EVALUATION OF THE CONTRIBUTION METAGENOMIC SHOTGUN SEQUENCING HAS IN ASSESSING POLLUTION SOURCE AND DEFINING PUBLIC HEALTH AND ENVIRONMENTAL RISKS

by

Brian Mercer

A Thesis Submitted to the Faculty of

Charles E. Schmidt College of Science

In Partial Fulfillment of the Requirements for the Degree of

Master of Science

Florida Atlantic University

Boca Raton, FL

August 2020

EVALUATION OF THE CONTRIBUTION METAGENOMIC SHOTGUN SEQUENCING HAS IN ASSESSING POLLUTION SOURCE AND DEFINING PUBLIC HEALTH AND ENVIRONMENTAL RISKS

by

Brian Mercer

This thesis was prepared under the direction of the candidate's thesis advisor, Dr. Nwadiuto Esiobu, Department of Biological Sciences, and has been approved by all members of the supervisory committee. It was submitted to the faculty of the Charles E. Schmidt College of Science and was accepted in partial fulfillment of the requirements for the degree of Master of Science.

SUPERVISORY COMMITTEE:

Nwadiuto Esiobu, Ph.D.
Thesis Advisor

James X. Hartmann (Aug 20, 2020 15:32 EDT)

James Hartmann, Ph.D.

Donald S McCorquodale (Aug 22, 2020 12:30 EDT)

Don McCorquodale, Ph.D.

Sarah Milton (Aug 22, 2020 14:36 EDT)

Sarah Milton, Ph.D.
Interim Chair, Department of Biological Sciences

Teresa Wilcox, Ph.D.
Interim Dean, Charles E. Schmidt College of Science

Robert W. Stackman Jr., Ph.D.
Dean, Graduate College

August 24, 2020
Date

ACKNOWLEDGEMENTS

The author wishes to express sincere gratitude to all his committee members Dr. James Hartmann and Dr. Don McCorquodale for their guidance throughout this process. The author would like to especially thank his advisor Dr. Nwadiuto Esiobu for her guidance, patience, helpfulness, and mentorship during this program. The author is grateful for Lyle Johnson and Florida Spectrum Environmental Services for financial support on this work. Last but not least, the author expresses great thanks to his family and wife for their emotional support.

ABSTRACT

Author: Brian Mercer

Title: Evaluation of the Contribution Metagenomic Shotgun Sequencing has in Assessing Pollution Source and Defining Public Health and Environmental Risks

Institution: Florida Atlantic University

Thesis Advisor: Dr. Nawdiuto Esiobu

Degree: Master of Science

Year: 2020

State-approved membrane filtration (MF) techniques for water quality assessments were contrasted with metagenomic shotgun sequencing (MSS) protocols to evaluate their efficacy in providing precise health-risk indices for surface waters. Using MSS, the relative numerical abundance of pathogenic bacteria, virulence and antibiotic resistance genes revealed the status and potential pollution sources in samples studied. Traditional culture methods (TCM) showed possible fecal contamination, while MSS clearly distinguished between fecal and environmental bacteria contamination sources, and pinpointed actual risks from pathogens. RNA MSS to detect all viable microorganisms and qPCR of fecal biomarkers were used to assess the possible environmental risk between runoff drainage canals and a swamp area with no anthropogenic impact. Results revealed higher levels of pathogenic bacteria, viruses, and virulence and antibiotic resistance genes in the canal samples. The data underscore the potential utility of MSS in

precision risk assessment for public and biodiversity health and tracking of environmental microbiome shifts by field managers and policy makers.

EVALUATION OF THE CONTRIBUTION METAGENOMIC SHOTGUN SEQUENCING HAS IN ASSESSING POLLUTION SOURCE AND DEFINING PUBLIC HEALTH AND ENVIRONMENTAL RISKS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

This study was conducted in two parts: the first aimed to compare and contrast aspects of traditional culturing methods for monitoring microbial water quality, versus an emerging method using metagenomic shotgun sequencing (MSS). The second part was to verify the methodology and use RNA MSS to characterize the difference in microbial communities in drainage canals versus a pristine swamp area and gauge the environmental impact. Metagenomic shotgun sequencing provided an advanced in-depth look at the microbiome of each environmental water sample, thus providing a more accurate risk assessment of each site. Each sample of metagenomic data was mined for specific parameters such as: quantification of all microorganisms (bacteria, viruses, fungi, protists, and virulence/antibiotic resistance genes). This data was then compared to traditional culturing quantification of fecal coliforms and *Enterococci spp.,* as well as qPCR for fecal biomarkers, and final analysis of water quality was compared between all methods. In the second part of this study we used extracted RNA MSS to capture all of the viable organisms present in each sample. We first verified this novel methodology by analyzing a duplicate sample, and then using it to gauge the differences in microbial communities of drainage canal samples and a pristine swamp land with no anthropogenic impact.

2.0 CHAPTER 1 METAGENOMIC SHOTGUN SEQUENCING PROVIDES PREVALENCE DATA FOR REAL PATHOGENS, AND SOURCE-TRACKING INDICIES USEFUL IN PUBLIC HEALTH RISK ASSESSMENT OF ENVIRONMENTAL WATERS

2.1 Abstract

State-approved membrane filtration (MF) techniques for water quality assessments were contrasted with metagenomic shotgun sequencing (MSS) protocols to evaluate their efficacy in providing precise health-risk indices for surface waters. Samples from a freshwater receiving pond (ABI-1002) and two upstream storm water ditches (ABI-1003) and (ABI-1004) yielded alarmingly high Fecal coliform MF densities of 220, >2000 and >2000 CFU/100ml respectively. The indicator, Enterococcus bacteria exceeded allowable limits in all but the equipment control (ABI-1001). Using MSS, the relative numerical abundance of pathogenic bacteria, virulence and antibiotic resistance genes revealed the status and potential pollution sources of each ditch. High levels of Shigella sp. (0 (ABI-1001), 4945 (ABI-1002), 55,008 (ABI-1003), and 2221 (ABI-1004) genomic reads/100ml) correlated with virulence genes and antibiotic resistance genes found in fecal samples for ABI1003 and not ABI1004. Traditional culture methods (TCM) showed possible fecal contamination in two of the four samples, and no contamination in the others. MSS clearly distinguished between fecal and environmental bacteria contamination sources, and pinpointed actual risks from pathogens. Our data underscore

the potential utility of MSS in precision risk assessment for public and biodiversity health and tracking of environmental microbiomes shifts by field managers and policy makers

2.2 Introduction

For over 50 years, microbiological risk assessment of environmental samples has been based on monitoring the prevalence of indicator organisms that are generally not harmful, but indicate the possible presence of pathogenic bacteria, viruses and protozoans (EPA 822-R-10-005, 2010). The widely used indicator bacteria include easily cultured heterotrophic bacteria—Enterococcus spp. , Escherichia coli , Fecal coliforms and more recently Bacteroidetes sp.p, Clostridia spp., some phages and bio-marker genes (Esiobu et al., 2004; Wade et al., 2006; Noble et al., 2003; Coakley et al., 2015). The current regulations of surface water quality standards for Florida are referenced in document 62 - 302 (Water Quality Standards Variances. Epa.gov). Enteric organisms such as E. coli, Fecal coliforms , and Enterococci spp. are the only well-regulated indicators of microbiological health risks. The safety thresholds vary and are based on the class of water and organism being tested. Issues related to the sensitivity and specificity of the indicator detection systems have been continually revised and improved with advances in chromogenic media and one-step assays (Odonkor & Ampofo, 2013; Ferguson et al., 2013). However, numerous limitations remain. Reviews and discussions of these challenges abound in literature (Rochelle-Newall et al., 2015; Tan et al., 2015; Figueras & Borrego, 2010; Evangelista & Coburn, 2010), and include but not limited to questions about the suitability of enteric indicators for respiratory/skin illnesses, the interpretation of prevalence data in non-point source environments, the difficulty in distinguishing between environmental strains and real indicators (fecal strains), the lack of indicators for biodiversity in preserved areas and finally, the utility of the numbers (Esiobu et al., 2013).

The US Environmental Protection Agency (EPA) manual for monitoring (EPA 5.11 Fecal Bacteria, 2012; EPA-820-R-14-010, 2014; EPA-820-R-14-011, 2014) justified the indicator system because "it is difficult, time-consuming, and expensive to test directly for the presence of a large variety of pathogens; water is usually tested for coliforms and fecal Streptococci instead". In some instances, these relatively inexpensive culture techniques provide data considered sufficient to accurately assess (at least in part) the public health risk of water samples. On the other hand, interpretation of results from non-point sources of pollution is not always clear-cut. To address the many challenges that remain unsolved, emerging technologies such as the metagenomic shotgun sequencing, which can detect virulence genes and all microbial life forms- viruses, bacteria, fungi and protozoans in a single assay are being developed. These permit a direct rather than an indirect assessment of public health risks.

Techniques such as amplicon metagenomic analysis using 16S rRNA pyrosequencing (Gomez-Alvarez et al., 2012) are being used to rapidly and effectively monitor different disinfection treatments of drinking water samples. Similarly, Cabral et al. (2018) employed sequencing technologies to successfully characterize microbial communities in water ways while other recent studies (Mohiuddin et al., 2017; Roy et al., 2018; Cocolin et al., 2018; Li et al., 2018) have demonstrated the utility of next generation sequencing techniques and metagenomic shotgun sequencing in gaining insight to the complex problems associated with surface water quality, and direct detection of possible pathogenic organisms in beach sand. Furthermore, whole genome sequencing has been employed to study functional genes in potable water treatment systems and associated biofilms (Douterelo et al., 2018). Wide-scale use of these emerging sequencing

technologies for environmental monitoring will require experimentally modelled interpretations of the big-data generated, as well as comparative analysis of the results with validated culture methods and other epidemiological indices. In this study we determined and compared the microbiological water quality of four samples using the traditional EPA standard methods and metagenomic shotgun sequencing of genomic DNA extracted directly from water to evaluate their relative efficacy in predicting actual health risks to the public and biodiversity in the environment.

2.3 Methods

2.3.1 Sampling

Four water samples were collected in September of 2016 aseptically from: (A) deionized water used to ensure sampling equipment was contaminant free (ABI-1001)—(quality control), (B) a receiving freshwater pond used for recreational activities (ABI-1002), a storm water ditch whose effluent enters the freshwater pond (ABI-1003), and a second storm water ditch located the furthest from the pond and whose effluent merges with those of ABI-1003 before emptying into the pond (ABI-1004.) The exact locations of sample points are kept confidential for privacy purposes. All samples were collected at the same event after rainfall to ensure storm water ditches held water. Fecal coliform membrane filtration and Enterococcus membrane filtration samples were collected in 100 ml sterile vessels containing sodium thiosulfate, sufficient to neutralize any free chlorine present. Metagenomic shotgun sequencing samples were also collected from the same locations, but in 1 L amber bottles (sterilized by rinsing with 10% bleach) with no preservatives. All samples after collection were placed immediately on ice and stored at 4˚C in the laboratory until analysis.

2.3.2 Fecal Coliform Enumeration

Fecal coliform bacteria were enumerated following the standard methods SM9222D method (Clesceri et al., 1998). Filtration volumes were based on multiple years of historical data for these sample sites, and results were reported in Colony Forming Units (CFU)/100ml. Two dilutions were prepared for each sample (10 ml and 50 ml) following NELAP protocol of recovering 20 - 60 CFU. Blanks were also prepared to ensure sterility of membrane filter equipment using sterile deionized (DI) water. All samples were processed within the Florida Department of Environmental Protection (DEP) guidelines of 8 hours after collection.

2.3.3 Enterococci Enumeration

Enterococcus spp. were detected using the EPA 1600 method for enumeration of Enterococci from water (EPA-821-R-06-009, 2006). Dilutions were carried out as in Fecal coliform above, based on multiple years of historical data for these sample sites, and results were reported in CFU/100ml. Two dilutions were prepared for each sample (10 ml and 50 ml) following Standard Methods protocol of recovering 20 - 60 CFU. Blanks were also prepared to ensure sterility of membrane filter equipment using sterile DI water. All samples were processed within the Florida DEP guidelines of 8 hours after collection.

2.3.4 Metagenomic Shotgun Sequencing

Five hundred (500) ml of each water sample was thoroughly mixed and vacuum filtered through 0.2 µm pore-sized polycarbonate filters, and a replicate sample filtered with the remaining 500 ml. All cells were washed out and concentrated into 5 ml of sterile DI water by placing the polycarbonate filter into a 50 ml centrifuge tube containing 5 ml of sterile DI water and vortexing. The suspension of cells was then harvested by ultra-

centrifugation at 15,000 rpm in 1.5 ml aliquots. The pelleted cells were pooled into one tube before lysis and DNA extraction using the Qiagen DNeasy protocol according to manufacturer instructions (Qiagen, Valencia, CA). Metagenomic DNA concentration, purity and quality were then measured with NanoDrop 2000c and gel electrophoresis. After standardized library preparations, sequencing was outsourced to COSMOSID® (1600 E Gude Dr. Rockville, MD 20850 United States). DNA libraries were constructed using COSMOSID® proprietary Library Prep Kit. The PCR products were purified using 1.0Å~ speed beads and eluted in 15 µL of nuclease-free water, then quantified by PicoGreen fluorometric assay. Libraries were then pooled and loaded onto a high sensitivity chip run on a Caliper LabChipGX (Perkin Elmer, Waltham, MA) to estimate the size. Libraries were sequenced using Illumina NextSeq/HiSeq platform. Sequencing primers targeted a full repertoire of bacteria, fungi, viruses, and protozoan parasites; and also included virulence factors and antibiotic resistance markers. All sequences obtained were quality filtered and subsequently aligned and classified using the GenBook® database, a highly curated database comprising over 150,000 microorganisms' genomes and gene sequences. This database consists of both private and publicly available genomes from sources such as NCBI-RefSeq/WGS/SRA/nr, PATRIC, M5NR, IMG, ENA, DDBJ, etc. Sequence data were deposited in the NCBI Sequence Read Archive (SRA) under accession number SRP259849.

2.4 Results

Results obtained from this study are all reported as CFU/100ml unless otherwise stated, and the samples are labeled as ABI1001-ABI1004, with ABI1001 being the equipment

control (blank) as previously mentioned. Figure 1 displays the results obtained for traditional indicators—Fecal coliforms

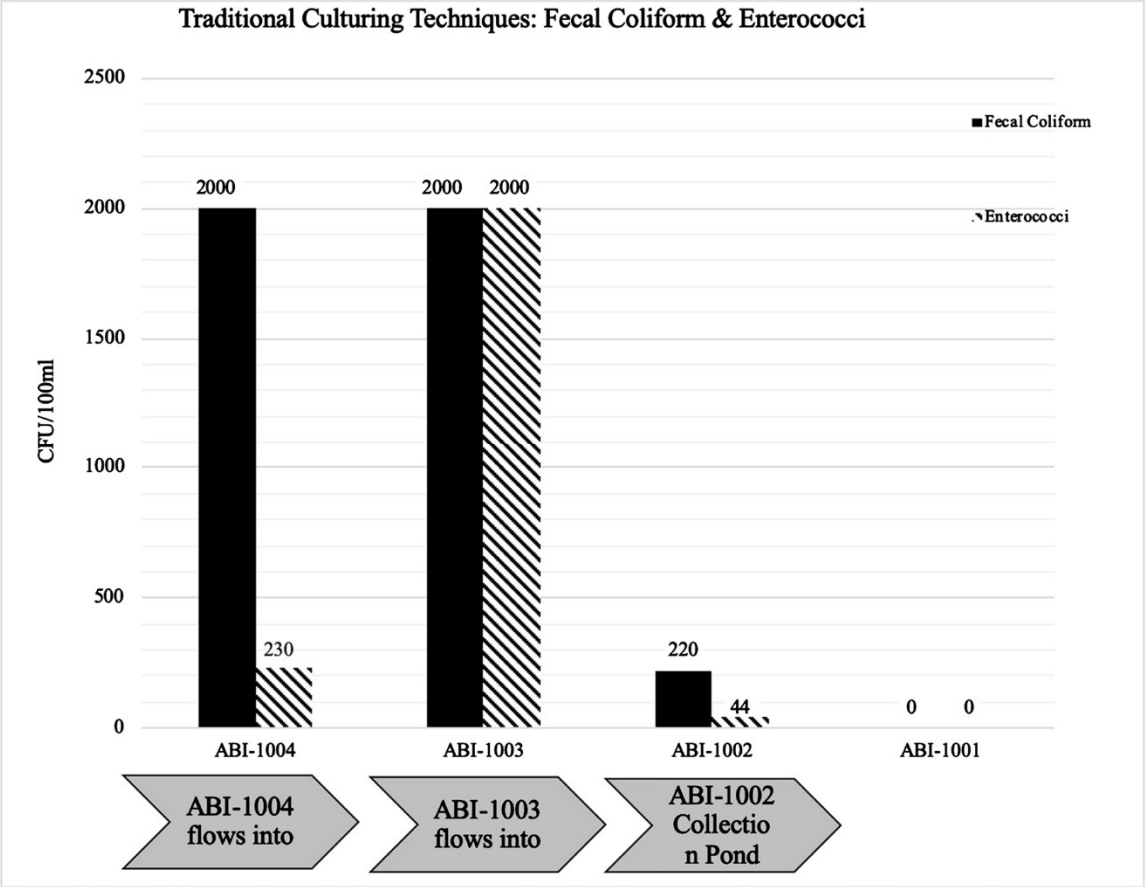

Figure 1. Population of classic indicator bacteria (Fecal coliforms and Enterococci) in a receiving pond and storm water ditches using the traditional culture techniques CFU/100ml (Colony Forming Units per 100 ml) based in standard Methods guidelines. ABI1001 is an equipment/reagent QC blank.

and Enterococci using traditional culture techniques—SM9222D and EPA 1600. The lowest run dilution for these samples was 10 ml, making the highest possible detection limit to be 2000 CFU/100ml. As such all results that are at 2000 CFU/100ml are presumed to be greater than 2000 CFU/100ml, because the plates were too numerous to count. Results reported for these culturing techniques are based on the most accurate number calculated from the volume of sample filtered. It is noteworthy that while Fecal coliforms (a nearly ubiquitous enteric group of bacteria in animals) exceeded detection

limits in ABI-1003 and 1004; enterococcal levels were contrasting; being significantly less in the latter. The numbers in the receiving pond and control samples were within expected levels.

Figure 2 shows the distribution of the traditional indicator organisms in each sample using metagenomic shotgun sequencing calculations (genomic reads/100ml). Indicator organisms labeled as coliform represents all the genera of the coliform group (Citrobacter, Enterobacter, Klebsiella, and Escherichia) found in each sample. The Enterococcus value is the sum of all species of Enterococcus found in each sample. Species for each category were added up and calculated based on relative abundance, then converted to genomic reads/100ml. MSS numbers sharply contrast the values obtained from culture techniques. Whereas ABI 1003 and 1004 showed a similar trend in Figure 1, MSS detected more than 24,000

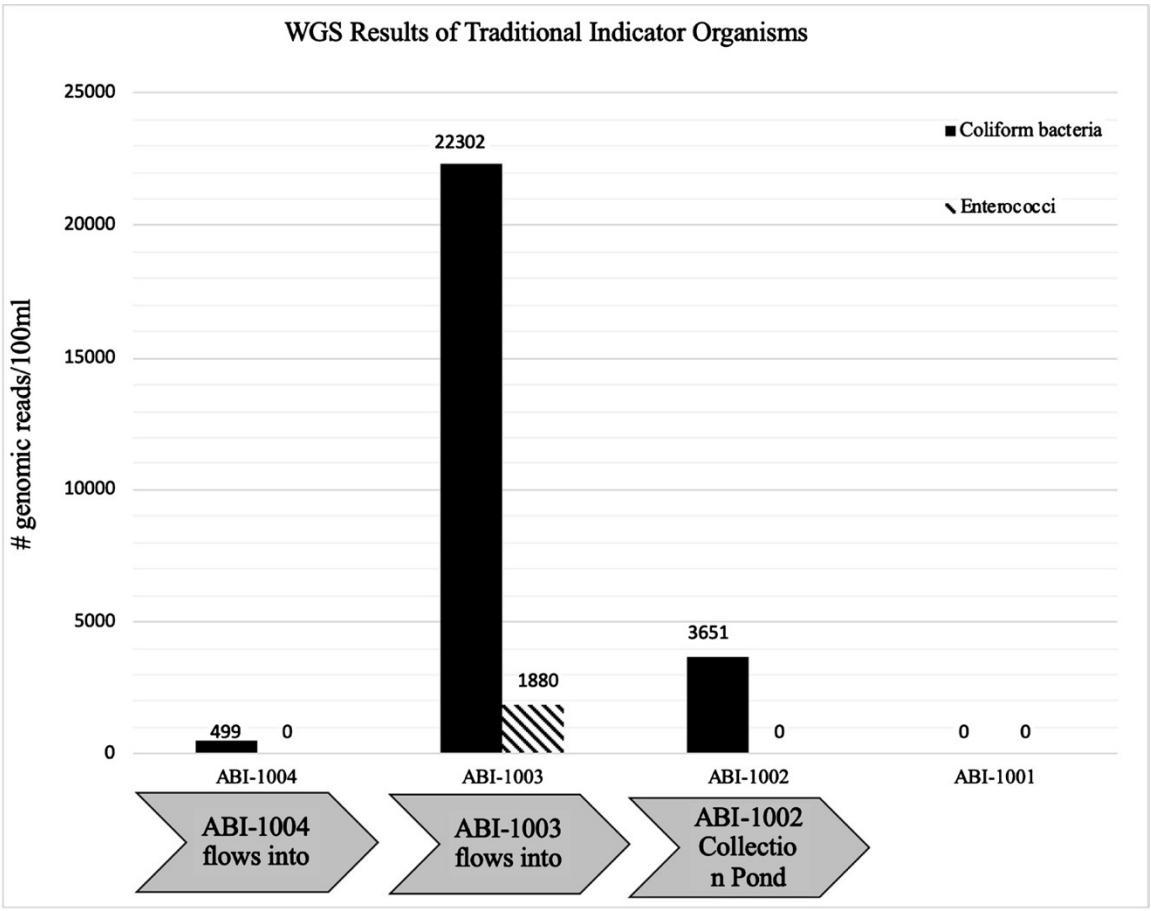

Figure 2. The traditional indicator organisms detected using metagenomic shotgun sequencing MSS. All genera of the Coliform group, and all species of the Enterococcus genus were summed up for a final count.

Fecal coliforms in ABI 1003; only 499 for ABI 1004 and more than 3000 for the receiving pond. Although the equivalency of genomic reads and colony forming units is not necessarily linear, the remarkable difference between both techniques underscore the challenge with sensitivity and specificity of risk assessment assays.

In Figure 3, the prevalence of some potentially pathogenic bacteria species found in each sample is presented in genomic reads/100ml of sample to allow comparisons with results obtained from traditional methods. Shigella sp. were detected at the following densities (0 (ABI-1001), 4945 (ABI-1002), 55,008 (ABI-1003), and 2221 (ABI-1004) genomic reads/100ml). With the exception of Pseudomonas aeruginosa which is ubiquitous in

nature, residing in soils, water, and human skin; sample ABI 1003 contained 100 to 1000 times more potential pathogens than ABI 1004 even though both samples were incorrectly rated of similar pollution trend based on traditional indicator culture approach (Figure 1). Most of the pathogens detected in high numbers include Vibrio cholerae and Staphylococcus spp. and other enteric organisms. None of the common pathogens was detected in the receiving pond.

In addition to specific pathogens, indicator genera, a single assay of MSS also detects virulence and antibiotic resistance genes. In Figure 4(a), the distribution and abundance of commonly known bacterial virulence genes found in the samples

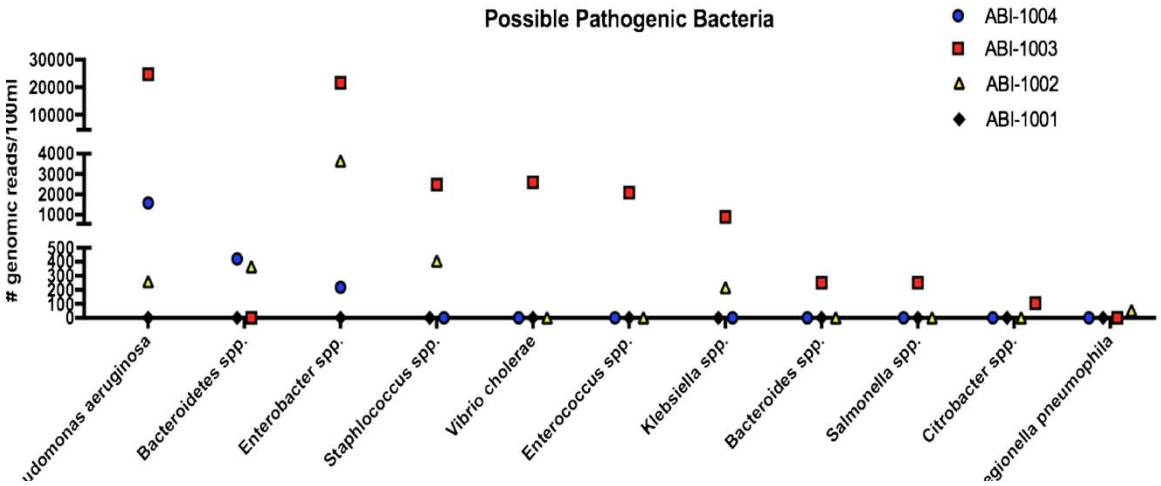

Figure 3. A representation of the potential pathogenic bacteria in each sample expressed in genomic reads/100ml of sample. Those expressed as spp. had multiple species detected, and all were added together

are displayed Sample ABI-1003 contained additional virulence genes, too numerous to include in one figure. Figure 4(b) displays a strikingly important distribution of antibiotic resistance genes associated with the samples. ABI-1003, again, had much more diversity and prevalence of antibiotic resistance genes than could be included in this chart. There is a clear correlation between virulence/antibiotic resistance genes and the density of Shigella sp. and other potential pathogens presented in Figure 3. The absence or relatively low density of these genes in sample ABI1004 is congruent with results

displayed in Figure 2 and Figure 3 but sharply contrast traditional culture results which recorded high abundance of Fecal coliforms and enterococci (Figure 1).

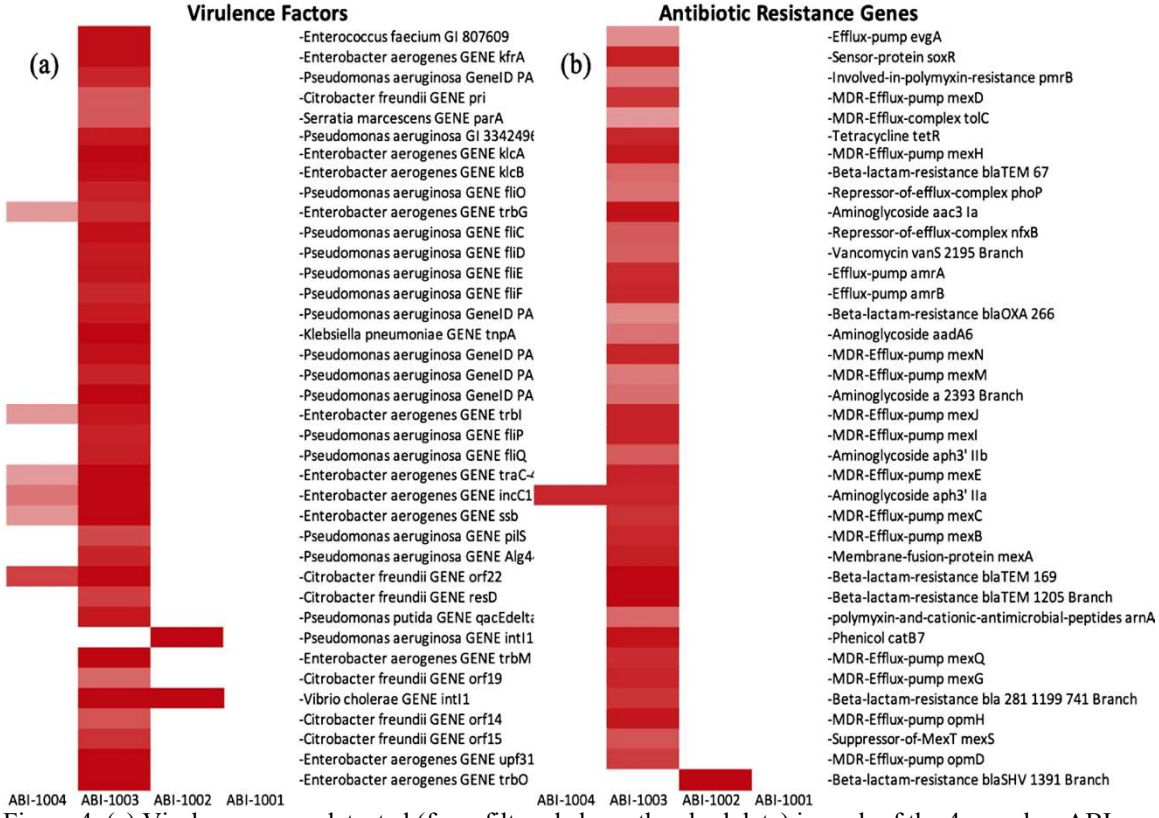

Figure 4. (a) Virulence genes detected (from filtered above threshod data) in each of the 4 samples. ABI-1003 represents the top 10% of virulence genes detected. The more maroon the bar is, the higher incidence that gene was detected. (b) A represents the top 50% of antibiotic resistance genes detected. The more maroon the bar is, the higher incidence that gene was detected.

The relative occurrence of bacteria phyla detected in all samples is shown in Figure 5. This reveals the relative community structure and diversity of the environmental water sample. Each value is the sum of all the taxonomic ranks for that phylum, present in the sample. The receiving pond-ABI 1002 and sample ABI 1004 have the highest natural bacteria diversity while ABI1003 seemed enriched with Proteobacteria, many of which are enteric organisms.

A summary of the various parameters derivable from the TCM and the MSS are presented in Table 1. Whereas TCM yields numerical values of mere indicators of the

potential presence of pathogens, the MSS quantifies indicators, pathogenic bacteria and viruses as well as virulence and antibiotic resistance genes together provide a highly reliable measure of the not just the quality but also the source of pollution.

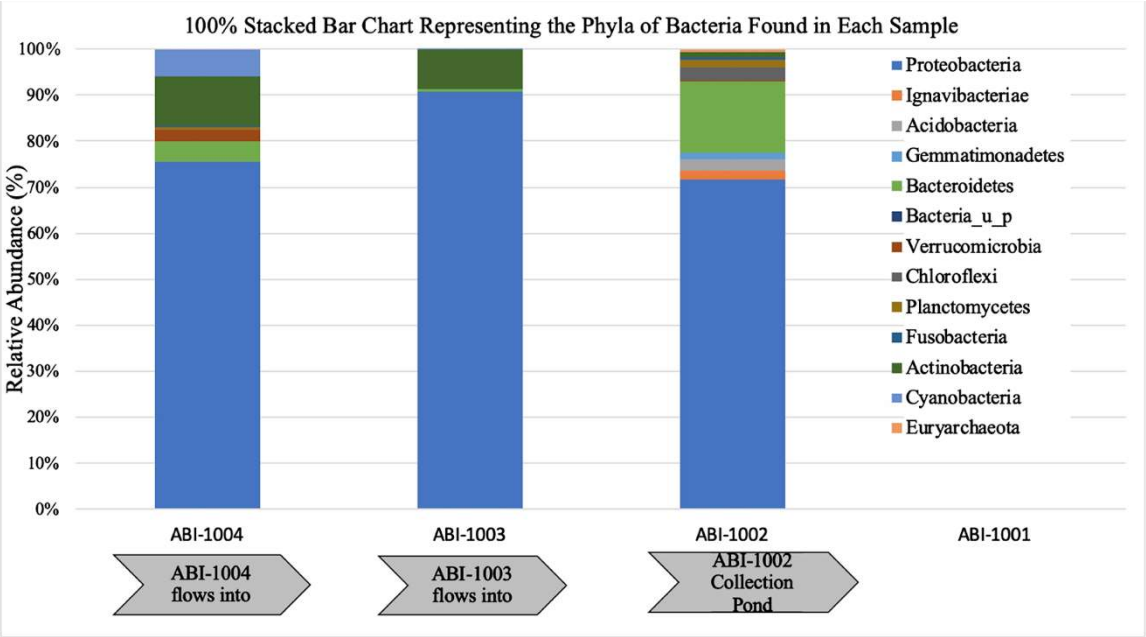

Figure 5. Bacteria community structure of a fresh water receiving pond and two storm water drains and a control, all bacteria phyla detected are expressed in percent relative abundance of total genomic reads. These results correspond to the total 500 ml of each sample.

2.5 Discussion

2.5.1 Traditional Culture Methods

Using traditional culturing methods alongside metagenomic shotgun sequencing allowed us to compare these methods for the first time to understand the microbial community indices versus membrane plate count results. The traditional culturing methods (Figure 1) showed that samples ABI-1003 and ABI-1004 exceeded the Florida Fecal coliform limit of 800 CFU/100ml on any one day) as per SM 9222D for Fecal coliform organisms (Clesceri et al., 1998). Furthermore, ABI-1002 showed relatively low concentrations of traditional indicator organisms, and traditional test results were within the acceptable criteria of Florida surface water limits. These methods are beneficial due to their cost

effectiveness and quick production of results. Although this information is quick and easy to obtain, traditional indicator organisms tend to adapt and survive for prolonged periods in the humid and warm Florida climate (Bonilla et al., 2007; Hartz et al., 2008). Besides, the phenotypic and biochemical assays are prone to yield false positive results due to the enormous versatility of environmental bacterial metabolism. Also, many environmental factors can influence the results of these methods such as: runoff of soil from rainfall, debris entering sampling point, and natural conditions such as shade from sunlight.

2.5.2 Metagenomic Shotgun Sequencing

It was possible to obtain a plethora of data from metagenomic shotgun sequencing not discernible from the traditional culture methods used in this study. MSS results enabled the construction of a clear risk assessment by more than one layer of evidence: human pathogens, human pathogen relative abundance, antibiotic resistance genes, virulence genes, viruses, and also all other microorganisms such as environmental bacteria. The relative abundance of human associated pathogens can indicate source and potential risk of an environmental sample (Wade et al., 2006). All of the parameters obtained in a single test (including strain level detection) allowed for potential source-tracking of any fecal pollution present in a given sample. Metagenomic shotgun sequencing can also be applied to define a microbiome for particular environmental water bodies. Understanding the water body microbiome condition could allow for fewer repeats of traditional culture testing and create a more comprehensive fingerprint for long term monitoring of changes in user behaviors and climate.

Metagenomic shotgun sequencing data for traditional indicators are depicted in Figure 2. Agreeing with the TCM (Figure 1) MSS showed very high levels of indicator organisms

in ABI1003. However, it sharply contrasted the results for samples ABI1002 and ABI1004, where elevated levels in the TCM were low in MSS and vice versa. Furthermore, where traditional methods limit total indicator bacteria levels to an estimated maximum value (Figure 1), the MSS data in Figure 2 provided numeric levels, exceeding 20,000 reads/100ml in some cases. Figure 2 also demonstrates the pattern of indicator bacteria over space, where higher levels were found in ABI-1003 and present in lesser amounts at the pond (ABI-1002). Indeed, no possible pathogenic bacteria were present at the pond (ABI-1002) at higher levels than were found in the two storm water ditches (Figure 3), and in several cases, pathogenic bacteria strains found in the ditches were not present in the pond. This degree of specificity could be very useful for environmental or public health scientists investigating non-point pollution sources. Data in Figure 3 showed some potentially harmful bacteria in the samples, but ABI-1003 contained the highest density of pathogenic bacteria. Organisms such as Vibrio cholerae have been found in known contaminated surface waters in Haiti (Roy et al., 2018) using MSS technology. This organism was found in high quantities in sample ABI-1003 suggesting some source of fecal contamination. In Figure 4(a) and Figure 4(b), sample ABI-1003 clearly contained the most virulence and antibiotic resistance genes. This is a striking confirmation of the high prevalence data of actual pathogens for ABI-1003 while most of the other samples contained environmental microorganisms (Figures 1-3). The most remarkable evidence of the potential fecal contamination of sample ABI 1003 is the abundance and type of virulence and antibiotic resistance genes associated with human enteric organisms. A correlation plot based on sequence reads shows that virulence genes had a low correlation with viral load except for the fecal polluted sample ABI-1003. On

the other hand, a high pathogen density almost always correlated with elevated total bacterial population in exactly the same way as low pathogen counts. While this is not surprising because most of these organisms (pathogen and non-pathogens) are largely heterotrophic, the result underscores the problem of relying on just numbers in risk assessment. Distinguishing fresh pollution by pathogenic strains of bacteria and physiologically similar ecotypes requires rigorous discriminatory tests like the MSS which provides multiple indices of the water quality. Most of the bacteria found in sample ABI-1003 in Figure 3 are enteric bacteria of fecal origin. Close-up scrutiny revealed that only one Bacteroides spp. was found in ABI-1003, the classic Bacteroides fragilis str S6L5. This strain of Bacteroides fragilis is also possibly associated with human gastro-intestinal disease, although research is still being conducted. No specific human Bacteroides sp. was detected in samples ABI-1002 and ABI-1004. The only Bacteroides were members of the Bacteroidetes phylum with no specific annotation, suggesting environmental animal sources. This deduction is fully congruent with the virulence and antibiotic resistance profiles in Figure 4 which depicts a clear-cut difference between the samples, with only ABI-1003 having more virulence genes associated with enteric bacteria. Taken together, all the results from the MSS hone in on a human fecal bacteria presence at site ABI-1003.

Analysis of taxonomic data from the sequence results of ABI-1002, contrasted the results obtained from the traditional culture methods. It is interesting to note in Figure 2 that no Enterococcus spp. were reported in the major taxa found in samples ABI-1002, and ABI-1004 using metagenomic sequencing, while traditional methods detected low numbers of Enterococci This is possibly due to a low sensitivity of the MSS assay where only a few

nanograms of DNA is used for sequencing, regardless of amount of sample DNA available. This limitation could result in organisms and entities in very low relative abundance not being detected. Another interpretation of the apparent discrepancy is that the organisms were well below risk levels and within safety thresholds, and so not detecting such microbes would not alter the utility of the MSS. Nevertheless, experimental determination of the actual threshold sensitivity test will enable its large-scale application in food and environmental quality control.

2.5.3 Benefits and downfalls of TCM and MSS

Unlike TCM, metagenomic shotgun sequencing allows for coupling multiple layers of risk assessment indices obtained concurrently to make predictions with a very high level of confidence. The profile of virulence genes detected in ABI 1004 is more diverse and abundant than the relatively uncontaminated ABI 1002, and both samples were completely devoid of the genes present in ABI 1003. When coupled with data in Figure 3, it becomes apparent that the former contained non-pathogenic naturally occurring organisms. Given the larger dataset of pathogenic bacteria, virulence genes, and antibiotic resistance genes associated with human enteric organisms, the MSS flags ABI-1003 as a suspect.

Table 1 is a summary of the detection capabilities of the traditional methods vs. metagenomic shotgun sequencing. Despite its versatility and specificity, another limitation of metagenomic shotgun sequencing of total DNA from a sample is the possibility of sequencing DNA from dead non-viable cells. This can skew assessments of public health risks and source-tracking. Notwithstanding, coupling information from viruses, virulence factors and antibiotic resistance should allow appropriate conclusions

to be reached. Sequencing environmental RNA instead of DNA, in order to establish only viable organisms in the samples would eliminate any uncertainties. Also increasing the depth of sequencing could increase sensitivity of the MSS technique (Rashkin et al., 2017).

Test Protocol	Metagenomic Shotgun Sequencing				EPA Approved Traditional Culturing Techniques	
Expected Results	Levels and diversity of Traditional Indicator Bacteria, Pathogenic Bacteria, Virulent Genes, Antibiotic Resistant Genes				Levels of Fecal Colifom Bacteria, *Enterococci spp.*, Possible Contamination	
Sample ID	**Traditional Indicators (genomic reads/100ml)**	**Known Pathogens (genomic reads/100ml)**	**Total Number of Virulence Genes**	**Total Number of Antibiotic Resistance Genes**	**Fecal Coliform (CFU/100ml)**	***Enterococci spp.* (CFU/100ml)**
ABI1004 (Storm Water Drain)	499	2600	14	1	>2000	230
ABI1003 (Storm Water Drain)	24182	58700	289	70	>2000	>2000
ABI1002 (Freshwater Collection Pond)	3651	4750	6	1	220	44
ABI1001 (Equipment Blank/ Control)	0	0	0	0	0	0

Table 1. Overview of samples and parameters Assessed by detection capabilities of traditional culturing techniques and Metagenomic shotgun sequencing.

There are various benefits and downfalls for both methods tested. Traditional culture techniques can be performed for a couple hundred dollars per sample, providing indicator results for bodies of waters with 24 - 48 hours. Normal laboratory reporting turnaround time for commercial labs for traditional methods ranges from 7 - 21 days. Metagenomic shotgun sequencing of water samples is a niche that is yet unravelling. Low demand for this type of in-depth testing greatly increases the cost of analysis to over a couple thousand dollars for this intense analysis. Turnaround time is about 4 weeks at best currently. Test time and cost will sharply decline as the test becomes more popular. The wealth of information obtained from MSS is hard to ignore in a progressive scientific world where precision and confident prediction of risk are critical. An entire microbiome

of water bodies can be characterized using this method when sampled over a period of time, giving far greater sensitivity for assessing the public health of a body of water, and also providing insight of potential sources of contamination. Once established, any shift in the microbiome of the sample site will provide detail as to what potential problems may have occurred (fecal pollution, protentional algae blooms, nutrient influx, change in natural conditions, etc.). The utilization of new technologies to assess public health impacts from water bodies is an overlooked benefit.

2.6 Conclusion

This study confirmed the gross limitations of the traditional culture gold standard for assessing water quality of surface waters to include false positives for indicator bacteria and lack of source-tracking components. The numbers of real positives did not always correlate with actual pathogen presence. However, the TCM is a relatively inexpensive assay with short turnaround time. Metagenomic shotgun sequencing technology on the other hand provided a powerful resolution for a water body where state-mandated thresholds for safe recreation were exceeded, and for which a non-point or point source was not clearly evident. Multiple risk assessment parameters obtained from the relatively expensive MSS analysis include specific identification/quantification of all microbiological entities, virulence genes and antibiotic resistance genes; allowing for a robust health risk evaluation in one step. In addition, the comprehensive data on the microbiomes of the niche provide important base-line reference for early detection and intervention in cases of anthropogenic or climate change perturbation. The need to define and standardize the sensitivity of MSS (assuring the detection of only viable bacteria, even at low relative abundance) is critical to its wide applications in environmental health

management. Until the high cost of the MSS is tampered, environmental managers could utilize this powerful technology for decision-making on preserved and protected surface waters as well as major drainage and hydrological activities that could impact biodiversity other than humans. In all other cases, a periodic application of MSS could be coupled with routine TCM monitoring of sites, thereby establishing a baseline for MSS and creating a database overtime for sample sites, to have a reference of protentional changes to the microbiome.

3.0 RNA METAGENOMIC SHOTGUN SEQUENCING AND MICROBIAL SOURCE TRACKING TO EVALUATE PUBLIC HEALTH AND ENVRIONMENTAL RISKS OF DRAINAGE CANALS VS. PRISTINE CONSERVED SWAMP AREA

3.1 Abstract

This study aimed to ensure the methodology of RNA metagenomic shotgun sequencing (MSS) as a tool to assess the public health and environmental risk, and also gauge the impact of diverting drainage canals into a pristine swamp area with no anthropogenic impact using this technology to view all viable organisms. First, we analyzed a drainage canal 720A(BCS03) and its duplicate 720B(BCS03 DUP) to ensure the methodology. Results showed that over 95% of the bacterial community was relatively the same. Two bacteria dominated the sample and duplicate *Bdellovibrio bacteriovorus* 13812 Branch and Bacteroidetes Phylum 2536 641 comprising over 88% of the sample. Next, we used this powerful tool to see the differences in microbial community between 9 water samples, 5 drainage canals 349-01(BCS02), 349-02(BCS01), 398-01(BCS06), 398-02(BCS04), 720A(BCS03), 720B(BCS03 DUP) and 4 samples from a pristine swamp area 429-01(KMN), 719-01(2NATDNA J4), 767-01(CBI-MID), 854-01(ESL-OUT). There was a vast difference between the drainage canals and pristine area, there were far more possible risk bacteria, viruses, and virulence factors in the drainage canals. Using this technology to assess the possible risk of diverting the canals into the pristine area showed concerning levels of a Cyanobacteria *Microcystis aeruginosa*, environmental

pathogenic virus, and virulence factors. RNA MSS was confirmed reproducible method and proved an effective technique to detect environmental risk factors in environmental water samples. More monitoring of these samples must occur to determine the risk of diverting the drainage canals into the pristine area, however the initial results seen show possible environmental risk.

3.2 Introduction

Water quality microbial risk assessment has changed little since it has been regulated at the state level. Traditional culture methods (TCM) such as Fecal coliform membrane filtration, Enterococci membrane filtration, and E. coli membrane filtration are among the most used culture methods for indicator organisms (Esiobu et al. 2004, Wade et al. 2006, Noble et al. 2003, Bonilla et al. 2007, Odonkor et al. 2013). These mentioned tests are designed to be quick and cost effective to determine if indicator organisms are present and in what quantity, which could be indicative of the microbial water quality. Molecular DNA technologies have recently increased in popularity (as a research tool) to track the source of fecal contamination using biomarkers (Coakley et al. 2015), and these tests would be paired with traditional indicator methods (TIM). Studies of microbial source tracking (MST) seek to uncover the source of contamination in a body of water either potable or non-potable (Bauza et al. 2019, Kongprajug et al. 2019, Vadde et al. 2019). Many studies have been conducted to characterize the microbial communities of water bodies using 16S sequencing (insert 16S references here), while these methods serve to distinguish the different microbial communities, many of them only characterize the bacteria at the phylum level (Gomez-Alvarez et al. 2012, Cabral et al. 2018). Other studies employ next-gen sequencing techniques to understand surface water quality and

target pathogenic organisms (Mohiuddin et al. 2017, Roy et al. 2018, Cocolin et al. 2018. Li et al. 2018, Breton-Deval et al. 2019). These technologies are useful in understanding the microbial population and can pinpoint specific public health risks however, they simply scratch the surface in terms of fully characterizing the microbiome (bacteria, viruses, parasites, virulence genes, antibiotic resistance genes) of a body of water. In our previous study we used a novel technique (MSS) to assess the microbial risk of 3 water samples. We compared this novel technique to the traditional culture methods, and the overwhelming wealth of information we uncovered was staggering (Mercer et al. 2020). Using MSS we could see all bacteria, viruses, fungi, parasites, virulence factors, and antibiotic resistance genes in each sample. Not only did we see organisms that have not been seen before in public health risk assessment screening, but we were also able to track the source of pollution in one of the 3 samples. The complexity introduced by non-point pollution sources in South Florida's water ways requires well designed microbial risk assessment protocols to evaluate human risks associated with exposure to a given environment. On the other hand, danger to the environment (native plants, and native animals and biodiversity) are not assessed by currently available traditional testing protocols used to monitor public health risk. Protecting the microbiome of a niche is as crucial as the animal sustainability. Bacteria, virus, fungi, and parasites co-exist naturally in environmental water ways and swamps. Introducing new organisms to a well-established balance could affect to the environment in unpredictable ways. For example, some cyanobacteria and fungi have the ability to bloom under favorable conditions, causing repercussions for the entire ecosystem. Viruses and parasites have the ability to

infect plants and animals and could pose a risk to biodiversity if contaminated water sources are inserted without adequate treatment.

In this study we aimed to use RNA MSS as a tool to characterize the microbial communities of different types of bodies of water. Choosing to use RNA would give us an in-depth look of the living microbial community at the time of collection, because RNA is transcribed by living organisms and is not stable for extended periods of time. First, we wanted to establish a baseline microbiome for these sample sites. Next, we wanted to see the vast differences between runoff canals versus a pristine are with no anthropogenic impact. Finally, we wanted to gauge the environmental and public health impact of diverting runoff canal waters into a pristine area.

3.3 Methods

3.3.1 Sampling

In this project, a total of nine samples and one duplicate (DUP) were analyzed. Five samples (designated BCS) from drainage canals and storm water treatment area's numbered 349-01 (BCS02), 349-02 (BCS01), 398-01 (BCS06), 398-02 (BCS04), 720A (BCS03), 720B (BCS03 DUP) while 4 samples were obtained from pristine environments of the reserve, free from any known anthropogenic impact 429-01 (KMN), 719-01 (2NATDNA J4), 767-01 (CBI-MID), 854-01 (ESL-OUT), see Table 2 for description of samples. The exact locations of sample points are kept confidential for privacy purposes. All samples were collected in the month of September 2018, not all on the same day due to location and accessibility. Metagenomic shotgun sequencing (MSS) samples were collected aseptically in 1L amber bottles (sterilized by rinsing with 10% bleach and then autoclaved) with no preservatives. All samples after collection were placed immediately

on ice and stored at 4 °C in the laboratory until analysis. Samples for RNA MSS were filtered and extracted upon arrival to the laboratory.

Sample Numbers	Sample Description	Remarks
349-01(BCS02)	Drainage canal	Baseline data on advanced microbiological quality. Site already has well established traditional method data. RNA shotgun sequencing, and qPCR analysis targeting fecal biomarkers
349-02(BCS01)	Drainage canal	Baseline data on advanced microbiological quality. Site already has well established traditional method data. RNA shotgun sequencing, and qPCR analysis targeting fecal biomarkers
398-01(BCS06)	West feeder canal	Baseline data on advanced microbiological quality. Site already has well established traditional method data. RNA shotgun sequencing, and qPCR analysis targeting fecal biomarkers
398-02(BCS04)	North feeder canal	Baseline data on advanced microbiological quality. Site already has well established traditional method data. RNA shotgun sequencing, and qPCR analysis targeting fecal biomarkers
720A(BCS03)	Canal near STA "Confusion Corner"	Baseline data on advanced microbiological quality. Site already has well established traditional method data. RNA shotgun sequencing, and qPCR analysis targeting fecal biomarkers
720B(BCS03)	Canal near STA "Confusion Corner" Duplicate of sample 720A	Baseline data on advanced microbiological quality. Site already has well established traditional method data. RNA shotgun sequencing, and qPCR analysis targeting fecal biomarkers
429-01(KMN)	Pristine Reserve Areas of the Big Cypress devoid of any known anthropogenic impacts	Baseline data on microbiological quality. RNA shotgun sequencing, and qPCR analysis targeting fecal biomarkers.
719-01(2NATDNAJ4)	Pristine Reserve Areas of the Big Cypress devoid of any known anthropogenic impacts	Baseline data on microbiological quality. RNA shotgun sequencing, and qPCR analysis targeting fecal biomarkers.
767-01(CBI-MID)	Pristine Reserve Areas of the Big Cypress devoid of any known anthropogenic impacts	Baseline data on microbiological quality. RNA shotgun sequencing, and qPCR analysis targeting fecal biomarkers.
854-01(ESL-OUT)	Pristine Reserve Areas of the Big Cypress devoid of	Baseline data on microbiological quality. RNA shotgun sequencing, and

Table 2. Sample ID's and descriptions of environmental samples collected.

3.3.2 RNA Metagenomic Shotgun Sequencing

Metagenomic shotgun sequencing samples were analyzed the same as previously published (Mercer et al. 2020). First all samples were brought to room temperature to ensure microbes were active and producing RNA. Five hundred (500) ml of each water sample (sample ID's in Table 2) was thoroughly mixed and vacuum filtered through 0.2µm pore-sized polycarbonate filters, and a replicate sample filtered with the remaining 500 ml. All cells were washed out and concentrated into 5mL of sterile DI water by placing the polycarbonate filter into a 50mL centrifuge tube containing 5mL of sterile DI water and vortexing. The suspension of cells was then harvested by ultra-centrifugation at 15,000 rpm in 1.5 ml aliquots. The pelleted cells were pooled into one tube before lysis and RNA extraction using the GeneJET RNA purification kit following the manufactures instructions (Thermo Scientific™, Waltham, MN). Metagenomic RNA concentration, purity and quality were then measured with NanoDrop 2000c. Reverse transcription of the extracted RNA and sequencing was outsourced to COSMOSID® (1600 E Gude Dr. Rockville, MD 20850 United States.) The metagenomic RNA was reverse transcribed using a commercially available kit to produce metagenomic cDNA. After standardized library preparations, sequencing was also performed by COSMOSID® (1600 E Gude Dr. Rockville, MD 20850 United States.) cDNA libraries were constructed using COSMOSID® proprietary Library Prep Kit. The PCR products were purified using 1.0X speed beads and eluted in 15µL of nuclease-free water, then quantified by PicoGreen fluorometric assay. Libraries were then pooled and loaded onto a high sensitivity chip run on a Caliper LabChipGX (Perkin Elmer, Waltham, MA) to estimate the size. Libraries

were sequenced using Illumina NextSeq/HiSeq platform. Sequencing primers targeted a full repertoire of bacteria, fungi, viruses, and protozoan parasites; and also included virulence factors and antibiotic resistance markers. All sequences obtained were quality filtered and subsequently aligned and classified using the GenBook® database, a highly curated database comprising over 150,000 microorganisms' genomes and gene sequences. This database consists of both private and publicly available genomes from sources such as NCBI-RefSeq/WGS/SRA/nr, PATRIC, M5NR, IMG, ENA, DDBJ, etc.

3.3.3 qPCR for Fecal Biomarkers

A total of three qPCR primers were used, including two that amplified human associated Bacteriodetes species (BacH and BacHum) and one for Bifidobacterium adolescensis (W257) commonly found in human fecal samples. Total metagenomic DNA was extracted from 500ml of water in each sample. All samples were subjected to each of three primers in duplicate. For positive control quality assurance, a gene fragment representing the targeted DNA sequence was obtained from the GeneBank used for each primer pair and also used to create a standard used to estimate the prevalence of the bacterial genes in the environmental water sample. The qPCR cycling conditions were used based on the manufacturer recommendations and qPCR kit instructions.

3.4 Results

Results of samples are designated as 349-01 (BCS02), 349-02 (BCS01), 398-01 (BCS06), 398-02 (BSC04), 720A (BCS03), 720B (BCS03 DUP) with 720B (BCS03) being the duplicate of 720A (BCS03) pertaining to drainage canals, and 429-01 (KMN), 719-01 (2NATDNA J4), 767-01 (CBI-MID), 854-01 (ESL-OUT) pertaining to pristine area.

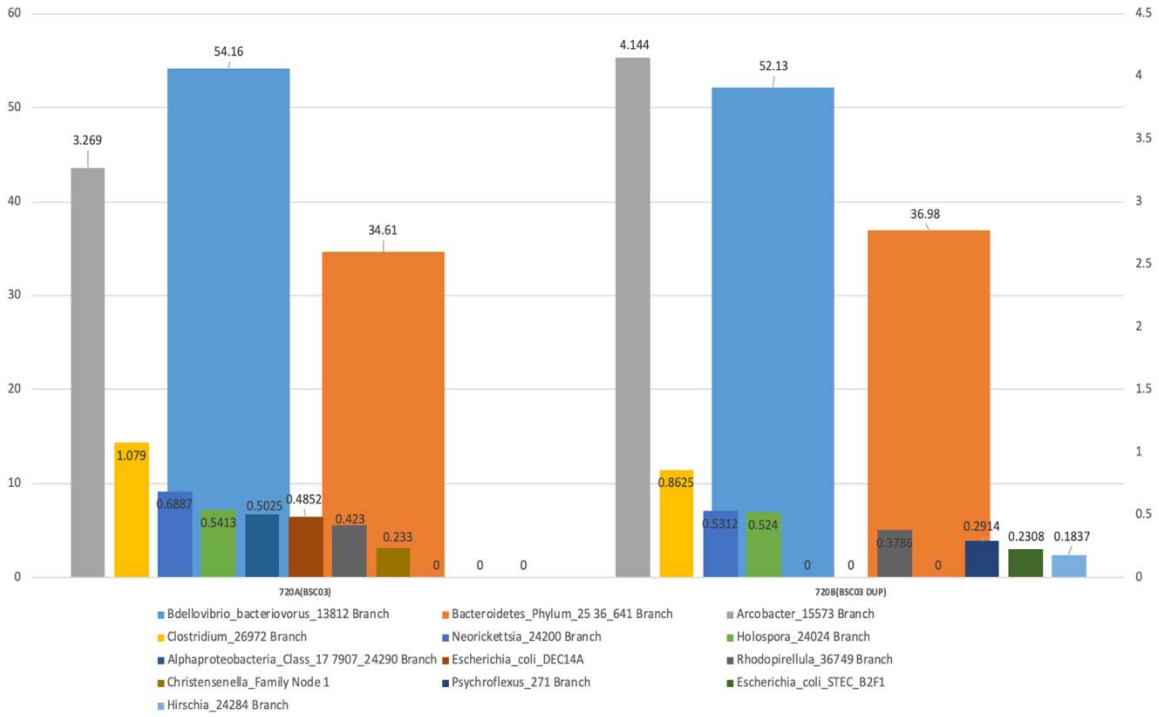

Figure 6. shows the top 95% relative abundance of bacteria of samples 720A(BCS03) and 720B(BCS03 DUP), the latter being the duplicate of the first sample.

Two bacteria clearly dominate these samples Bdellovibrio bacteriovorus 13812 Branch and Bacteroidetes Phylum 2536 641 Branch. Also, the relative abundance for each organism is relatively the same when considering the amount of sample filtered (500ml). Important to note that they both contained Escherichia coli however they were different species DEC14A in 720A(BCS03) 0.4852% and STEC B2F1 in 720B (BCS03) 0.2308%. Six organisms differed between the two samples, however all were under 0.5% of the total bacteria relative abundance.

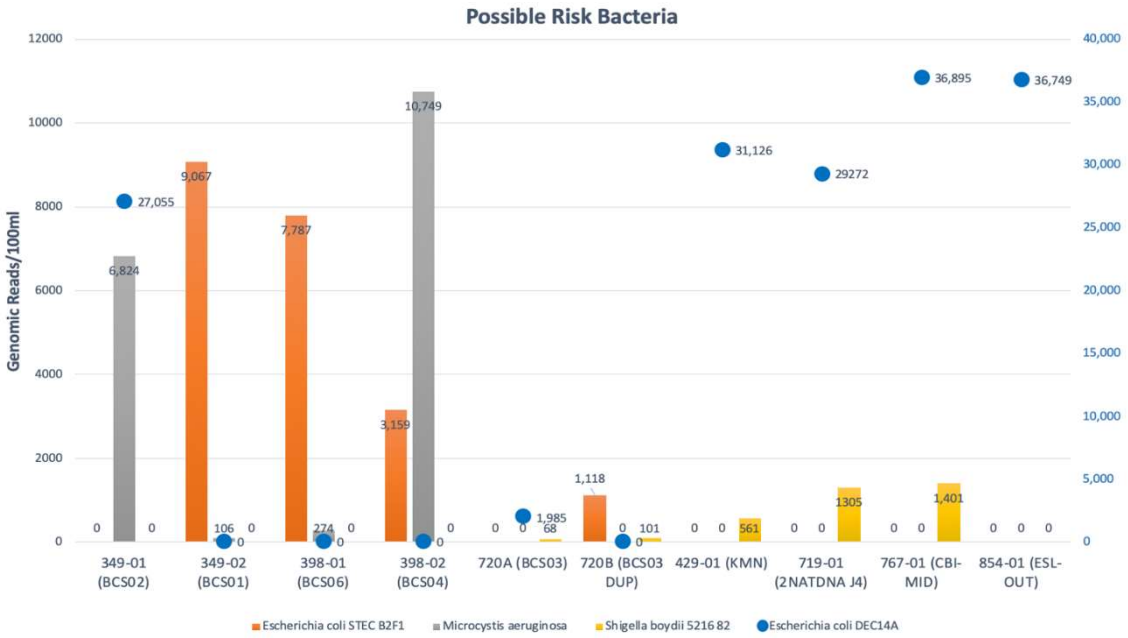

Figure 7 displays the results of possible environmental and public health risk bacteria found in each of the 10 samples recorded as genomic reads/100ml of sample.

Escherichia coli DEC14A quantities detected were too numerous to group with the other organisms, therefore this organism was placed on a separate y-axis. Surprisingly higher levels of Escherichia coli (E. coli) were detected in the pristine area. Also, noteworthy, the Cyanobacteria Microcystis aeruginosa was only detected in the drainage canal waters and not in the pristine area.

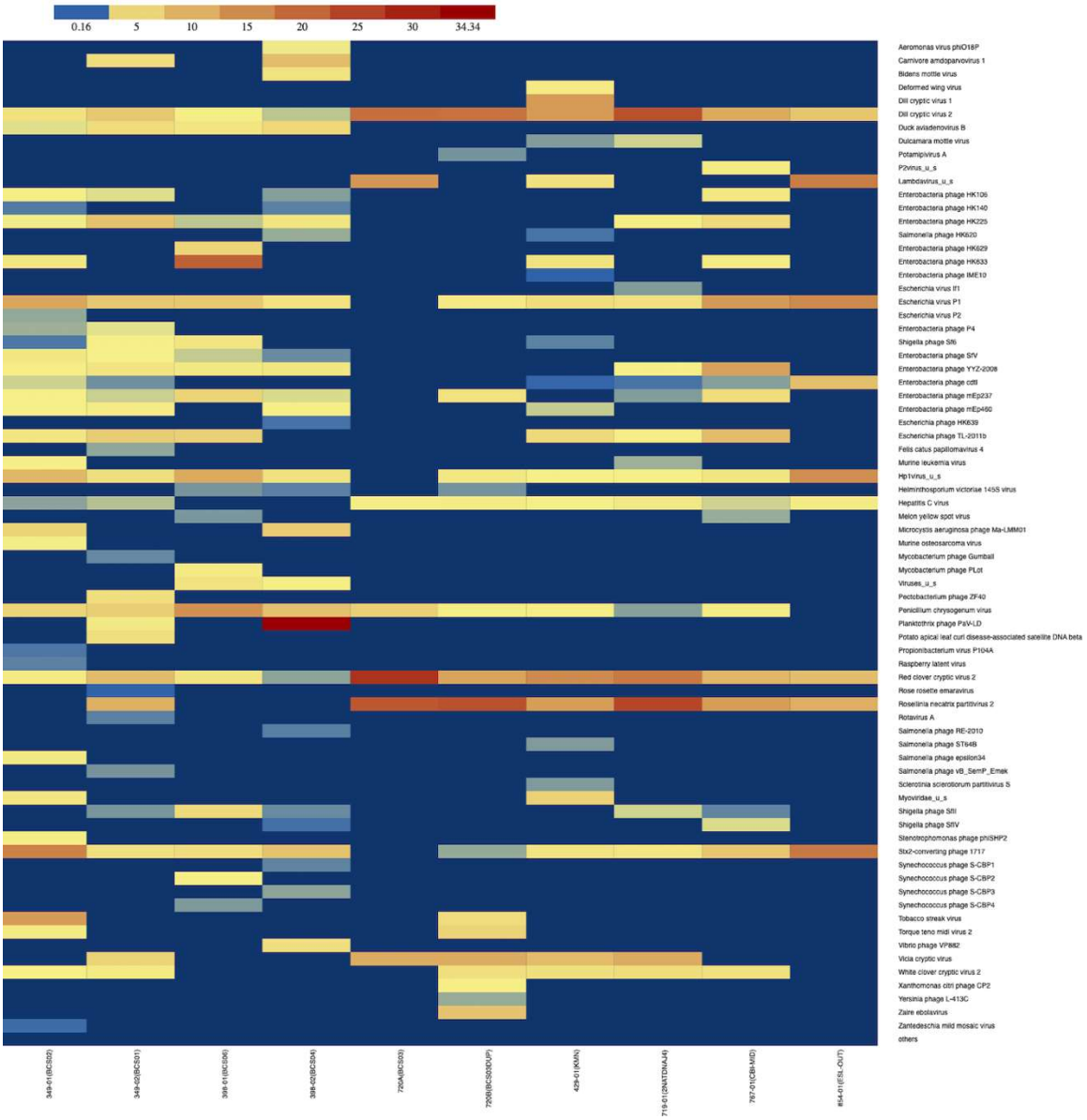

Figure 8 shows the relative abundance of viruses found in each sample.

The more maroon each bar the higher relative abundance that organism is present. The darker blue indicated not detected for that particular virus. These results are based on the total 500ml of sample that was filtered and extracted. Analysis of the occurrence of viruses in all samples clearly shows a high diversity of viruses present in canal samples versus very low diversity present in natural samples.

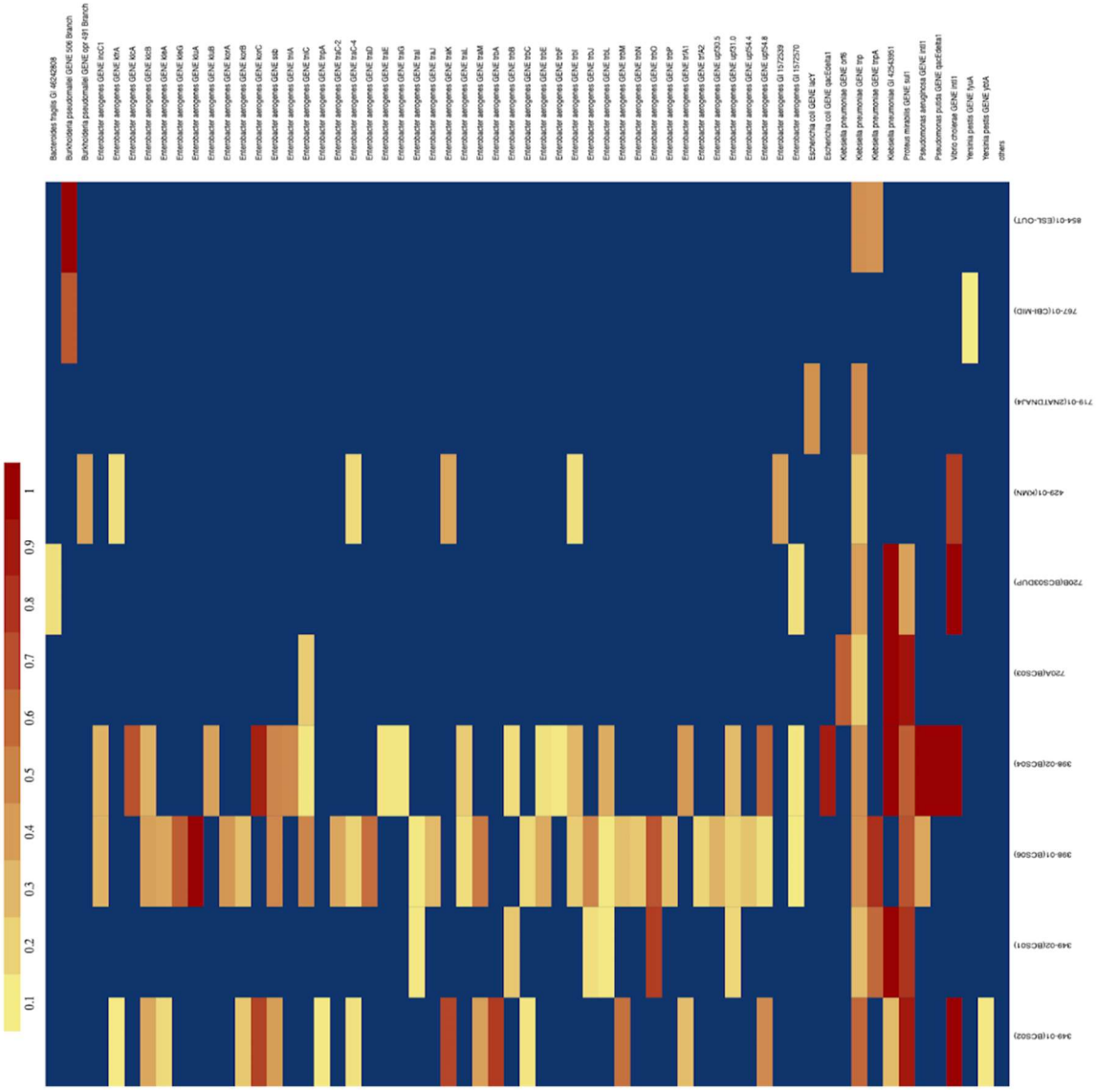

Figure 9 shows the distribution and total matches of virulence factors detected in each environmental sample.

Total matches are calculated by the total amount of shared + unique kmers found in each sample out of the total possible number of shared + unique kmers found in the database that has been calculated for that particular organism. The closer the result is to 1 means that gene is fully represented in that sample. It is clear to see the greater abundance of

virulence factors detected in the drainage canal samples versus the amount detected in the pristine area.

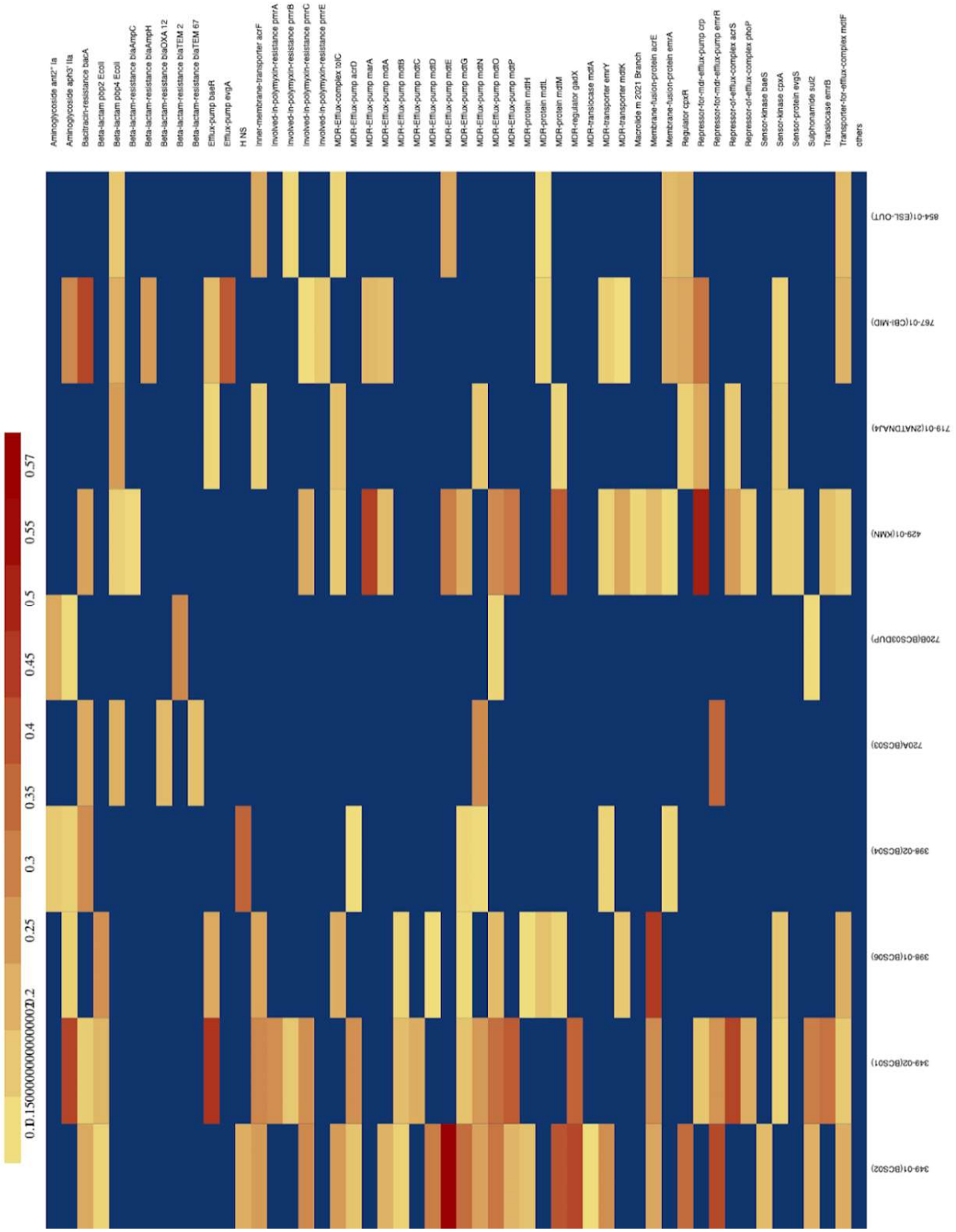

Figure 10 shows the diversity and total matches of antibiotic resistance genes found in all samples.

Figure 10 shows the diversity and total matches of antibiotic resistance genes found in all samples.

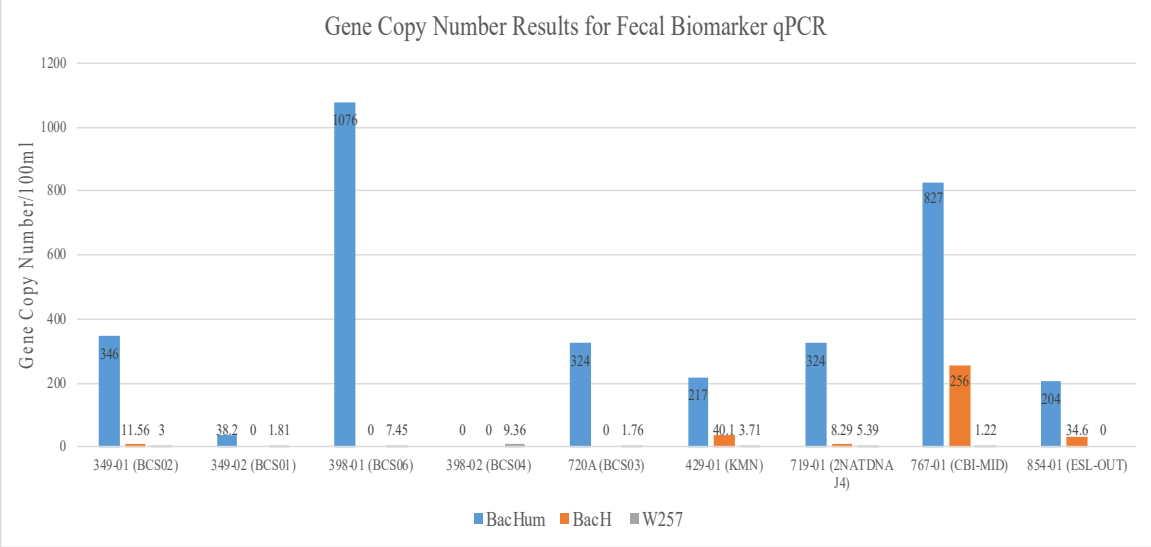

Figure 11 qPCR data representing 3 fecal biomarkers (BacHum, BacH, W257).

3.5 Discussion

Using RNA metagenomic shotgun sequencing (MSS) as a tool to measure the water quality gave us the advantage of seeing viable organisms in each sample. Previous studies that focused on quantitative polymerase chain reaction (qPCR), 16S sequencing, next-gen sequencing, and metagenomic shotgun sequencing all used extracted DNA from drinking water or environmental waters as mentioned in the introduction. One potential problem with using extracted DNA is organisms detected are not confirmed viable and could possibly be dead thus could be of little importance depending on the outcome desired. When analyzing metagenomic data and data mining for specific organisms it is most warranted to know all the organisms present were alive at the time of collection. In doing so, we can determine if possible pathogenic or environmentally invasive organisms can have an effect on public or environmental health respectively.

In this study we were determined to establish a baseline for the novel RNA MSS technique on these 9 water samples. To obtain that goal, we would have to ensure this method is reproducible, thus we analyzed a duplicate sample randomly (Figure 6). In order to assure the most accurate results when analyzing the duplicate, one 1L amber bottle was used, shaken vigorously, and filtered at the same time in different filter funnels. After filtration, extraction and sequencing were carried out as two separate samples. Figure 6 shows the top 95% relative abundance of bacteria for the sample 720A(BCS03) and its duplicate 720B(BCS03 DUP), a drainage canal sample. Clearly, we see these samples are dominated by two organisms at over 88% Bdellovibrio bacteriovorus 13812 Branch and Bacteroidetes Phylum 2536 641. Bdellovibrio bacteriovorus are ubiquitous in nature and is expected to be seen in this environment. Bacteroidetes Phylum 2536 641 was detected at the phylum level meaning the organism detected was not in the database, and this was the closest related relative. All the organisms detected in Figure 6 are expected to be seen in a drainage canal, and this sample and its duplicate are extremely close in the overall composition of bacteria. Therefore, we can deem this method sound and the results trustworthy.

Next we wanted to see the differences in microbial community if any between drainage canals and a preserved pristine area with no anthropogenic impact. In Figure 7 we can see a clear difference between the drainage canal samples and the samples collected from the pristine area. It is interesting to see the high levels of Escherichia coli DEC14a in the pristine area. Previously this organism was sequenced and uploaded to NCBI with its origin being from humans. The pristine area has had no human interactions, and the sample crew were the first people in this area. Leading us to believe that this strain of E.

coli is probably from some animals living in the area. Another organism found and previously sequenced from humans was Shigella boydii 5216 82 which was found in much higher concentrations in the pristine samples, pointing towards this organism being found in other animals as well. Microcystis aeruginosa is a cyanobacteria that blooms like a Fungi and could pose risk to the environment. It should be noted this organism is only found in the canal water samples 349-01(BCS02) 6,824genomic reads/100ml, 349-02(BCS01) 106genomic reads/100ml, 398-01(BCS06) 274genomic reads/100ml, and 398-02(BCS04) 10,749 genomic reads/100ml, and not found in the pristine area samples. The implications of introducing this organism to a pristine site where it is not commonly found deserves a close monitoring. It may die off or bloom depending on multiple interacting environmental conditions.

Analysis of the occurrence of viruses (Figure 8) in all samples clearly shows a high diversity of viruses present in canal samples versus very low diversity present in natural samples. While the vast majority of these viruses are bacteriophages; the high density of plant viruses like the Dill cryptic virus and the Red clover viruses are noteworthy. In addition, Planktothrix phage which infects a wide range of cyanobacteria (important primary colonizers and base of natural systems) is highly abundant in canal sample 398-02. This may be a reflection of the elevated counts of one of its hosts – Microcystis. However, its impact on naturally occurring cyanobacteria and subsequent environmental implications are unknown given the data present. Analysis of virulence factors (Figure 4) between all sample points clearly shows greater abundance of virulence factors in canal samples compared to the natural samples. Remarkably, some of the most abundant genes (KleG and similar ones) are on mobile elements and plasmids in canal samples and could

easily be shared with compatible environmental bacteria. On the other hand, highly elevated values of a Burkholderia pseudomallei GENE 506, associated with acute virulence and intracellular survival of Burkholderia pathogens was exclusively found in the natural environment and is significant.

At first glance of the results of antibiotic resistance genes (Figure 9) it appears there is relatively the same amount of antibiotic resistance genes found in both the canal and natural areas. However, after further data mining most of these genes are efflux pumps. Being these samples are from extracted RNA, these genes are turned on in the environment. Meaning some bacteria have active efflux pumps in the natural environment. This is expected in samples as diverse as bodies of water. Bacteria will have these pumps to pump out unwanted materials. Also, another antibiotic resistance gene of note is the Bacitracin-resistance bacA gene. This was also found in both the canal and natural samples. Bacitracin is produced by Bacillus subtilis common bacteria found in this type of environment.

Finally, we wanted to gauge the public health and environmental impact of possibly diverting the drainage canal waters into the pristine protected area. Looking at the profile of possible environmental and public health risk viable bacteria in across all samples (Figure 7), it should be noted that a key Cyanobacteria Microcystis aeruginosa was detected in relatively high numbers (349-01(BCS02) 6,824genomic reads/100ml, 349-02(BCS01) 106genomic reads/100ml, 398-01(BCS06) 274genomic reads/100ml, and 398-02(BCS04) 10,749 genomic reads/100ml) in canal samples. This organism is well known for blooming in the south Florida climate, and causing detrimental effects to the

Florida water ways. The important point is that this organism was not detected in any of the natural samples. Being that the organism was detected using the powerful RNA sequencing tool, there is no doubt it is viable. It is not known what could happen if this organism is introduced into a body of water where it is not commonly found, but surely it raises concerns and warrants close monitoring. Also, virulence factors detected in the canal samples were far greater than that detected in the natural samples (Figure 8). Not knowing the environmental implications, this is still not desired for any body of water. Using this cutting-edge RNA whole genome sequencing technology, we were able to pinpoint all viable organisms with strong confidence. Showing that there are a few concerns for environmental factors that could play a role in changing the microbiome of the natural body of water. Most importantly, we have established a baseline for these natural waters (also proved reproducibility using a duplicate sample) and can now employ this method for continuous monitoring of this environment. One downfall of this method is this technology is most effective if used overtime; however, the cost of this novel technique is still relatively expensive in water quality microbiology. More data points have to be collected to make definite confirmations of environmental or public health risk and establish a reference for each sample point. Once established, we could gauge the differences and similarities each time to determine if the results seen in this experiment of drainage canals versus pristine area would have significant results of discharging into that area.

4.0 REFERENCES

Bauza, V., Madadi, V., Ocharo, R. M., Nguyen, T. H., & Guest, J. S. 2019. Microbial Source Tracking Using 16S rRNA Amplicon Sequencing Identifies Evidence of Widespread Contamination from Young Children's Feces in an Urban Slum of Nairobi, Kenya. *Environmental Science & Technology, 53*(14), 8271-8281.

Bonilla, T. D., Nowosielski, K., Cuvelier, M., Hartz, A., Green, M., Eiobu, N., McCorquodale, D. S., Fleisher, J. M., & Rogerson, A. (2007). Prevalence and Distribution of Fecal Indicator Organisms in South Florida Beach Sand and Preliminary Assessment of Health Effects Associated with Beach Sand Exposure. Marine Pollution Bulletin, 54, 1472-1482.

Breton-Deval, L., Sanchez-Flores, A., Juárez, K., & Vera-Estrella, R. (2019). Integrative study of microbial community dynamics and water quality along The Apatlaco River. *Environmental Pollution, 255*, 113158.

Cabral, B. C. A., Hoffmann, L., Budowle, B., Ürményi, T. P., Moura-Neto, R. S., Azevedo, S. M. F. O., & Silva, R. 2018. Planktonic microbial profiling in water samples from a Brazilian Amazonian reservoir. *MicrobiologyOpen*, *7*(2).

Clesceri, L. S., Greenberg, A. E., & Eaton, A. D. (1998). Standard Methods for the Examination of Water and Wastewater (20th ed.). Washington DC: APHA American Public Health Association.

Coakley, Tricia, Brion, Gall M., & Fryar Alan E.. 2015. Prevalence of and Relationship between Two Human-Associated DNA Biomarkers for Bacteroidales in an Urban Watershed. *Journal of Environment Quality*, vol. 44, no. 5, p. 1694.

Cocolin, L., Mataragas, M., Bourdichon, F., Doulgeraki, A., Pilet, M.-F., Jagadeesan, B., &Phister, T. (2018) Next Generation Microbiological Risk Assessment Meta-Omics: The Next Need for Integration. International Journal of Food Microbiology, 287, 10 17.

Douterelo, I., Calero-Preciado, C., Soria-Carrasco, V., & Boxall, J. B. (2018). Whole Metagenome Sequencing of Chlorinated Drinking Water Distribution Systems. Environmental Science: Water Research & Technology, 4, 2080-2091.

Environmental Protection Agency (2006). Method 1600: Enterococci in Water by Membrane Filtration Using Membrane-Enterococcus Indoxyl-ß-D-Glucoside Agar (mEI). EPA-821-R-06-009

Environmental Protection Agency (2012). EPA 5.11 Fecal Bacteria

Environmental Protection Agency (2014). Overview of Technical Support Materials: A Guide to the Site-Specific Alternative Recreational Criteria TSM Documents. EPA 820-R-14-010.

Environmental Protection Agency (2014). Site-Specific Alternative Recreational Criteria Technical Support Materials for Alternative Indicators and Methods. EPA-820-R-14-011.

Esiobu, Nwadiuto, Mohammed R., Echeverry A., Green M., Bonilla T., Hartz A., McCorquodale D., Rogerson A. 2004. The application of peptide nucleic acid probes for rapid detection and enumeration of eubacteria, Staphylococcus aureus and Pseudomonas aeruginosa in recreational beaches of S. Florida. *Journal of Microbiological Methods*, vol. 57, no. 2, 2004, pp. 157–162.

Esiobu, N., Green, M., Echeverry, A., Bonilla, T. D., Stinson, C. M., Hartz, A., & Mccorquodale, D. S. (2013). High Numbers of Staphylococcus aureus at Three Bathing Beaches in South Florida. International Journal of Environmental Health Research, 23, 46-57.

Evangelista, K. V., & Coburn, J. (2010). Leptospira as an Emerging Pathogen: A Review of Its Biology, Pathogenesis and Host Immune Responses. Future Microbiology, 5, 1413-1425.

Ferguson, D. M., Griffith, J. F., McGee, C. D., Weisberg, S. B., & Hagedorn, C. (2013). Comparison of Enterococcus Species Diversity in Marine Water and Wastewater Using Enterolert and EPA Method 1600. Journal of Environmental and Public Health, 2013, Article ID: 848049.

Figueras, M. J., & Borrego, J. J. (2010). New Perspectives in Monitoring Drinking Water Microbial Quality. International Journal of Environmental Research and Public Health, 7, 4179-4202.

Gomez-Alvarez, Vicente, Revetta, Randy P., & Santo Domingo, Jorge W. 2012. Metagenomic analysis of drinking water receiving different disinfection treatments. *Applied and Environmental Microbiology*, 1 Sept. 2012.

Hartz, A., Cuvelier, M., Nowosielski, K., Bonilla, T. D., Green, M., Esiobu, N., & Rogerson, A. (2008). Survival Potential of and Enterococci in Subtropical Beach Sand: Implications for Water Quality Managers. Journal of Environment Quality, 37, 898.

Kongprajug, A., Chyerochana, N., Somnark, P., Kampaengthong, P. L., Mongkolsuk, S., & Sirikanchana, K. 2019. Human and animal microbial source tracking in a tropical river with multiple land use activities. *International Journal of Hygiene and Environmental Health, 222*(4), 645-654.

Li, B., Saingam, P., Ishii, S., & Yan, T. 2018. Multiplex PCR coupled with direct amplicon sequencing for simultaneous detection of numerous waterborne pathogens. *Applied Microbiology and Biotechnology*, *103*(2), 953–961.

Mohiuddin, M. M., Salama, Y., Schellhorn, H. E., & Golding, G. B. 2017. Shotgun metagenomic sequencing reveals freshwater beach sands as reservoir of bacterial pathogens. *Water Research,115*, 360-369.

Noble, R., Moore, D., Leecaster, M., Mcgee, C., & Weisberg, S. 2003. Comparison of total coliform, fecal coliform, and enterococcus bacterial indicator response for ocean recreational water quality testing. Water Research, 37(7), 1637-1643.

Odonkor, S. T., & Ampofo, J. K. 2013. Escherichia coli as an indicator of bacteriological quality of water: an overview. Microbiology Research. *4*(1), 2.

Office of Water, Environmental Protection Agency (2010). Quantitative Microbial Risk Assessment to Estimate Illness in Freshwater Impacted by Agricultural Animal Sources of Fecal Contamination. EPA 822-R-10-005.

Rashkin, S., Jun, G., Chen, S., & Abecasis, G. R. (2017). Optimal Sequencing Strategies for Identifying Disease-Associated Singletons. PLOS Genetics, 13, e1006811.

Rochelle-Newall, E. E., Nguyen, T. M., Le, T. P., Sengtaheuanghoung, O., & Riboizi, O. (2015). A Short Review of Fecal Indicator Bacteria in Tropical Aquatic Ecosystems: Knowledge Gaps and Future Directions. Frontiers in Microbiology, 6, 308.

Roy, M., Arnaud, J., Jasmin, P., Hamner, S., Hasan, N., Colwell, R., & Ford, T. 2018. A Metagenomic Approach to Evaluating Surface Water Quality in Haiti. International Journal of Environmental Research and Public Health, 15(10), 2211.

Metagenomic Approach to Evaluating Surface Water Quality in Haiti. International Journal of Environmental Research and Public Health, 15, 2211.

Tan, B. F., Charmaine, N., Jean, P. N., Lay, L. L., Karina, G. Y.-H., & Janelle, T. R. (2015). Next-Generation Sequencing (NGS) for Assessment of Microbial Water

Quality: Current Progress, Challenges, and Future Opportunities. Frontiers in Microbiology, 6, 1027.

Vadde, K. K., Mccarthy, A. J., Rong, R., & Sekar, R. 2019. Quantification of Microbial Source Tracking and Pathogenic Bacterial Markers in Water and Sediments of Tiaoxi River (Taihu Watershed). Frontiers in Microbiology, 10.

Wade, T. J., Calderon, R. L., Sams, E., Beach, M., Brenner, K. P., Williams, A. H., & Dufour, A. P. 2006. Rapidly Measured Indicators of Recreational Water Quality Are Predictive of Swimming-Associated Gastrointestinal Illness. Environmental Health Perspectives, 114(1), 24–28.

Water Quality Standards Variances. Epa.gov.