012).

Sample Preparation & Sequencing

The samples bound for genetic analysis were filtered through a 0.45μm nylon filter. DNA extraction conducted using a *Qiagen DNeasy PowerSoil Kit*. Sequencing was performed on the *Illumina MiSeq* platform. Proof of theory establishing that sequencing on the MiSeq platform accurately reflects a known bacterial community was established by Caporaso et al. (2012). The MiSeq output, containing the DNA sequences, was post processed using QIIME2 –an open source, Unix based command line program specifically designed for microbial community analysis (Bolyen et al., 2018). Within QIIME2 the software package DADA2 was used to remove chimeras and reads with a quality score <25, The advantage of DADA2 over other denoising techniques is that it infers sample sequences exactly, without coarse-graining into OTUs, and has high resolution –resolving differences of as little as one nucleotide (Callahan et al., 2016). Using exact sequences offers more flexibility than ASVs. By nature, exact sequences are "stable identifiers" and can be compared to any 16s rRNA database (Thompson et al., 2017). Taxonomy was determined for each ASV, by comparing the sequence to the Silva 132 learned classifier. The feature table, taxonomy file, and phylogenetic tree was exported from QIIME2 for downstream analysis in *R Studio* with the *PhyloSeq* and *Vegan* packages.

The PhyloSeq package was used to analyze α-diversity. Alpha diversity is the diversity (including species richness and evenness) with each site or sample (Whitaker, 1972) and was assessed using Shannon and Inverse Simpson indices. A Kruskal–Wallis test was used to compare α-diversity at each site and site type. The Vegan package was used to analyze β-diversity. Beta diversity is comparative diversity between sites, this assesses the similarity/dissimilarity of diversity between different sites. The Bray-Curtis Distances for β-diversity were calculated using Vegan. To assess relatedness between populations, Principle

component analysis (PCoA) was be done using Vegan, which incorporates phylogenetic signals in the 16S rRNA data. A Kruskal-Wallis test was used to determine if the diversity metrics differed across sites, and group sites with similar diversity measures together. Several, Multiple Least Square Regression analyses will be run to look for a possible correlations between microbial community and chemo-physical water quality data (Campbell et al, 2015; O'Connell et al, 2018). Canonical Correspondence Analysis (CCA) was used to identify possible correlations between species abundance and chemo-physical water quality data. To compare diversity between site types and identify taxa leading to significant differences, a SIMPER similarity percentage table was generated using *Vegan*.

Results

There were 19,680 bacterial taxa identified across all 146 samples. The alpha rarefication plot illustrates the plateau in α-diversity reached for each site type (Figure 4). The plateau signifies that an asymptote was reached during sequencing, and therefore adequate sampling depth was attained. This result indicates that within the sequencing run, no new taxa were being sequenced.

When looking at site type, the canal and canal mouth sites were statistically identical; and the bay and ocean influenced sites were statistically identical (p=6.406e-05) (Figure 5). A similar pattern is apparent Alpha diversity was also visualized with an NMDS plot (Figure 7). Looking at the sites individually, average α-diversity at MR01 (Miami River mouth), MR03 (the canal site) and BL01 (at the mouth of Black Creek) are statistically the same (Figure 6). These three sites also had some of the highest α-diversity observed in the study. Alpha diversity at BB37 (the ocean influenced site) is statistically distinct from all the other sites (p= 3.728e-03) (Figure 6). Site BB37

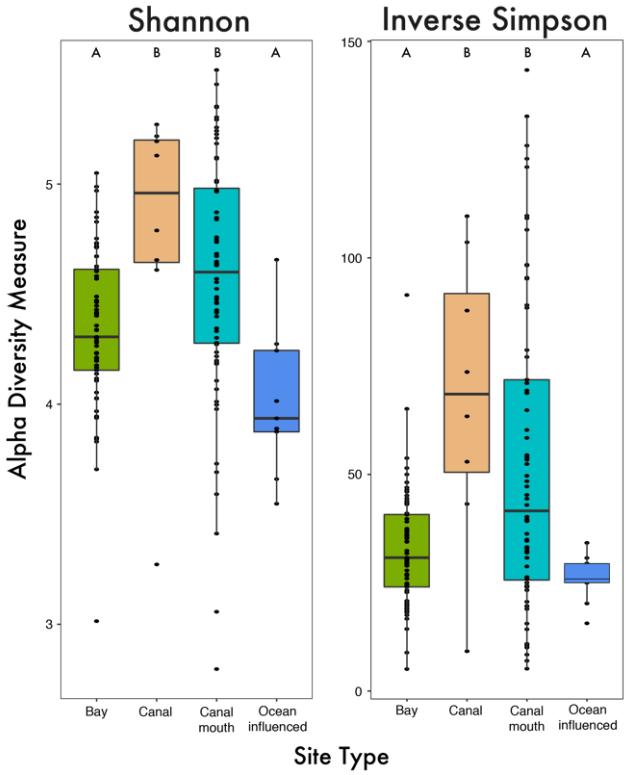
had the lowest α-diversity observed in the study. The remaining sites statistically fall in between these two extremes, sharing some combination of the "ocean influenced type" and the "Miami River type" sites. The ocean influenced site (BB37) had the least variability in α-diversity; while the Little River site (LR01) had the widest range of α-diversity, recorded in the study (Figure 6).

A PCoA (Principal Coordinates Analysis –used to asses dissimalarity) comparing β-diversity determined that MR01 and MR03 were statistically identical (Figure 8). Site BB37 (the ocean influenced site) is distinct from the all other sites. The remaning sites posessed characteristics of both the Miami River sites and the ocean influenced site ($p= 7.649e-03$).

A canonical correspondence analysis returned R^2 values for salinity, temperature, percent dissolved oxygen, nitrate/nitrite, and total phosphorus as 0.050, 0.063, 0.073, 0.082, and 0.086 – respectively. For the same test, the Akaike information criterion (AIC) –to measure goodness of fit– returned as 323.69, 322.73, 322.14, 321.67, and 322.00 (Fgure 9). Salinity had the highest AIC score, and total phosphorius had the highest R^2 value. The salinity curves for all site types followed the same general pattern (Figure 10). Salinity at the ocean influenced site was the most stable, averaging around 34. The more confined body of water generally indicated higher variability in salinity. For example, at the canal site, the salinity was the most varable –ranging from 2-25. Looking at total phosphorus, the same patern of open waters being more stable and more confined waters being more varable is seen (Figure 11). A sharp spike in total phosphorus, acompinied by a decline in salinity, was observed in December.

Taxonomic data were transormed for rank abundance. The top 20 most abundent taxa for each site type were visualized with stacked bar grahs and hirearcheral charts. The stacked bar graphs were generated for Order (Figure 12) and Family level (Appendix 3) taxonomy. The

198 lowest identifyable taxa was used on the hirearcheral chart. In most caseses this was Family level
199 (Appendix 4).
200



201

202 **Figure 5′:** shows the α-diversity by site type. Alpha diversity at the canal and canal mouth sites are statistically the
203 same. Alpha diversity at the Ocean influenced and Bay sites are statistically the same (p=6.406e-05). Alpha diversity
204 was plotted on a graph and a boxplot was overlaid. Each point on the plot represents one sampling event. The thick
205 black line with in the box represents the average α-diversity for the site. The higher the line, the higher the average α-
206 diversity.
207

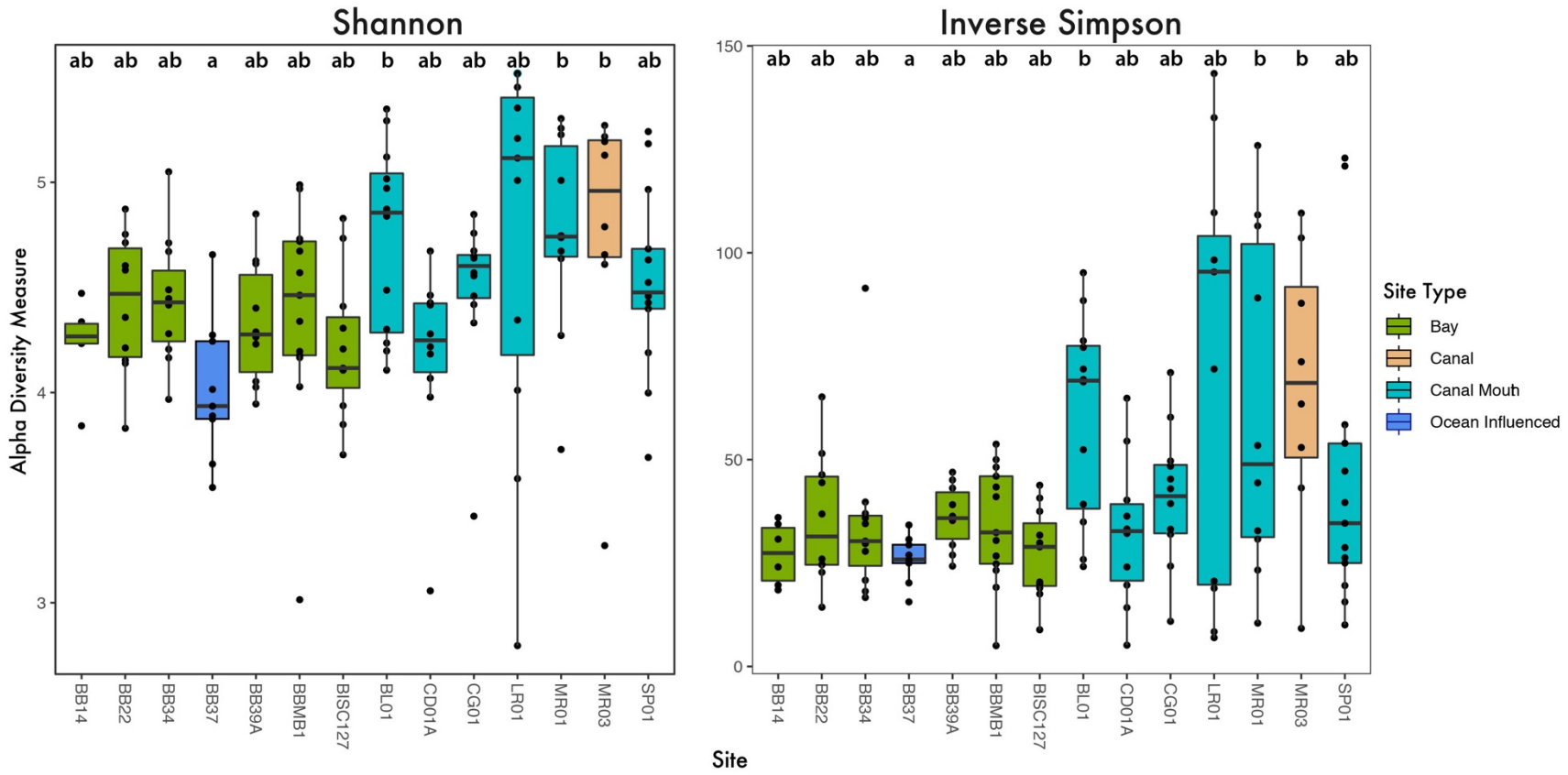
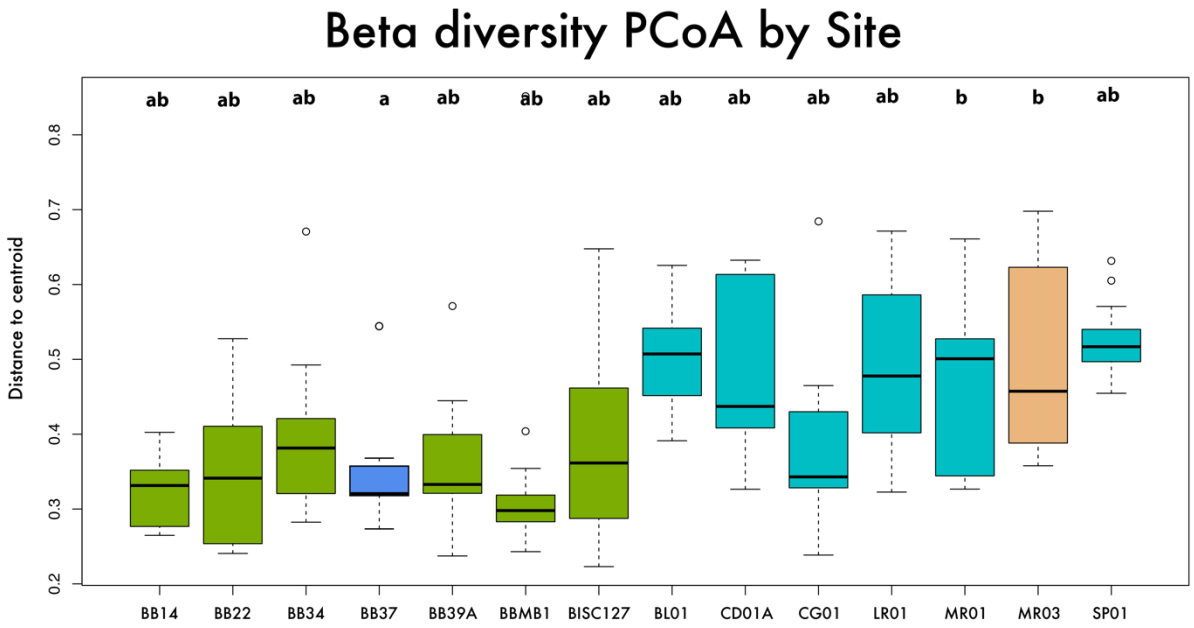


Figure 6′: Box plots illustrating alpha diversity at each site. The sites could be grouped into two statistical groups: **A)** the oceanic site (BB37) **B)** the Miami River sites (MR01 & MR03) and the Mouth of Black Creek (BL01). **AB)** the rest of the sights share properties of both groups (p= 3.728e-03). Color indicates the site type. The Shannon and Inverse Simpson indices were used to measure alpha diversity. Alpha diversity was plotted on a graph and a boxplot was overlaid. Each point on the plot represents one sampling event. The thick black line with in the box represents the average α-diversity for the site. The higher the line, the higher the average α-diversity. These plots were generated using the Phyloseq package in R studio.



209

210 **Figure 8′:** Boxplot generated from a Principal Coordinates Analysis using Bray-Curtis dissimilarity metric for β-
211 diversity. The letters at the top of the chart mark the statistical group each site belongs too. Bacterial β-diversity in
212 the bay followes a similar pattern to the one observed in α-diversity.

213

214 # Discussion

215 Overall, α and β-diversity microbiomes were fairly homogeneous across a majority of the
216 study area. In regard to α-diversity, we found three groups of sites. The first group consists of the
217 canal site and two of the canal mouth sites: MR01, MR03 and BL01, which have statistically the
218 same α-diversity. These three sites are distinct from BB37, the apparently most oceanic
219 influenced site in its own group with relating to α-diversity. The remaining 11 sites have
220 statistically identical α-diversity. Regarding β-diversity the sites once again could be organized
221 into three groups, based on statistical significance. The first group consisted of MR01 and MR03.

The second group was the oceanic influenced site BB37. The remining 11 sites were statistically identical to each other, in regard to β-diversity, meaning they have proportionally the same amount of unique taxa present. As predicted, the ocean influenced site (BB37) had the lowest α- and β-diversity (Figure 5 & 8). This is likely because of the relatively stable conditions at the site, as it is regularly flushed with oceanic water. The mouth of the Little River had the most variability in α and β-diversity (Figure 5 & 8).

Beta diversity (Figure 8) at the two Miami River sites (MR01 & MR02) were identical. The Miami River is the most urban and industrialized river in the study; therefore it stands to reason it would be highly influenced by these land uses. Site BB37 is the most seaward site and it is regularly flushed with oceanic water. Therefore, its reasonable for this site to be an outlier because it would be less influenced by land. The β-diversity at the remaining sites possess traits of a combination of the Miami River and oceanic site. Biscayne bay is regularly flushed with semi-diurnal mixed tides. This mixing combined with the less urbanized land use, outside of Miami's urban core, probably accounts for the patterns observed in this study.

A canonical correspondence analysis revealed that salinity and total phosphorous had the greatest impact on β-diversity. Salinity drove most of the horizontal separation between the samples, and total phosphorous drove more of the vertical separation between the samples – along the axes. The oceanic influenced site (BB37) had the most stable α and β-diversity. Salinity and total phosphorous were also most stable at this site. The increased variability in the diversity metrics at the other stations is attributed to the increased variability of these abiotic factors as well. It should be noted that while salinity and phosphorus significantly affected bacterial community, the strength of the effect was not particularly strong. O'Connell et al. (2018) determined that salinity and temperature were the main factors driving bacterial community.

This study supports the finding that salinity significantly affects bacterial community, but temperature did not seem to play as important role in determining bacterial community.

Abundant & distinguishing taxa

An analysis of similarity (SIMPER) was used to determine which taxa were responsible for distinguishing the sites from each other. There were no Archaea in the top 20 most abundant taxa for each site. Likewise, no Archaea appeared in the SIMPER analysis either. Simper analysis revealed the main taxa responsible for the difference between the oceanic and canal mouth sites were Cryomorphaceae, Rhodobacteraceae, SUP05 cluster, and the NS5 marine group . All these taxa were more abundant at the oceanic site, indicating they are marine taxa. Rocca et al. (2019) suggests that marine taxa may be more resilient in brackish conditions. The preponderance of marine taxa in estuarine conditions in this study supports that finding. The family Cryomophaceae is non-monophyletic (Bowman, 2014). Its members are generally secondary producers and inhabit locations relatively rich in organic carbon (Bowman, 2014). The family Rhodobacteraceae are a common family of bacteria in marine environments (Simon et al, 2017). All species in the family are obligate aerobic, chemoheterotrophs (Rosenberg, 2014). Many marine members of the family use aerobic anoxygenic photosynthesis –meaning they use light to produce ATP, but the process does not result in the release of O_2 (Simon et al, 2017). The five species within the family Rubritaleaceae are not distinguishable based on 16S analysis alone (Rosenberg, 2014). Bacteria from the SUP05 cluster seem to play an important role in the nitrogen and sulphur cycles (Shah et al, 2017). Members of the NS5 marine group are heterotrophs associated with phytoplankton blooms (Seo et al, 2017). NS5 marine group

members possess enzymes for catalyzing many phytoplankton-derived macromolecules (Seo et al, 2017). Overall, the most abundant taxa fit in niches responsible for carrying basic nutrient cycling processes you would expect to find in a marine habitat.

Cyanobacteria are a major, diverse group of photosynthetic bacteria that can inhabit freshwater and a wide range of salinities (Cohen & Gurevitz, 2006). Cyanobacteria were most abundant at the bay sites and least abundant at the oceanic influenced site. Cyanobacteria species can function as aerobic photoautotrophs; anaerobic photo-autotrophs; photoheterotrophs; or chemoheterotrophs (Cohen & Gurevitz, 2006). Many cyanobacteria are known to be N_2 fixers (Arrigo, 2005). In some primarily oligotrophic waters, their contributions to available nitrogen is significant; while in other areas their contribution to N_2 fixation is quite low (Arrigo, 2005). Because the resolution of taxa identified in this study is largely limited to Family level, it is difficult to identify the implications of the presence of various cyanobacteria in the samples. Through SIMPER analysis Cyanobiaceae ASV40 and Cyanobiaceae ASV4 were identified as being a distinguishing taxa between bay sites and the canal site. Cyanobiaceae ASV40 and Cyanobiaceae ASV4 were found predominantly at more saline sites, so presumably they represent saltwater tolerant taxa. *Fluviicola spp.* (in the Family Cryomorphaceae) was also identified through SIMPER analysis as being a distinguishing taxa between The canal (MR03) and oceanic (BB37) sites. The name *Fluviicola* translates as "river dweller" (Woyke et al., 2011). So perhaps unsurprisingly *Fluviicola*it was completely absent from the oceanic site, and present in relatively large numbers at the canal site. Other members of the genus are known to be predominantly fresh water bacteria (O'Sullivan et al., 2005 & Yang et al., 2014). *Fluviicola spp.* was found in moderate abundance at the bay and canal mouth sites, supporting the idea that those sites are influenced by a combination of oceanic and fresh water factors (Appendix 7).

290 Several members of the family Flavobacteriaceae were key in distinguishing the sites from
291 each other. Flavobacteria (family Flavobacteriaceae) are one of the most abundant organisms in
292 aquatic habitats (McBride, 2014). Unsurprisingly flavobacteria were one of the most abundant
293 bacteria observed in this study. While they were still present at fresher sites flavobacteria were
294 much more abundant at more saline sites. No species of flavobacteria are known to be
295 photosynthetic; nearly all species are aerobic chemoorganotrophs (McBride, 2014). Some aquatic
296 flavobacteria are typically not free floating, they rather grow on a surface —i.e. floating organic
297 matter (McBride, 2014). Some are known pathogens for fish (Chen et al, 2017) and possibly
298 sponges (Mulheron, 2014). Typically, flavobacteria are associated with flocculent —as such they
299 are important decomposers in aquatic habitats (McBride, 2014).
300 Most taxa were not identifiable to species or genus level, however one relatively abundant
301 taxon was *Shewanella frigidimarina* (Family Shewanellaceae), which was the 8th most common
302 (relative abundance= 1.5%) bacteria at the mouth of the Miami River (MR01). *S. frigidimarina* is
303 capable of using a wide verity of molecules as an electron acceptor in the electron transport chain
304 of cellular respiration, including: oxygen, iron, manganese, uranium, nitrate, nitrite and
305 fumarate (Copeland et al., 2006). Therefore, it is frequently used in bioremediation (Copeland et
306 al., 2006). At the upstream Miami River site (MR03) the most abundant bacteria (1.9% relative
307 abundance) belong to the family Methylococcaceae. Bacteria in this family are chemoautotrophs
308 that metabolize methane (Bowman, 2014). Methylococcaceae are obligate methane and methanol
309 metabolizers. These molecules are their only carbon and energy source as they are unable to use
310 other substrates containing carbon-carbon bonds (Bowman, 2014). These methane loving
311 bacteria play a critical roll in carbon cycling and Earth's homeostatic processes (Bowman, 2014).
312 Methylococcaceae have also been used in bioremediation applications, because of their ability to

sequester large amounts of methane (Bowman, 2014). Donnelly (2018) observed Methylococcaceae in high abundance in urban canals in urban Ft. Lauderdale, FL. While both *S. frigidimarina* and Methylococcaceae are beneficial, their presence in high abundance suggests the location is highly polluted. The Miami River is the most urbanized river in the study, therefor finding bacteria which exploit heavy metals and methane is not surprising.

Bacteria in the family Enterobacteriaceae can be used as an indicator of anthropogenic pollution (Leite et al., 2018). Of the 116 samples that had detectible *Enterococci* (through traditional culture methods) only 11 of those samples had detectable Enterobacteriaceae through 16S analysis. Of those 11 samples 3 were under the EPA limit of 20 MPN per 100mL for *Enterococci* (US EPA, 2012). The discrepancy between culture methods and 16S analysis can likely be attributed to holding time. The holding time for the 16S samples ranged from 24hrs – 120hrs, and likely exceeded the EPA's maximum holding time of 30hrs (US EPA, 1982). *Enterococci* blooms from rain events are typically short lived and sampling strategies should have high temporal resolution, to adequately detect presence of the bacteria (Aranda et al., 2016).

Currents & Hydrology

Water transport in the bay is principally tidal influenced (Wang, et al., 2003). However small subtidal currents that are not easily measured, strongly influence residence time of water (Wang, et al., 2003). Wind over the shallow bay follow two distinct seasonal patterns (Wang, et al., 2003). Prevailing winds in the summer are gentle Southeasterlies. In the winter winds are generally Southeasterly, but stronger, and they are occasionally interrupted by clockwise rotating winds associated with passing cold fronts (Wang, et al., 2003). Tides in the bay are mixed-semi-

diurnal; having two high tides and two low tides each day —with the two highs being of unequal zenith and the two lows of unequal nadir (Smith, 2001). The tidal range in Biscayne Bay is well below 1.0m (Smith, 2001). As one moves south in the bay, tidal range decreases (Wang, et al., 2003). The region of the bay north of Key Biscayne sees less exchange with the ocean. Region of the bay between Key Biscayne and the Ragged keys is, for the most part, unencumbered by islands and therefore is well flushed with oceanic water.

Precipitation is the dominant source of freshwater to the bay; followed by canal input and ground water discharge (Stalker et. al, 2009). The overall volume of water introduced to the bay varies from the wet to dry seasons, but the ratio of water introduced by these three sources remains constant (Stalker et al., 2009). Historically, the volume of groundwater discharged into the bay was much higher than it is today (Stalker et al., 2009; Cantillo et al., 2000). This is mainly due to anthropogenic alteration of the water table (Stalker et al., 2009; Cantillo et al., 2000). Over the study period Black Creek, Snapper Creek, Miami River, and Little River were each responsible for delivering hundreds of millions of cubic meters of fresh water into the bay. The Cutler Drain and Coral Gables Waterway conducted much less water –on the order of tens of millions of cubic meters (Appendix 11).