University of Massachusetts Amherst
ScholarWorks@UMass Amherst

Masters Theses | Dissertations and Theses

October 2019

A Novel Infield Metagenomic Approach to Evaluating Surface Water Quality in Lake Warner

Brooke Stebbins

Follow this and additional works at: https://scholarworks.umass.edu/masters_theses_2

 Part of the Environmental Health Commons

Recommended Citation

Stebbins, Brooke, "A Novel Infield Metagenomic Approach to Evaluating Surface Water Quality in Lake Warner" (2019). *Masters Theses*. 856.
https://scholarworks.umass.edu/masters_theses_2/856

This Open Access Thesis is brought to you for free and open access by the Dissertations and Theses at ScholarWorks@UMass Amherst. It has been accepted for inclusion in Masters Theses by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.

A NOVEL INFIELD METAGENOMIC APPROACH TO EVALUATING SURFACE WATER QUALITY IN LAKE WARNER

A Thesis Presented

by

BROOKE STEBBINS

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

MASTER OF SCIENCE

September 2019

Public Health
Environmental Health Sciences

© Copyright by Brooke Stebbins 2019
All rights reserved

A NOVEL INFIELD METAGENOMIC APPROACH TO EVALUATING SURFACE WATER QUALITY IN LAKE WARNER

A Thesis Presented

by

BROOKE STEBBINS

Approved as to style and content by:

_____

Timothy Ford, Chair

_____

Laura Vandenberg, Member

_____

Alexander Suvorov, Member

_____

Emily Kumpel, Member

_____

Laura N. Vandenberg, Graduate Program Director

_____

Timothy Ford, Department Chair,
Environmental Health Sciences

ACKNOWEDLGEMENTS

I would like to thank my advisor Dr. Timothy Ford for his constant support and understanding through my master's thesis. This research project was made possible because of his encouragement and wealth of knowledge of the field. I would also like to extend my gratitude to my committee members, Dr. Laura Vandenberg, Dr. Alexander Suvorov, and Dr. Emily Kumpel for their support and guidance. Additionally to Dr. Laura Vandenberg as the Academic Advisor, who provided insight, life lessons, and laughs. She truly is an asset to the Department of Environmental Health Sciences. I would like to extend my gratitude to Dr. Cristina Martin, who has provided me with laboratory knowledge and allowed me to discuss any issues that occurred during this process. Also, to Patrick Wittbold for his willingness to help even when that involved collecting sludge at the wastewater plant, I appreciated the valuable advice and time he was able to lend to me. My sincerest appreciation to my husband, Kody. Thank you for your constant love, patience, and support through this stressful time. Finally to my friends and family, thank you for your time and love.

ABSTRACT

A NOVEL METAGENOMIC APPROACH TO EVALUATING SURFACE WATER QUALITY IN LAKE WARNER

SEPTEMBER 2019

BROOKE STEBBINS, B.S., UNIVERSITY OF MASSACHUSETTS AMHERST

M.S., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Timothy E. Ford

In January 2010, a magnitude 7.0 earthquake devastated Haiti, one of the poorest countries in the Western Hemisphere. Haiti's weak sanitation infrastructure and limitations in the public health system made the country susceptible to the spread of waterborne diseases. Following the earthquake, cholera rapidly spread through Haiti, killing 4,672 people in 5 months and leaving thousands hospitalized (MSNBC, 2010). Cholera is an infectious diarrheal disease caused by the pathogen, *Vibrio cholerae*, which results in severe dehydration with a high mortality risk. The source of the epidemic was traced to the Artibonite River, the island's longest and most essential drinking water source (Encyclopaedia Britannica. n.d.). The origin of the contamination was later discovered to be unsanitary conditions left from United Nations peacekeepers from Nepal. Eight years later, cholera cases are still prevalent, although numbers have declined recently due to aid from private organizations (Dowell, S.F. et al 2011, Katz, J.M. 2013). However, with climate-related increases in ocean water temperatures, scientists expect hurricanes to intensify and increase damage to developing countries (Center for Climate and Energy Solutions. n.d.). Natural disasters promote the spread of waterborne illness by

isolating people from safe drinking water and destroying public health infrastructure such as happened with the cholera outbreak in Haiti (Funari, E. et al 2013). To prevent future waterborne disease epidemics in such areas with limited resources, it would be beneficial to improve environmental surveillance through development of rapid, reliable, and portable detection methods for waterborne pathogens.

The advent of high-throughput sequencing technologies has enabled the detection and characterization of microbial communities in their natural environments, an approach known as metagenomics. Metagenomic sequencing, unlike more traditional PCR methodologies, is capable of sequencing thousands of organisms in a sample. This metagenomic shotgun sequencing approach detects the abundance of microbes and bacterial diversity in the environment (Illumina, n.d.). The Oxford Nanopore MinION is a shotgun sequencing device that is optimal for portable, rapid detection of the microbial diversity in an environmental sample (Oxford MinION, n.d.). This handheld device has enormous potential for field use in emergency preparedness and disease response, particularly in developing countries where more advanced analytical equipment may be inaccessible due to lack of facilities or damaged infrastructure. Having access to quick, infield assessment technology for rapidly emerging outbreaks would be beneficial to a disease-specific public health response.

Current protocols recommend that DNA is extracted from environmental samples as rapidly as possible after collection. If cooling is available with an insulated ice chest, samples may be transported/stored for periods ranging from 6 to 24 hours. The shorter timeframes minimize unwanted shifts in microbial structure (U.S. Geological Survey, 1997, WHO, n.d.). Access to cold storage in remote areas is unlikely, and the use of

liquid preservation methods could assist in maintaining quality of DNA, and hence produce more accurate data in metagenomic analyses. In the absence of cold storage facilities, infield filtration coupled with preservation techniques are necessary to maintain samples integrity for transport to laboratory facilities.

This thesis aimed to develop an infield filtration and sequencing protocol, coupled with the Oxford Nanopore MinION sequencing platform, to identify the potential bacteria, viruses, protozoa, fungi, antimicrobial resistance (AMR), pathogenic strains, and virulence associated genes for use in remote locations. Five locations across Lake Warner, Massachusetts were used for method development, coupled with Millipore Sterivex filters for field filtration to determine the most effective method for sample preparation in remote locations. Additionally, a chemical preservation method was assessed using dimethyl sulfoxide, disodium EDTA, and saturated NaCl (DESS). A study by Gray et al, found that liquid preservation methods (DNAgard, RNAlater, and DESS) outperformed the card-based preservatives (FTA cards and FTA Elute cards) in terms of bacterial recovery (Gray, M.A., et al 2013). DESS was selected for investigation in this thesis because of the low cost compared to the other liquid-based preservatives.

Lake Warner in Hadley, Massachusetts, which is heavily used for fishing and boating activities, flows into the Connecticut River via the Mill River. Historically, the lake experienced high *Escherichia coli (E. coli)* levels due to pollution from primary effluent released in the 1950s from the Amherst Wastewater Treatment Plant (Johnson, J., 2015). Similar to *Vibrio cholerae*, *E.coli spp* is a waterborne bacteria caused by fecal contamination. Although most *E.coli* are natural inhabitants of the gastrointestinal tract, pathogenic serotypes can result in severe complications in vulnerable populations such as

kidney failure in children and the elderly adults. (Todar, K., 2012). Lake Warner was chosen for the method development because of its history of *E. coli* pollution and recreational traffic as well as its general accessibility for study.

Designing a methodology for rapid detection of pathogenic bacteria using a metagenomic approach could help improve surveillance for environmental pathogens that pose future epidemic risk. These tools are becoming increasingly important for prediction and response to waterborne diseases as climate impacts increase the frequency, intensity, and duration of extreme weather events that damage critical infrastructure for vulnerable populations (van Aalst, M.K. 2006).

TABLE OF CONTENTS

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER 1

LITERATURE REVIEW

1.1 Background

1.2 Waterborne Diseases

1.2.1 *Vibrio Cholerae*

Cholera is a diarrheal disease transmitted by the fecal-oral route that is often asymptomatic. However, 10% of the infected population will develop a severe case. The symptoms of severe cholera include diarrhea, vomiting, and cramps, which will lead to dehydration and hypovolemic shock due to the sudden loss of bodily fluids, and potential death if left untreated. Cholera stems from the toxigenic bacterium *Vibrio cholera* of either serogroup O1 or O139. Vibrios are Gram-negative (bacteria that will not maintain a crystal violet stain during Gram-staining used for bacterial differentiation) that are often resistant to antibiotics, lysozymes produced by immune cells, and detergents (Pierce, N.F., 1972) and when digested most vibrios will die in the stomach due to their sensitivity to acid (CDC, 2018). The pathogenic serotypes release cholera enterotoxin into the intestine which binds to the enterocytes (intestinal cells) and moves into the cytosol to activate reactions to produce cAMP (secondary messenger) by adenylate cyclase stimulation (Figure 1.1). Cystic fibrosis transmembrane conductance regulator (CFTR) is activated by the increased levels of cAMP resulting in diarrhea by the efflux of water and ions from infected enterocytes (Thiagarajah, J.R., et al 2005). These pathogenic serotypes are identified using three criteria: the O group 1 specific antiserum will be missing agglutination (grouping of particles in presence of an antibody or complement),

O group 139 specific antiserum agglutination, and capsule presence (Finkelstein, RA. 1996).

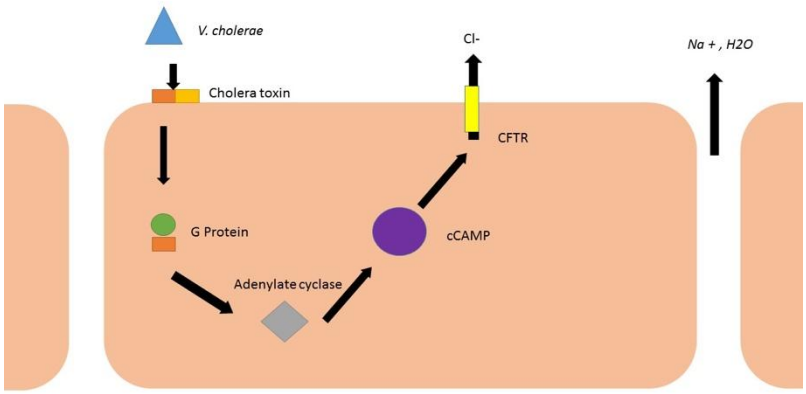

Figure 1.1. The process of *V. Cholerae* in human cells.

Dr. Timothy Ford's research team from UMass Amherst is optimizing the rapid approaches to monitor the water quality in Haiti after the deadly outbreak of cholera that followed the 7.0 Richter scale earthquake that caused a total of 817,000 cases of illnesses and 9,749 deaths from 2010-2016 (Katz, J.M. 2013, Roberts, M. 2011). Organizations such as the U.S. Centers for Disease Control and Prevention (CDC), The United Nations Children's Fund (UNICEF), the World Health Organization (WHO), as well as other health associations were able to raise money to provide stable response systems and surveillance for diseases; however, maintaining these monitoring programs are too costly for many developing countries, preventing these areas from completely eliminating the threat of waterborne diseases such as cholera (MSNBC. 2010). Last year, 159 deaths and 13,681 cases occurred globally; though the incidences of waterborne illnesses and death have declined, these numbers are expected to grow due to the predicted increase in

frequency of heavy rainfall and hurricanes (Dowell, S.F. et al 2011). Haiti and other developing countries will likely be unable to handle the increasing flooding events due to the poor water and sanitation infrastructure causing the population to be exposed to contaminated drinking water, further increasing the chances of waterborne illnesses.

1.2.1.1 *Escherichia coli*

Escherichia coli is a gram-negative coliform bacterium that includes potentially harmful strains, although most serotypes are harmless. It is found in the normal gut microbiota and inhabits the lower intestine of endotherms where it provides benefits to the host such as vitamin K_2 production and inhibition of the growth of pathogenic bacteria (Singleton, P. 1999, Vogt, R.L. 2005). *E. coli* is released into the environment through defecation and can thrive in an aerobic environment for approximately. three days, though some researchers have suggested the bacteria can last longer outside of a host (Vogt, R.L. 2005). Routes of exposure to virulent strains of *E. coli* through ingestion can lead to a variety of health concerns including urinary tract infections, neonatal meningitis, gastroenteritis, hemorrhagic colitis, Crohn's disease, and on rare instances, bowel necrosis (Eckburg, P.B. et al 2005, Lim, J.Y. et al 2005). The pathogenic strains of particular concern due to high mortality risk are *E. coli* O157:H7 and O104:H4, both of which release shiga toxins that halt protein synthesis by cleaving an adenine nucleobase from the 28S RNA (60S subunit), a similar mechanism to ricin toxicity (Reid, G. et al 2001). Inhibition of protein synthesis results in cell apoptosis in the kidney and other tissues; Enterohemorrhagic *Escherichia coli* (EHEC) also produces intimin, an attaching

and effacing protein, and this virulence factors contribute to severe diarrhea (Hartland, E.L. et al 2013).

Recreational water use guidelines are set by the U.S. Environmental Protection Agency for swimming and have been established for *E. coli* at >126 cfu/ 100mL. The geometric mean of five samples must be fewer than >126 cfu/ 100mL over a 30 day period to be considered safe for swimming. Presence of high levels of *E. coli* indicates fecal pollution which could be from numerous sources such as agricultural runoff, sewage, and/or septic leakage (Johnson, J. 2015). Researchers from Montana State University monitored a heavily recreationally used river, the Little Bighorn River, on the Crow Indian Reservation for *E. Coli*, more specifically *E. coli* O157:H7. The study found that a site downstream of the Crow Fair swimming hole site tested positive for Shiga toxin 1 gene and intimin and concluded that concentrated animal feeding operation (CAFO) manure may be the cause of the high levels of fecal pollution. Presence of these virulence factors are a potential public health threat that require monitoring and further source tracking (Hamner, S. et al 2015).

1.3 Culturing

Culturing methods are considered the standard for microbial identification and possible diagnosis. The samples of interest are inoculated on various media that will aid in the identification and presence of specific organisms. Although this method is largely used for monitoring the public's safety, culturing is limited in its capacity to identify and quantify the microorganisms in the environment due to its poor sensitivity. The culturing standard also requires time for incubation, which can be especially dangerous when a

pathogenic bacteria is present in a public swimming hole and rapid detection of waterborne pathogens is required. A more sensitive approach is needed to identify these pathogenic microbes for the public's health (Zhou, Y. et al 2016). In particular cases, waterborne infections can have severe adverse health effects if left untreated, especially for vulnerable populations such as children and pregnant women. The non-culturing method of metagenomics was established to eliminate the limitations of culturing and allow researchers the ability to analyze the microbial diversity present in the environment.

1.4 Sequencing

Advances in sequence-based testing have provided researchers with a more efficient means of microbial identification. Phenotypic testing remains the method of choice for bacteria and fungi, while the sequence-based testing is increasingly used to identify the presence of antiviral resistance (Pandya, S, et al 2017).

1.4.1 16S ribosomal RNA amplicon

In the early 2000s, scientists started to identify the community of microbes in the human gut (e.g. the gut microbiome) by using whole genome sequencing. This research project is called the Human Microbiome Project (HMP), with the main objective of comparing the characteristics of the human microbiome between diseased and healthy individuals. Understanding the diversity of the human gut microbiota provides insights on the microbial ecology that may help to understand the role of gut bacteria in promoting different disease states (D'Agata, E., et al. 2015, Sandle, T 2016). Though a combination of methods has been used to identify these microorganisms, most of the work focuses on

using the 16S ribosomal RNA amplicon, as culturing techniques fail to identify >90% of microbial species (Ranjan, R. et al 2016). This sequencing method uses amplification of the 16S rRNA region by PCR. The 16S ribosomal RNA gene is widely used for prokaryote analysis because the gene codes for the bacterial ribosome, 30S subunit, found in all bacterial species. Research and clinical laboratories utilize this sequencing method to characterize and identify pathogenic strains present in a sample because it can discriminate between similar bacterial species (Wang, X. et al 2015, Kothari, R. et al 2018). However, there are limitations to the method, 16S rRNA gene annotation was established by a presumed connection with the operational taxonomic unit (OTU). Particular genes are identified based on predictions using the OTUs rather than directly sequencing; therefore, using this approach generates analyzes that are less specific on the species level and more accurate on the genus or phyla level. Factors such as horizontal gene transfer can also prevent precise gene identification from 16S rRNA amplicon sequencing, making it an unreliable methodology (Ranjan, R. et al 2016).

1.4.1.1 Whole Genome Shotgun Sequencing

An alternative method using taxa classification databases different from 16S rRNA amplicons for genomic assessment is whole genome shotgun sequencing (WGS). This technique shears large fragments of DNA into smaller fragments that are cloned and randomly sequenced to produce a genomic library. The downside of this method is that it requires a substantial amount of data analysis and compared to the 16S rRNA amplicon technique the overall expense is greater (Ranjan, R. et al 2016, Clark, D.P. et al 2013). In the Ranjan et al. 2016 study, researchers compared the two sequencing techniques using stool specimens on both MiSeq and HiSeq 2000 instruments for WGS and 16S amplicon

to determine the superior method for microbiome analysis. The overall findings suggested that the WGS approach had considerable advantages even though the 16S rRNA amplicon approach is more cost efficient and has an established data archive. WGS produced double the amount of identified bacterial species per read with 4,100 versus 2,050 species, and also provided data on the viruses, fungi, and protozoa, unlike the 16S rRNA amplicon method. Due to the differences in taxa classification process, the WGS approach was able to detect 2,441 species that the 16S rRNA amplicon method did not detect, as well as identifying more diversity (Ranjan, R. et al 2016). Utilizing the WGS approach for microbial community analysis provided a wealth of knowledge about microbial populations from both clinical and environmental samples (Chen, K. et al 2005).

1.5 MiSeq Illumina Sequencing

The Illumina short-read, high through-put sequencing method is an example of the second (or next) generation sequencing technologies which are currently the most commonly used platforms. The HiSeq2000 has the ability to produce 600 Gb of paired-end 100 base pair reads in a timeframe of ten days and has become more affordable as more laboratories adopt the technology. The MiSeq produces 5 million 150-base paired-end reads with 1.5 Gb sequenced within a day. The high-throughput yield capability allows researchers to identify the vast microbial diversity in an environmental or clinical sample (Illumina, 2018). Both HiSeq 2000 and MiSeq generate similar results except for the scale of microbial diversity; determining the sequencer would depend on the budget and the experimental question (Cao, MD. et al 2016, Caporaso, J.G. et al 2012).

1.6 Oxford Nanopore MinION

The Oxford nanopore-based, MinION, is an example of third generation sequencing technologies. This portable handheld device can read long DNA lengths making it ideal for sequencing environmental samples. Nanopore sequencing uses an ionic current to separate macromolecules by size and configuration. Nanopores (with a diameter of 10^{-9} meters) are formed in an electrically resistant membrane by pore-forming proteins such as α-hemolysin, a pore-forming toxin secreted by Staphylococcus aureus that binds to the outer membrane of cells, causing apoptosis (Oxford Nanopore Technologies, n.d., Liu, Z., et al 2016, Stoddart, D. et al 2009). When DNA, RNA, or protein pass through these pores, or even come in contact with them, they disrupt the ionic current that allows for identification of the molecule. These α-hemolysin pores can identify the four bases using three recognition sites, R1, R2, and R3, that can decipher between the bases located in the lower section of the two 5 nm sections (Figure 1.2) (Oxford Nanopore Technologies, n.d.).

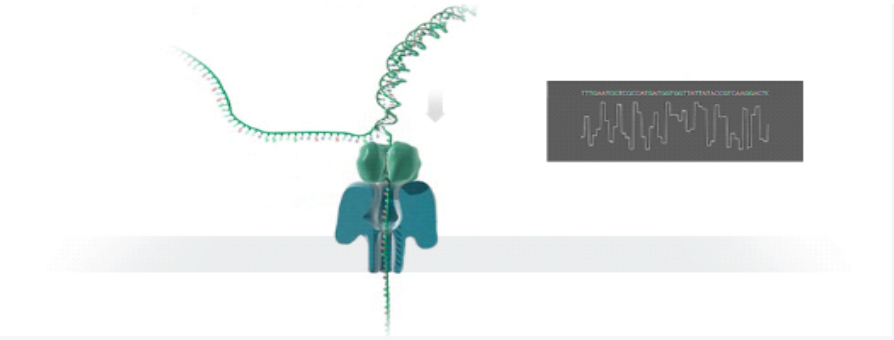

Figure 1.2. Sequencing of DNA by a nanopore. Taken from the Oxford Nanopore Technologies website.

The Oxford Nanopore MinION has the potential to be a valuable monitoring tool for detecting waterborne pathogens, especially in remote locations given its' portability, but the method still requires optimization. In contrast to Illumina, the Oxford Nanopore MinION generates lower through-put, and the cost for reagents and one flow cell can be upwards of $1,000; however, there is the option of multiplexing the samples, where up to 12 samples are barcoded and combined to run through one flow cell. The potential for multiplexing establishes the Oxford Nanopore MinION as a more cost-efficient method at $80 per sample as opposed to $1,000; however, multiplexing decreases the reads generated. A major advantage of the Oxford Nanopore MinION would be initially identifying the pathogen for an outbreak followed by the high through-put Illumina sequencing (Roy, M. et al 2018). Establishing the Oxford Nanopore MinION as an emergency preparedness method would benefit remote areas in developing countries in remote areas where reliable access to equipment such as the Illumina is unlikely. The benefits of sequencing environmental samples for metagenomic analysis using the whole genome sequencing Oxford Nanopore MinION would enable rapid field results. The

device can rapidly sequence enough to identify species and strains with fewer than 500 reads, which is beneficial for providers who require drug resistance profiling within hours (Cao, MD. et al 2016). Additionally, it is capable of producing 500 reads in 10 minutes (Oxford Nanopore Technologies., n.d.); however, a recent study comparing Oxford Nanopore MinION and Illumina found that the Oxford Nanopore MinION had a significantly higher error rate (Lu, H. et al 2016).

1.7 Metagenomic research

Sequencing of metagenomic samples can be utilized in a variety of applications such as researching antibiotic resistance genes (ARG) through wastewater. Wastewater is an important source of antibiotic-resistant bacteria (ARB) and ARG in the environment and has been referred to as a "hotspot" for horizontal gene transfer (HGT), the process where genetic information is transferred from one species to another (Karkman, A., et al 2018). Over a four-year timeframe, Yang et al. collected activated sludge from the Hong Kong Wastewater Treatment Plant (Shatin) and found aminoglycoside and tetracycline resistance genes to be the most abundant; however, the ARGs were seasonal with spikes of sulfonamide-, multidrug-, and chloramphenicol-resistance genes observed more often in the winter compared to summer months. This approach using environmental samples for metagenomic analysis proved to be an effective tool for identifying and surveying ARGs (Yang, Y. et al 2013). Other environmental samples were assessed for ARGs by Chen and colleagues (2013) when they investigated river and sea sediments for antibiotics such as ofloxacin, sulfamethazine, tetracycline, norfloxacin, and erythromycin. The overall finding was that the sea sediment contained a broader diversity

of ARG compared to the river sediment which consisted of more ARGs connected to human activity (Chen, B. et al 2013).

Additionally, a metagenomic approach was utilized as a surveillance tool for potential waterborne pathogens in Haiti. Roy et al. (2018) collected surface water samples at five sites in January 2018 near the origin of the 2010 cholera outbreak; this outbreak continues to plague the country today. CosmosID bioinformatic platform results showed a broad diversity of bacteria present in the water samples (see Figure 1.3, a Krona visualization showing the gram-negative *Proteobacteria* as the dominant phylum comprising 84% of the bacterial diversity followed by *Gammaproteobacteria* at 44%, *Alphaproteobacteria* at 41%, and *Betaproteobacteria* at 11%). Within the *Gammaproteobacteria* class, there were two bacterial genera present that are of concern to human health: *Acinetobacter* and *Legionella.* This preliminary study provides a solid foundation for further analysis into utilizing sequencing of metagenomic samples as a environmental monitoring tool. The researchers were able to detect converting phages for Shiga toxins and cholera toxins, revealing the possibility that these diseases could re-emerge. These findings also demonstrate the need for continuous monitoring in Haiti and other countries that historically have experienced outbreaks (Roy, M.A. et al 2018).

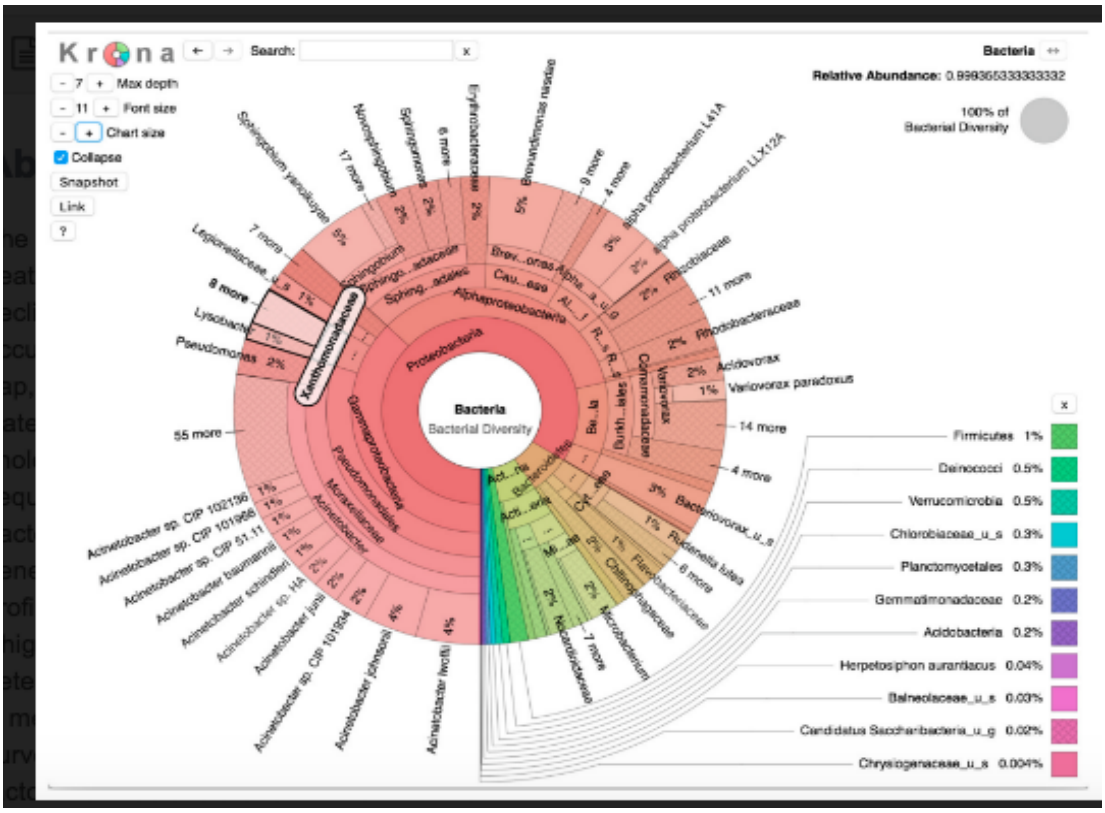

Figure 1.3. Krona visualization. Total bacterial diversity, representing gamma diversity, among all samples from the January 2018 time point (Roy, M.A. et al 2018). Permission given by the lead author.

1.8 Conclusions

No present day methodology exists for a reliable, rapid approach to detect waterborne pathogens in remote locations that encapsulates the filtration, transportation, and processing of samples. In circumstances where resources are limited, such as in developing countries, there is a lack of effective monitoring techniques that could be utilized in emergency situations. Designing and implementing a method with current technology could improve surveillance for developing countries and allow prediction and an early response to outbreaks of infectious diseases. During the cholera epidemic in

Haiti, a nationwide monitoring program was initiated that included registering cases and epidemiological surveys that tracked the origin of the disease (Piarroux, R. et al 2011). However, the monitoring program was slow as more cases of cholera erupted; furthermore, the Haitian government cannot afford to maintain this program. If affordable emergency preparedness methods for detection of waterborne pathogens and community-based outreach were implemented– these epidemic occurrences may be better controlled.

1.9 Thesis Overview

Better detection of waterborne pathogens such as *Vibrio cholera* could be of benefit to millions of people, especially in regions that currently lack the resources for monitoring and control. Increased temperatures and severe rain events anticipated due to climate change will contribute to the transmission of waterborne diseases. Although bacterial culturing methods can be utilized for water quality assessment of particular bacterial species – these methods cannot provide a comprehensive assessment of pathogenic strains, converting phages, and virulence and antimicrobial resistance genes. Given the previous comparison studies on sequencing of environmental samples using a metagenomic approach, the WGS provides a better representation of the microbial diversity in a given sample. However, it is not without its limitations in terms of cost, the need for filtration of fairly large volumes of water, sample transportation, isolation of high purity DNA, library preparation and sequencing. Each step needs optimization for use in a field setting such as Haiti, which is the focus of this thesis.

It is difficult to find a body of water with comparable conditions in the United States to the Artibonite River (Figure 1.5), the suggested source of the cholera epidemic in Haiti. In addition, many obstacles prevented water samples from being shipped to the

U.S. from Haiti, including the current political climate and risk of sample degradation. Instead, a local lake, Lake Warner, was used for the method development, Lake Warner is a 68 acre heavily used body of water located in northern Hadley, Massachusetts, in the Mill River and Connecticut River watersheds (Figure 1.4). The lake is primarily used for fishing and recreational water sports such as kayaking, boating, and canoeing. The Mill River flows into Lake Warner and begins in Puffers pond, located in Amherst Massachusetts, and winds through agricultural farming lands, highways, and the UMass campus before it empties into the lake. The Mill River inlet to Lake Warner was on the Section 303(d) list of the U.S. Federal Clean Water Act for high levels of *E.coli* bacteria pollution in 2002, 2007, 2012, and 2014. Lake Warner has had a history of high levels of *E.coli*, although more recently the levels have dropped due to work by Friends of Lake Warner and Mill River, a nonprofit organization for the preservation, restoration, and use of the lake and river. However, there are still concerns about the water quality in various parts of the lake, especially the mouth of the Mill River. Lake Warner is monitored for bacteria levels (*E.coli)* by culturing and follows the maximum 235 colonies per 100 mL rule by the Massachusetts state standards for primary and secondary contact. To our knowledge, no sequencing has been performed on any of the samples collected from this lake for monitoring purposes. Additionally, water quality monitoring is only conducted from May to September. High *E.coli* reads were found in the inlet of Mill River in Lake Warner. The source of this bacterial population is unknown, but it has been suggested that aging sewer and septic tanks may be the cause in 2002, 2007, 2012, and 2014 (Johnson, J. 2015). Lake Warner was selected as the model for the method development in this thesis due to the historically high levels of *E.coli* and wastewater pollution.

A reliable, rapid detection method for waterborne pathogens could help inform treatment decisions for drinking water sources. The Oxford Nanopore MinION device is a portable, handheld sequencer that could sequence samples within hours, compared to the Illumina and other second-generation sequencing platforms, which require upwards of 55 hours. In emergency situations, the Oxford Nanopore MinION could be utilized for clinical purposes as well as environmental monitoring in developing countries. However, the optimal preservation method is transporting and storing samples on ice for a maximum of 6 hours. Many remote areas with limited resources require more than 6 hours of transportation to the nearest facilities for sample processing. An alternative preservation method that could prevent any potential DNA degradation would be necessary in these circumstances. Dimethyl sulfoxide, disodium EDTA, and saturated NaCl, known as DESS, is a preservation method that could significantly prevent degradation of microbial DNA. DESS preservation would enable samples to be transported from remote locations without the need for cold storage during transport. This chemical method would be useful in remote areas of Haiti if options for cold storage are limited or nonexistent to transfer environmental samples to analytical facilities for assessment. Studies have assessed preservation methods for environmental bacterial samples with DNAgard, RNAlater, DESS, FTA cards, and FTA Elute cards. They found that liquid preservatives such as DESS were more effective in preserving the microbial DNA with all the liquid preservatives methods performing equally (Beknazarova, M. et al 2017, Gray, M.A. et al 2015). DESS is an affordable solution that could potentially replace cold storage during extensive transportation distances.

Figure 1.4. Map of Lake Warner, Hadley, MA the model used for the method development. (Source: Google Maps).

Figure 1.5. Section of the Artibonite River. The source of the cholera outbreak. (Source: Google Maps).

1.10 Thesis Aims

To address the increasing prevalence of waterborne diseases, particularly for developing countries, this thesis will focus on creating a reliable, rapid methodology in these remote areas by:

1) Evaluating the microbial diversity of the surface water from a local lake with a history of pollution and comparing two sequencing methods: Oxford Nanopore MinION and the MiSeq Illumina (Chapter 2)
2) Designing an infield filtration method for deployment in remote areas with no access to electricity (Chapter 3)
3) Evaluating preservation methods for transportation of samples requiring more than 6 hours of storage (Chapter 4)

CHAPTER 2

BACTERIAL DIVERSITY IN THE LAKE WARNER MODEL USING TWO SEQUENCING METHODS: OXFORD MINION AND MISEQ ILLUMINA

2.1 Background

Despite advances in technology, epidemics of waterborne infectious diseases still occur in the 21st century, such as the 2010 cholera epidemic in Haiti. The origin of the diarrheal disease was tracked to a major source of public drinking water, the Artibonite River (Piarroux et al 2011). Lake Warner and the Mill River located in Hadley, Massachusetts have previously exhibited high levels of coliform bacteria due to a long history of waste water pollution. Lake Warner is used as a model for this experiment to look at using WGS as a future monitoring tool for countries like Haiti. "State of the Lake" reports have indicated historically high *E.coli* counts near the inlet of the Mill River and Lake Warner (Johnson, J. 2015). Though progress has been made to lower the levels of *E. coli* pollution, no data have been collected on the bacterial diversity of Lake Warner or the river. Data on the microbial diversity would be beneficial to public health by providing a wealth of knowledge about bacteria, viruses, fungi, protozoa, antimicrobial resistance genes, and virulence genes, and could aid in source tracking of aquatic pathogens. In this comparison study, metagenomic sequencing of water samples conducted with two sequencing methods, coupled with the CosmosID bioinformatic analyses, will enable the development of a methodology that is potentially applicable for use in remote locations to identify all the microorganisms and genes of interest in that particular environment rather than any one particular organism. Second generation sequencing with instruments such as the Illumina is the most utilized method; however,

development of new technology (third generation) in the form of the Oxford Nanopore MinION sequencing device has the potential to yield rapid and actionable data in a portable, relatively inexpensive device (Roy, M.A. et al 2018).

2.2 Methods

2.2.1 Chemicals and Reagents

M-coliblue24 broth, DNeasy PowerWater kit (Qiagen), Sterivex PowerWater kit (Qiagen),

Rapid Barcoding Kit (RBK-SQK004), NEBNext Ultra FS Library Kit, NEBNext Multiplex Oligos (96 index), MiSeq Reagent Kit V2 (300 cycles), DMSO-EDTA-salt (DESS), 70% ethanol

2.2.1.1. Water Collection

Approximately 500 mL samples were collected in triplicate in sterile polypropylene bottles by hand dipping (with gloves) according to USGS protocols (U.S. Geological Survey 1997) from five locations around Lake Warner in September 2018 (Figure 2.1). Locations were selected based on the proximity to potential sources of anthropogenic pollution. Site 1 (42º23'.8 N, 72 º34'51" W) is located near the boat launch, where Friends of Lake Warner and Mill River frequently collect samples for bacterial levels by culturing (Johnson, J. 2013). Site 2 (42º23'13.0" N, 72º34'44.7" W) is downstream from the animal farm (home to cows, goats, chickens, and horses), and Site 3 (42º23'14.8" N, 72º34'47.3" W) is adjacent to an abandoned building. Site 4 (42º23'37.7" N, 72º33'56.6" W) is located near the second boat launch, and Site 5 (42º23'32.6" N, 72º33'40.7" W) is at the junction of the Mill River, where high *E.coli* counts were previously observed (Johnson, J. 2015). Primary wastewater samples were

collected in triplicate from Amherst Water Treatment Plant, Hadley, MA using a 500 mL metal cup with chain to dip down into the primary waste (provided by Amherst Water Treatment Plant) and deposited into sterile polypropylene bottles (while wearing gloves).

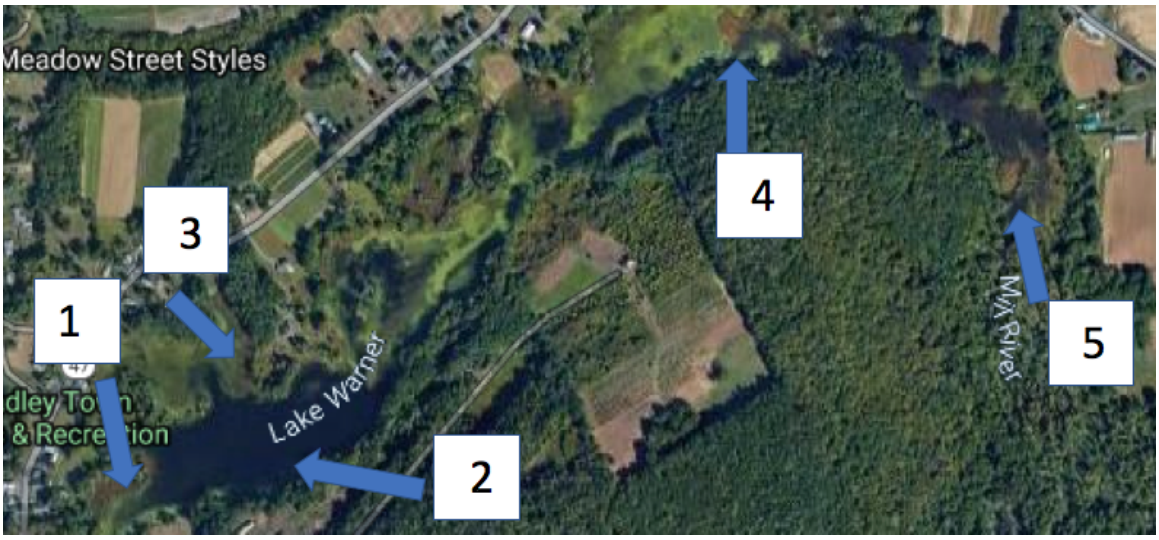

Figure 2.1. Map of Lake Warner located in Hadley, MA, with the 5 locations sampled at the end of September 2018. (Source: Google Maps).

2.2.1.1.1. Filtration

Approximately 400 mL of each triplicate lake water samples was filtered through Millipore Sterivex filter units with an attached vacuum system and the remaining 100 mL triplicate samples were enriched with m-Coliblue24 enrichment broth to specifically target coliform bacteria and *E.coli,* but filtered differently with a laboratory filtration system using membrane filters. Lake water triplicate samples for enrichment were filtered through 50 mm, 0.45 µm pore size, sterile membrane filters using a laboratory filtration system (GN-6 Metricel, Pall corporation, Ann Arbor, Michigan) for subsequent DNA isolation. Each filtration tower was thoroughly disinfected with 70% ethanol and

deionized water as well as air dried prior to each filtration. To enrich for the coliform group of bacteria, samples were grown following the manufacturer's protocol on m-ColiBlue24 using 100 mL from each triplicate sample from each location. Sterile 50 mm petri dishes containing absorbent pads were saturated with 2 mL of m-ColiBlue24 broth for enrichment of coliform and *E. coli* from Lake Warner. The enriched samples enhance coliforms and *E.coli* that could be potential public health risks that are difficult to detect in unenriched samples, due to the high concentrations of other naturally occurring bacteria.

The remaining 400 mL of each lake water triplicate sample was filtered through Millipore Sterivex filters units using a vacuum apparatus which included a caulking gun to push the water through the syringe into the filter, polypropylene tubing for the filtrate, and a vacuum to release the pressure from the filter unit. Due to the high amount of DNA yield required to sequence on the Oxford Nanopore MinION (400 ng), the samples were spiked. The spike consisted of filtered 50 mL triplicate primary wastewater through the 50 mm sterile membrane filters using the vacuum manifold and placed in sterile 50 mm petri dishes with 2 mL of m-Coliblue24 broth. The samples were incubated at 35ºC for 24 hours.

2.2.1.1.1.1. DNA isolation

The enriched and spiked samples were incubated at 35ºC for 24 hours followed by DNA isolation following the PowerWater DNeasy Kit (Qiagen) instructions. DNA quantity was measured on a Qubit Fluorometer (Thermo Fisher Scientific) performed by the Genomic Sequencing Laboratory (UMass Amherst), demonstrating that samples

contained <400 ng. The Oxford Nanopore MinION requirements for DNA yield are ≥400 ng, whereas the MiSeq Illumina requires only 0.1ng – 100ng of DNA. The unenriched samples were immediately isolated following the Sterivex PowerWater kit (Qiagen) instructions.

2.2.1.1.1.1.1. DNA library Preparation

The concentrations for every sample were significantly lower than the required amount for the Oxford Nanopore MinION library kit. The concentrations were also too low to utilize the isolated spiked DNA (507 ng/ul) without overwhelming the samples for the Oxford Nanopore MinION; therefore, isolated DNA from the preservation experiment (Chapter 4) was used for sequencing on both the MiSeq Illumina and Oxford Nanopore MinION. The 36 isolated DNA samples from the preservation experiment were spiked (507 ng/ul) with 1ul of the enriched wastewater because the concentrations were still lower than the required amount for the Oxford Nanopore MinION; however, the isolates were higher than the lake water isolates. Part of the spiked isolated DNA samples were library prepped following the NEBNext Ultra FS II Library Prep Kit for Illumina protocol with Multiplex Oligos (#E7335L). Since this experiment is comparing the findings between the two sequencing methods instead of identifying the bacterial community, the spike will not be accounted for after analysis. Metagenomic DNA produces long strands of DNA (>1.2 kb), but the preferred DNA range for MiSeq Illumina is 200-800 bp; therefore, the MiSeq fragmentation kit was selected to ensure proper DNA fragment sizes. The low DNA yield could be due to the DNA library kit selected that required many ethanol washing steps in PCR tubes that could have increased error.

Un-spiked sample triplicates from site 1 and 4 showed consistent concentrations and were chosen to run on only the MiSeq Illumina. Given that the protocol DNA range is significantly lower than the Oxford Nanopore MinION requirements, these samples will account for the diversity of Lake Warner while the preservation experiment samples will serve as the comparison for the two sequencing methods. The remaining amounts of the spiked isolated DNA were library prepped using the Oxford Nanopore MinION Rapid Barcoding Kit (SQK-RBK004) instructions with up to 12 indexes (preservation experiment DNA samples. The library DNA was quantity and quality checked using the Qubit Fluorometer and BioAnalyzer for the MiSeq Illumina samples (performed by the UMass Amherst Genomic Sequencing Laboratory) before sequencing.

2.2.2 Sequencing

- *MiSeq Illumina*

Samples were prepared for sequencing following the MiSeq Reagent Kit v2 Nano 300 cycle protocol (prep and pool samples <1 hour before MiSeq Illumina loading), with 251 base paired-end sequencing chemistry. The kit has an expected output of 500 Mb data and 2 million paired reads. The pooled samples (95) were sequenced for 48 hours.

- *Oxford Nanopore MinION*

Platform quality control tests were conducted using MinKNOW sequencing software before sequencing of samples to verify the number of active pores in the SpotON flow cell for the Oxford Nanopore MinION run. The three flow cells contained over 400 active pores in each. The 36 samples were split into 3 different runs, the first 12

samples were sequenced in the Oxford Nanopore MinION following the Rapid Barcode Kit (SQK-RBK004) for 2 hours. The 12 samples for the second run had the rapid sequencing adapter (RAP) added and preserved at -20ºC due to Oxford Nanopore Technologies (ONT) software update that resulted in the basecalling taking 24 hours after sequencing, rather than during sequencing, and therefore required overnight processing. However, after basecalling was completed, this second run could not be processed and was eliminated from the experiment.

2.2.3 Bioinformatics

The FASTQ files (for both) were sent to CosmosID, a microbial genomics platform that identifies virulence genes, pathogens, AMR, etc., using high performance data mining algorithms and GenBook, a database of 150,000 microbial genomes (CosmosID n.d.). The samples were de-multiplexed by CosmosID, meaning that each sample was barcoded with a sequence to identify the individual sample when they are pooled together for sequencing, this is called multiplexing.

2.2.4 Statistical Analysis

The resultant fastq files from the MinION and MiSeq Illumina were processed and analyzed using the CosmosID bioinformatic platform and GenBook, a genomic database with hundreds of millions of marker sequences and 150,000 microbial genomes. The database is organized in phylogenetic trees that accurately and precisely identify not only the microbial diversity within the sample, but the fungi, viruses, bacteria, protists, antibiotic resistance and virulence genes. The alpha diversity indices, beta diversity distance matrices, and species richness were calculated from the taxa abundance tables.

The Principal Component Analysis (PCA) was generated by clustering the samples by abundance to quantify the similarity using a covariance matrix of normalized data.

2.3 Results

2.3.1 Lake Warner Diversity Non-Spiked Samples

The results provided by CosmosID show the bacterial diversity in relative abundance (the number of each species), frequency (the number of reads that hit the species), and total percent match (the shared matches between species that accounts for the unique matches) (CosmosID, n.d.). The total relative abundance for bacteria identified by MiSeq Illumina for non-spiked site 1 unenriched (Figure 2.2) showed the predominate phylum of bacteria to be *Arcobacter* 15563 at 56% with total hit frequency of 6 and unique hit percentage of 6.19%. The second highest bacteria relative abundance was *Acinetobacter* 41878 at 8.47% with total hit frequency of 5 and unique hit percentage of 1.17%. The third highest was both *Empedobacter* and *Tolumonas* at 5.30% with total hit frequency of 16 and 13 along with the unique hit percentages of 0.29% and 45%, respectively.

Figure 2.2. Sunburst visualization of site 1 unenriched bacterial diversity sequenced on MiSeq Illumina.

The unenriched sample for site 4 (not spiked) bacterial diversity sequenced on MiSeq Illumina showed (Figure 2.3) the top bacteria was *Leclercia adecarboxylata ATCC 23216* at 36.90% with total hit frequency of 18,475 and unique hit percentage of 8.75%. *Enterobacter cloacae EcWSU1* at 20.91% was the second highest abundant

bacteria with total hit frequency of 461 and unique hit percentage of 1.19%. *Enterobacter sp. E20* at 8.66% had the third highest relative abundance with a total hit frequency of 1126 and unique hit percentage of 1.66%. Other bacteria of importance include *Escherichia coli* at 3.19%, *Klebsiella pneumoniae* at 7.17%, *Atlantibacter hermannii NBRC 105704* at 5.62%, and *Enterbacteriaceae bacterium ATTCC 29904* at 1.08%. A variety of bacteriophages were identified in site 4 (Figure 2.4) from the metagenomic analysis, with a number of phages associated with potentially pathogenic bacteria, notably *Escherichia virus Mu* at 11.25% relative abundance, *Escherichia phage phiV10* at 4.58%, *Shigella phage SfIV* at 1.81%, and *Shigella phage Sf6* at 1.45%.

Figure 2.3. Sunburst visualization of site 4 unenriched bacterial diversity sequenced on MiSeq Illumina.

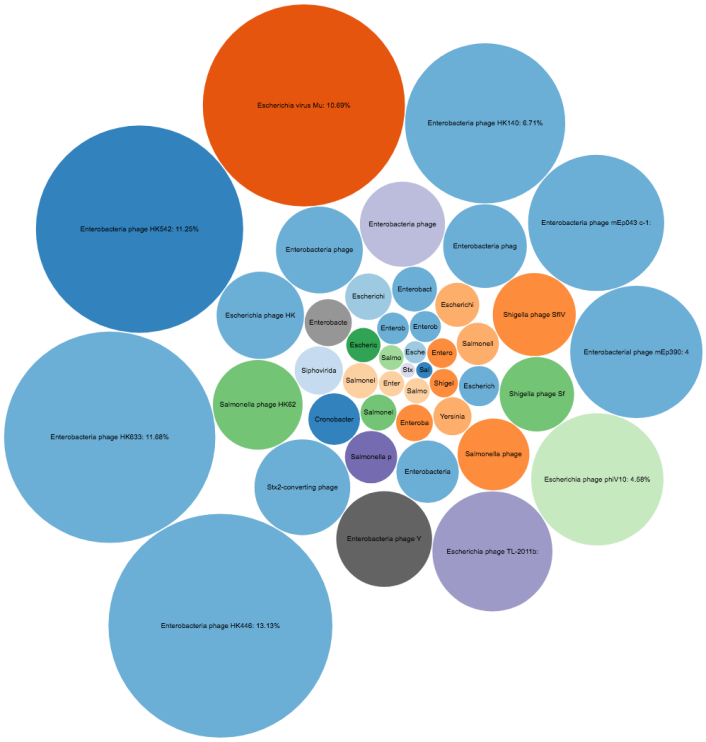

Figure 2.4. Bubble visualization of unenriched sample from site 4 of bacteriophages sequenced on MiSeq Illumina.

Among the clinically significant bacteria identified in the lake water metagenome from the non-spiked site 1 and site 4 (Table 2.1) were *Acidovorax spJHL_3, Aeromonas, Escherichia coli,* and *Exiguobacterium* which are potential opportunistic pathogens, with *Acidovorax spJHL_3* dominating in frequency at 229 along with *Aeromonas* at 69 and *Escherichia coli* at 45. The associated health effects from the pathogenic strains could cause sepsis, bloodstream infection, and Crohn's disease among other adverse health outcomes. Only one bacterial species is located in both site, *Acinetobacter* 41878, and the rest were only identified in one of the two locations.

Bacterial Species	Site	Relative Abundance (%)	Disease Association
Arcobacter 15563	1	56	Evidence of human and animal pathogen, underestimated, limited known
Klebsiella pneumoniae	4	20	Can cause destructive changes to human lungs; bloody sputum
Enterobacter sp.	4	8.7	Urinary infection; respiratory tract infection
Acinetobacter 41878	1,4	8.5	Associated with infections like skin infections and even meningitis
Aeromonas	1	4.9	Endophthalmitis; fish pathogen
Acidovorax spJHL3	1	2.6	Bloodstream infection; sepsis
Escherichia coli	4	1.8	Urinary tract infections; neonatal meningitis; gastroenteritis, and more potential severe outcomes
Exiguobacterium	4	0.2	Virulence factors that play a role in lethal infections by bacteria

Table 2.1. Clinically significant bacteria species identified in Lake Warner from locations 1 and 4 (official boat launch and unofficial boat launch) sequenced on MiSeq Illumina.

After the m-Coliblue24 enrichment of the site 1 water sample (not spiked), metagenomic analysis showed in Figure 2.5, the relative abundance of *Enterobacter cloacae UCI 49* was 13.29% with total hit frequency of 388 and unique hit percentage of 0.59%; *Klebsiella* 12.82% with total hit frequency of 13 and unique hit percentage of 0.98%; and *Raoultella ornithinolytica* at 11.43% with total hit frequency of 45 and unique hit percentage of 0.17%. A variety of bacteriophages were identified (Figure 2.6), with amongst the highest relative abundance being *Enterobacteria phage mEp460* at 33.10%, *Enterobacteria phage HK446* at 27.29%, along with *Shigella phage SfIV* at 1.03%, *Escherichia phage HK639* at 4.25%, and *Stx2* converting phage 1717 at 3%. While the *Stx2* converting phage was detected through analysis of the m-Coliblue24 enriched coliform group, it was undetected among the unenriched lake samples.

Figure 2.5. Sunburst visualization of site 1 m-Coliblue24 enriched sample bacterial diversity sequenced on MiSeq Illumina.

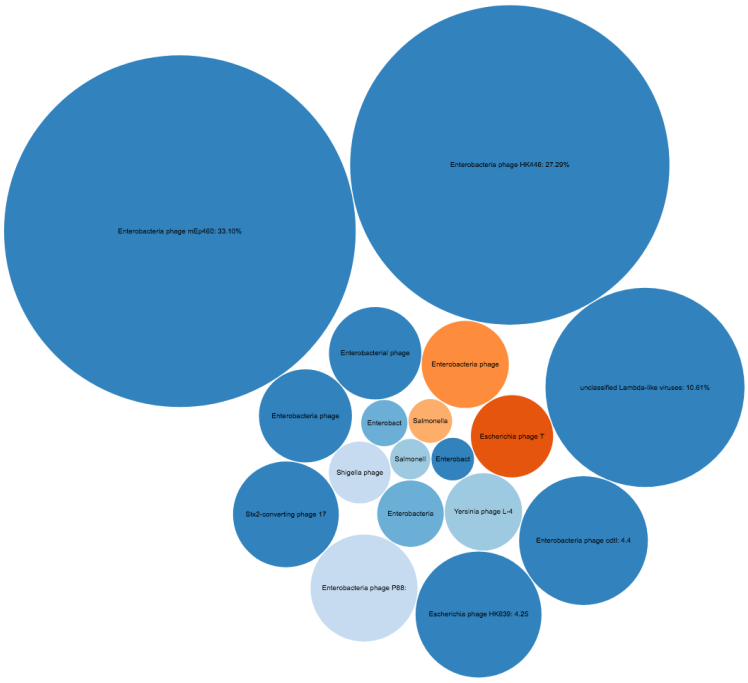

Figure 2.6. Bubble visualization of m-Coliblue24 enriched sample from site 1 of bacteriophages sequenced on MiSeq Illumina.

M-Coliblue24 enriched samples for non-spiked site 4 showed (Figure 2.7) a high relative abundance for *Polynucleobacter* at 83.09%, dominating the overall bacterial diversity along with total hit frequency of 34 and unique hits percentage 4.66%. Only three bacteria followed: *Flavobacterium 289* at 13.23%, *alpha proteobacterium SCGC AAA280-P20* at 2.13%, and *Limnohabitans* at 1.55%. No viruses were detected.

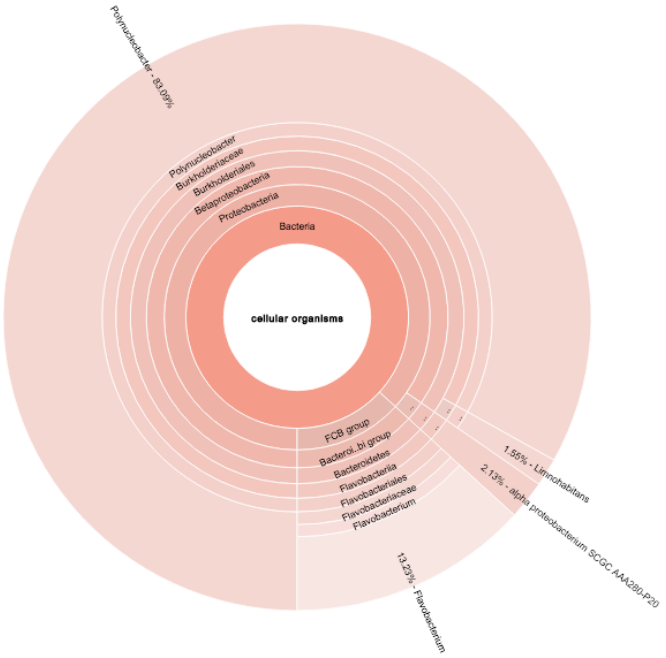

Figure 2.7. Sunburst visualization of the bacterial diversity of site 4 m-Coliblue24 enriched sample sequenced on MiSeq Illumina.

2.3.1.1 Spiked Samples for Oxford MinION and MiSeq Illumina Results

The same data sequenced by the Oxford Nanopore MinION and MiSeq Illumina are depicted in the PCA plot shown in Figure 2.8 for the total match percentage. This demonstrates the similarity and dissimilarity between the two methods. There are two outliers, the 10 day without preservation sequenced on the Oxford Nanopore MinION and the 7 day DESS preservation stored at the 95ºF sequenced on the Oxford Nanopore MinION. Given that the samples were the same, the points should be overlapping, yet the remaining points are sporadically distributed indicating that these two methods are producing dissimilar results in terms of bacterial species diversity in total match percentage. In Figure 2.9, one sample (immediate isolated) sequenced on both the MiSeq

Illumina and Oxford Nanopore MinION is compared for the total match percent of the bacterial diversity, the two points are different. The sample sequenced on Oxford Nanopore MinION in terms of percent match percentage shows different results than the MiSeq Illumina sequencing of the same sample.

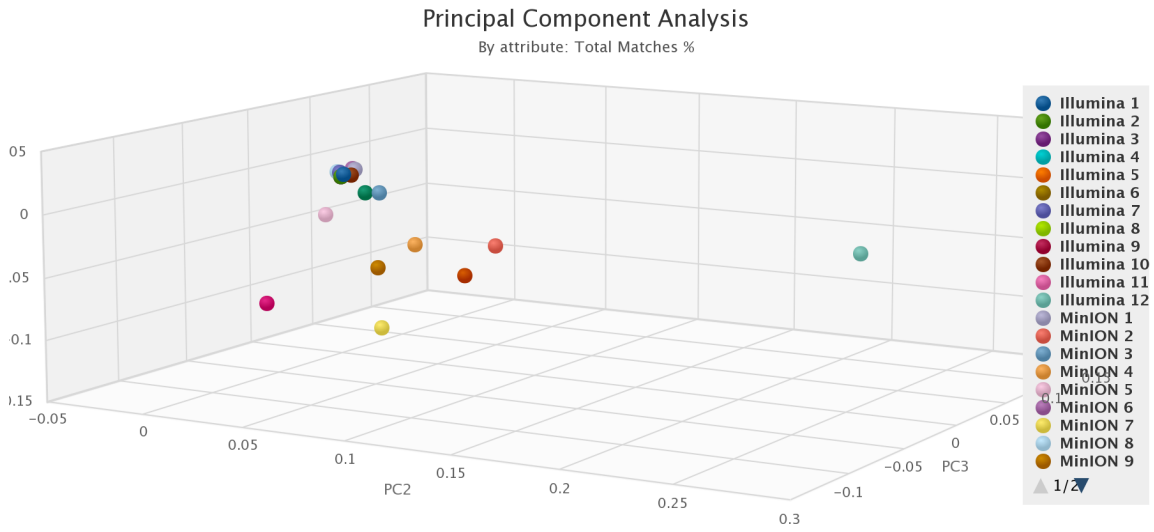

Figure 2.8. The total percent match of the bacterial diversity of the same data sequenced with both sequencers.

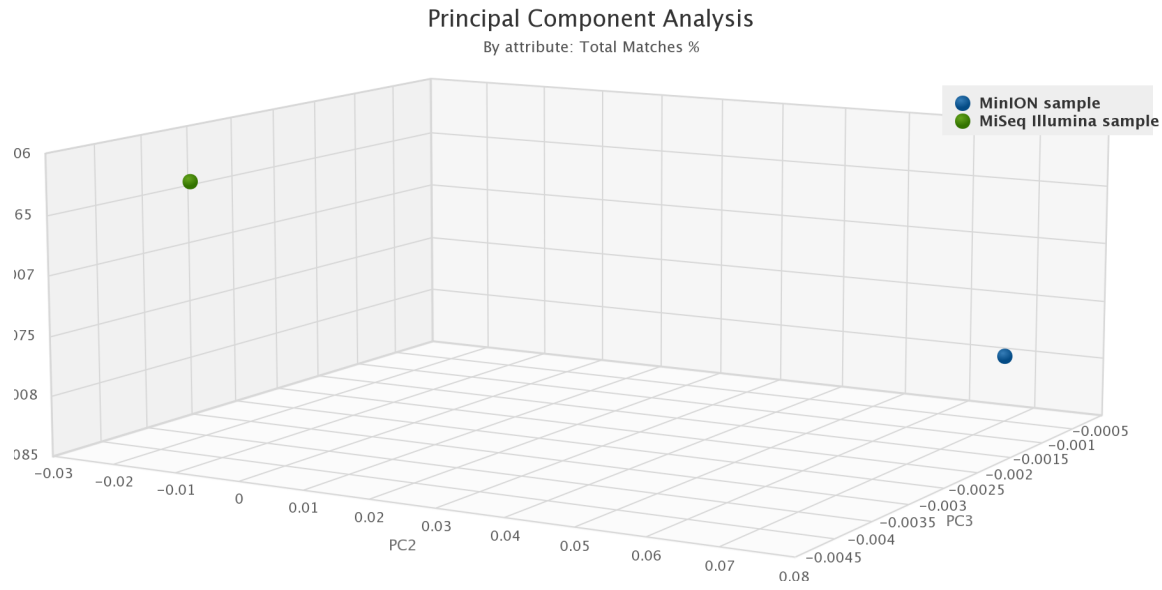

Figure 2.9. The total percent match of the bacterial diversity of one sample sequenced with both sequencers.

When the data are depicted as a PCA showing the relative abundance of bacterial species diversity in the same dataset sequenced by the MiSeq Illumina and Oxford Nanopore MinION (Figure 2.10), the results are spread throughout the plot with high variabilities. In Figure 2.11 one sample (immediate sample) is compared in terms of relative abundance of bacterial diversity sequenced on both Oxford Nanopore MinION and MiSeq Illumina. Similar to Figure 2.9, the PCA plot shows that the relative abundance of the bacterial diversity on the same sample sequenced on two different sequencers is dissimilar.

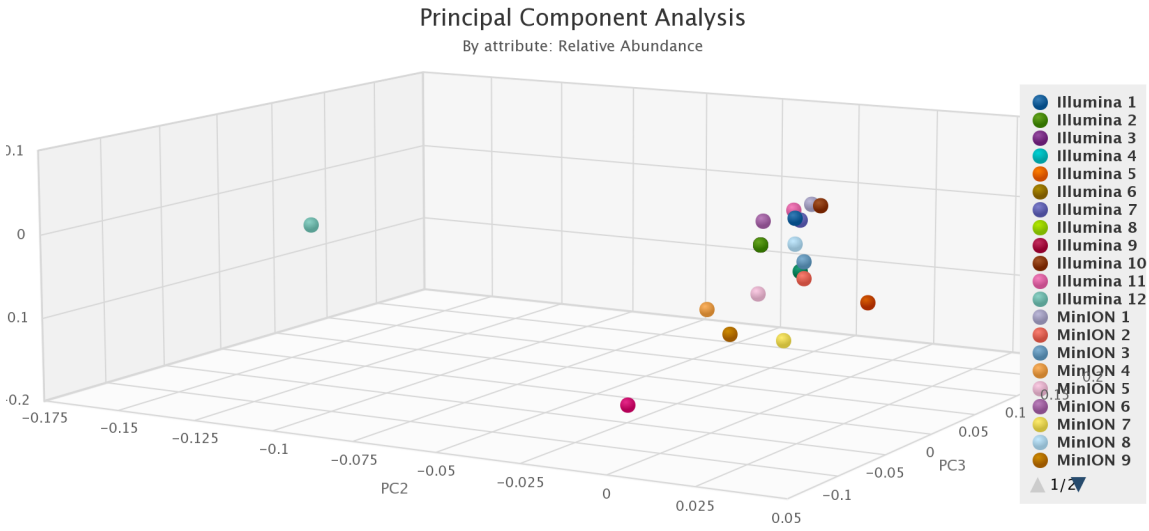

Figure 2.10. Relative abundance of the bacterial diversity of the same data set sequenced with both sequencers.

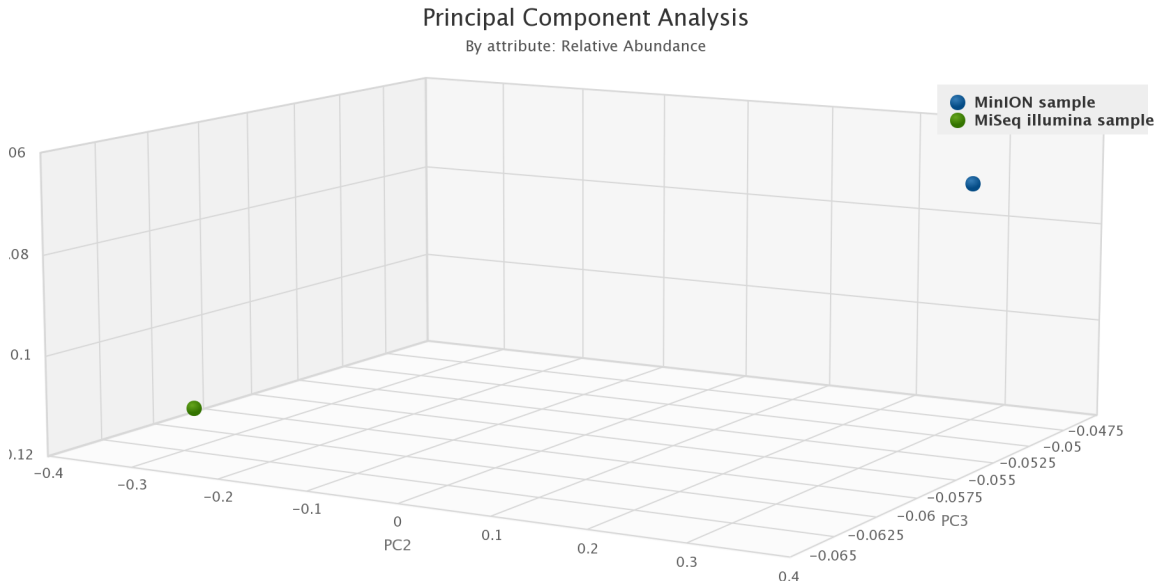

Figure 2.11. Relative abundance of the bacterial diversity of one sample sequenced with both sequencers.

Figure 2.12 and 2.13 show a specific sample presented as a sunburst of the microbial diversity that was sequenced on both MiSeq Illumina and Oxford Nanopore MinION, where the top relative abundant bacteria sequenced on the MiSeq Illumina were *Aeromonas media WS* at 8.88%, *Tolumonas* at 7.85%, and *Flavobacterium sasangense DSM 21067* at 6.69%. The MiSeq Illumina identified a total of 37 bacteria while the Oxford Nanopore MinION identified 50 bacteria with the top relative abundant bacteria being the *Acinetobacter baumannii* at 27.12%, *Klebsiella* at 9.99%, and *Acinetobacter sp. MDS7A* at 5.66% along with *Escherichia coli* at 5.63%. *Tolumonas* was identified at a lower level (0.80%) than the MiSeq Illumina sequence method identified as well as *Aeromonas media* (1.78%) and *Flavobacterium sasangense DSM 21067* (1.39%). Compared to the Oxford Nanopore MinION, the MiSeq Illumina did not identify *Acinetobacter baumannii, Cronobacter sakazakii, Enterobacter cloacae* complex, *Klebsiella* and more; however, the Oxford Nanopore MinION did not identify *Aeromonas* media, *Caloramator*, *Bdellovibrio bacteriovorus*, *Comamonas testosteroni*, and more.

Figure 2.12. Sunburst visualization of the immediate preservation isolation sequenced on the MiSeq Illumina.

Figure 2.13. Sunburst visualization of the immediate preservation isolation sequenced on the Oxford Nanopore MinION.

In Figure 2.14, the heat map provides a visual summary comparing a portion of bacteria identified by both sequencing methods. The colors indicate the range of values from lowest to highest observed, where red would represent the highest abundance. A similar abundance was found for the *Acinetobacter* and *Arcobacter*; however, the

remaining bacteria were dissimilar such as *Aeromonas* and the less than certain relative abundance determined in the analysis is indicated in blue.

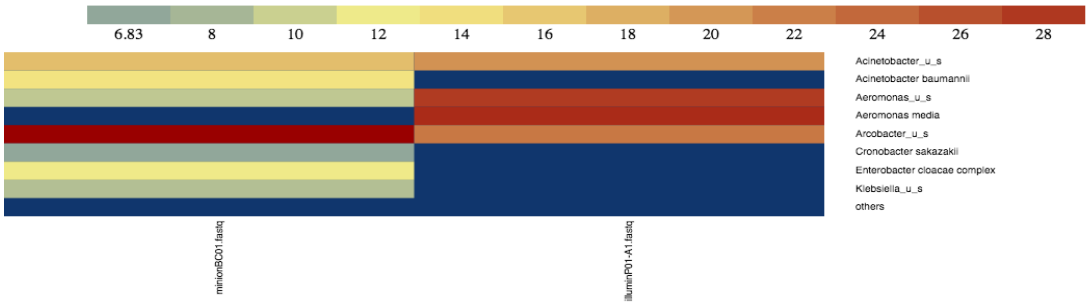

Figure 2.14. Heat map of the bacteria identified for sample 1, sequenced on both sequencers.

2.4 Discussion

Prior to this study, no sequenced data on the microbial diversity of Lake Warner existed; the only available data are on cultured samples from several locations throughout the lake. Although culturing is a low cost and relatively simple method to monitor for coliforms, it provides limited information on the total bacterial community present in a sample. Sequencing the environmental samples helps characterize the microbial diversity and identify genes and viruses of potential public health interest, such as the toxin converting phages. Roy et al. (2018) collected water samples from various locations near the epicenter of the 2010 cholera outbreak on the Artbonite River in Haiti. The researchers identified virulence associated and antibiotic resistance genes by using the metagenomic approach for environmental samples, through shipping samples to the U.S. for WGS using Ion Torrent instrumentation. The cholera toxin converting phage (CTXφ)

was found in one location used for bathing, washing clothing, and other daily activities where accidental consumption is highly possible. CTXφ is required for the *Vibrio cholerae* pathogenicity, and therefore this location could present risks for new cholera infections. In addition to identifying the converting phage for the cholera toxin, Shiga toxin converging phage was found in 70% of the samples collected and a high abundance of *E. coli* was observed in samples collected after the rainy season (July). However, samples with *Enterohemorrhagic E. coli 0157:H7* were detected at one site in a January sample (dry season), suggesting that fecal contamination is not simply related to the flooding events. This method could be a powerful tool for monitoring surface waters for potential pathogens that could contribute to waterborne disease outbreaks, and hence inform prevention (Roy, M.A. et al 2018). Use of the portable, handheld Oxford Nanopore MinION could provide an onsite screening tool to sequence data out in the field.

The PCA plot of percentage match of bacteria, comparing the two sequencing methods showed that the Oxford Nanopore MinION and MiSeq Illumina produced dissimilar results. The PCA plot of bacterial species relative abundance for the two methods were sporadically distributed with no similarity between the same sample sequenced on both sequencers. Another PCA plot with one sample sequenced on both MiSeq Illumina and Oxford Nanopore MinION confirmed these results with the two points being different. Given the MiSeq Illumina's ability to produce more reads than the Oxford Nanopore MinION, this sequencer should identify more bacterial species, yet it only identified 37 bacteria in sample 1 while the Oxford Nanopore MinION identified 50. These results could be due to errors created during the five ethanol washing steps for the

library preparation that may have affected the downstream MiSeq Illumina sequencing process. However, the heat map (Figure 2.14) showed that the Oxford Nanopore MinION produces more diversity with lower quantities of each bacterial species, while the MiSeq Illumina showed fewer species, but in larger quantities. These results are not surprising given that the MiSeq Illumina uses PCR and the Oxford Nanopore MinION does not. This additional PCR step in the MiSeq Illumina causes the lower level bacterial species to be lost while the bacterial species present in higher levels will be amplified (Caporaso. J. et al 2012).

An important aspect of this study in terms of surveillance of clinically significant bacterial species within a body of water is that only one clinically significant bacterial species, *Acinetobacter* 41878, was found in both locations (the official boat launch and the unofficial boat launch). These two locations experience high human activity, though only one of these locations, location 1 (official boat launch) is the only area that is tested for fecal coliforms (Johnson, J. 2015), interestingly location 4 (unofficial boat launch) was the only location that had levels of *E. coli* present. Fecal coliform is used as an indicator for water quality and Lake Warner follows the Massachusetts water quality standards for primary contact use, meaning that this body of water follows the standards for swimming (Johnson, J. 2015). These findings show the importance of testing for fecal coliform in not just one location throughout the lake, given that the lake experiences high levels of human activity and experiences nonpoint source pollution and internal loading from previous years of primary waste water pollution, more locations should be tested to ensure that the water quality is safe for the primary contact standards of Massachusetts (Johnson, J. 2015).

Another benefit of the Oxford Nanopore MinION is the quick and easy library prep protocol that can take as little as 30 minutes, decreasing the likelihood for error. Local individuals in countries such as Haiti can be easily trained to use the Oxford Nanopore MinION as a monitoring tool, linked to a waterborne disease prevention program. The MiSeq Illumina procedure is tedious and time consuming, which can increase user errors. In addition, it is impractical for use in isolated areas. Future research to establish a sterile in field procedure for the DNA isolation and library preparation, coupled with the portable, handheld Oxford Nanopore MinION would enable health departments to assess bodies of water after extreme weather events especially when people are isolated from safe drinking water.

A limitation of this study was the high DNA requirement to sequence samples on the Oxford Nanopore MinION, which requires at least 400 ng of DNA compared to the MiSeq Illumina, which requires as little as 0.1 ng of DNA. Lake Warner is cleaner than most bodies of water in developing countries and the amount of isolated DNA was low. However, more optimization of the sample collection and preparation is necessary before the Oxford Nanopore MinION can provide a reliable tool for rapid detection of microbial pathogens, antimicrobial resistance and virulence genes. It is concerning that established technologies such as the MiSeq Illumina provide different results from the Oxford Nanopore MinION. More comparative studies are necessary to establish the Oxford Nanopore MinION as an alternative for WGS in remote, low resource environments. The promise of the technology for rapid, actionable diagnostics following, for example, extreme weather events, make further research worthwhile. Even if access to the MiSeq Illumina or equivalent instrumentation was possible in countries like Haiti, library

preparation alone requires over a day to complete (depending on the number of samples), which does not include the sequencing time that could potentially take 24 to 48 hours. This is opposed to the Oxford Nanopore MinION, which is capable of sequencing within 2 hours. In Mitsuhashi. S. et al.'s (2017) study, the researchers used the Oxford Nanopore MinION to sequence a known mock bacterial community that consisted of 20 bacteria. The handheld sequencer detected all of the bacteria species, as well as the relative abundance within 5 minutes, similar to a 4 hour run with the Oxford Nanopore MinION. These results were compared with the 16S rDNA sequencing method, finding that both sequencing methods showed similar results (no comparison with Illumina was conducted). The researchers concluded the Oxford Nanopore MinION can be used for 2 hours to determine the bacterial composition for more complex samples (Mitsuhashi. S. et al 2017).

Additionally, seasonality data was not evaluated for this study on Lake Warner, because there was only one collection time during the end of September for the five locations with insufficient DNA yield for analysis by Oxford Nanopore MinION. Roy et al. (2018) reported a substantial difference in the microbial diversity between the winter and summer months in Haiti. Collection during the summer months especially after rainfall events could have produced enough DNA to sequence on the Oxford Nanopore MinION.

CHAPTER 3

ESTABLISHING FIELD FILTRATION SYSTEM FOR DEPLOYMENT ON LAKE WARNER MODEL

3.1 Background

Water collection requires transport of samples in sterile bottles and electric vacuum manifolds for in laboratory filtration. The common filter membranes used in the laboratory setting can only filter small quantities of water (≤500 mL) especially if the sample is highly turbid (≥200 mL). This is due to the small pore size of the filters and requires the use of multiple membranes to produce enough DNA for sequencing on the Oxford Nanopore MinION. Additionally, the use of multiple filter membranes for the same sample increases the likelihood of cross contamination during the preparation for DNA isolation. Millipore Sigma manufactures the Sterivex filter units, a filter encased in a sterile vial. This enclosed design of the Millipore Sterivex filter unit prevents cross contamination of the filter membrane, which allows for easy handling and transportation when the filter unit is capped on both ends. When coupled with a vacuum pump, Millipore Sigma states that the device can filter 1,000 to 2,000 mL of water (Millipore Sigma. n.d). However, out in the field, the lack of electricity requires researchers to rely on other electrical power-independent devices such as a syringe to push the water through the unit. Due to the design of the Sterivex filter unit, the filter can only sustain a certain amount of pressure with just the addition of the syringe and no vacuum for depressurization. Therefore, a vacuum system is required to allow the pressure to be released to use this device effectively. Previous trials were conducted in the Ford

Laboratory with the Sterivex unit and syringe without vacuum; in these tests, researchers were only able to pump a maximum of 200 mL of river water before the pressure resistance prevented further filtration. The Oxford Nanopore MinION requires a minimum of 400 ng of DNA, and previous trials in the Ford lab have found that 200 mL of local water samples did not meet the DNA requirement for processing on the portable sequencer (Nanodrop readings were >20 ng/ul); therefore, a battery vacuum system for in-field filtration is essential to conduct field research using the Oxford Nanopore MinION.

3.2 Methods

Triplicates of Lake Warner water were pumped through the Millipore Sterivex using a syringe without a vacuum pump and the amount of water pushed through the filter and time taken were noted. The same pond water was pumped through Millipore Sterivex using a syringe with an attached portable vacuum with a makeshift PVC pipe, tubing, and polypropylene Erlenmeyer flask to create the de-pressurizer as well as the water waste reservoir (Figure 3.1). The time and amount of water pumped through using the portable vacuum was noted.

3.2.1. Filter Cartridge

The Millipore Sterivex is a filter encased in a vial tube and is an ideal filter to ensure no cross contamination if water is properly filtered. Maintaining a sterile environment during the pre-sequencing processes is critical to produce data that accurately represents the diversity of the sampled body of water. The filter is capable of

filtering 1,000 mL to 2,000 mL when coupled with a vacuum manifold; this is significantly greater than typical laboratory filter membranes that are not capable of filtering through 1,000 mL of water unless multiple filter membranes are utilized and then pooled through submersion in PBS solution. The PBS solution method could possibly introduce contaminants into the sample because of the constant handling of the filter membranes, making them less attractive for use in the field. The drawback of the Millipore Sterivex is that it is not able to filter the 1,000 mL of water if the user only uses a syringe to push the water through, because the filter unit can only handle a certain pressure until it needs a depressurizing system. If a depressurizing system is not incorporated, then the back pressure will push the water back up the syringe resulting in loss of sample due to too much pressure.

3.2.1.1. Vacuum Method

A portable vacuum would be required out in the field to ensure enough DNA yield to sequence on the Oxford Nanopore MinION. The syringe method coupled with a vacuum creates a depressurizing system allowing more water to pass through the filter unit. A Milwaukee 18V cordless battery-operated vacuum that retails for roughly $90, is a cheap and effective piece of equipment for field use. The battery-operated vacuum is coupled with a PVC pipe and tubing that connects to the waste reservoir with a separate tube to the vacuum.

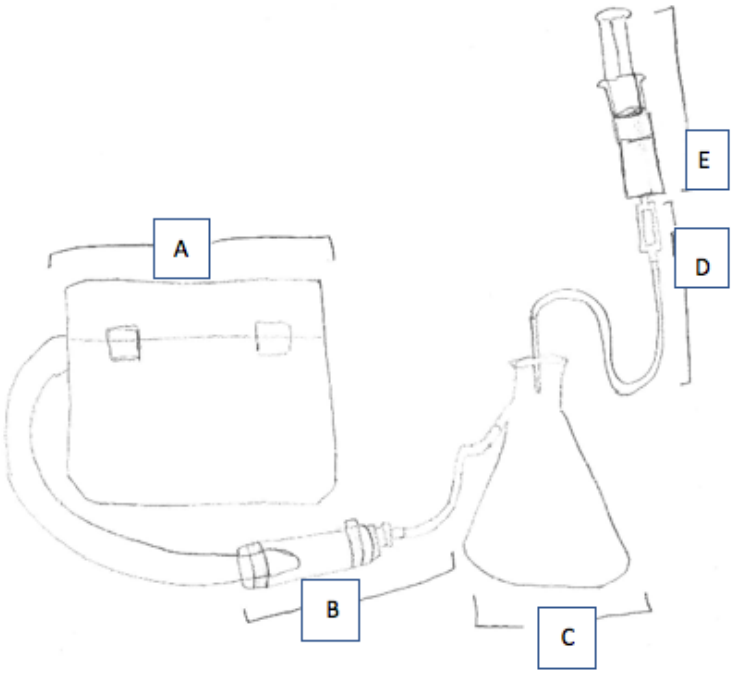

Figure 3.1. Portable filtration system design with (A) battery operated vacuum, (B) PVC pipe for attachment of vacuum hose to reservoir, (C) Erlenmeyer flask to serve as the filtrate reservoir, (D) Sterivex filter unit attached to tubing for filtrate waste, and (E) the syringe holding the water source to push through filter unit.

3.3 Results

The maximum amount of lake water capable of being pushed through the Sterivex unit with just the syringe was less than 200 mL due to the back pressure. Previous experiments have found that the DNA yield for 200 mL of filtered lake water was significantly less than the 400 ng required to sequence on the Oxford Nanopore MinION. After over 30 minutes of attempts to filter more than 200 mL without a vacuum pump, the filter unit was unable to handle more. The timeframe required to filter the 200 mL

averaged around 25 minutes. The more turbid the water sample, the less volume of water that could be pushed through the filter.

The portable vacuum manifold with the Sterivex filter unit was able to filter through 600 mL of lake water within 25 minutes and 1,000 mL in less than an hour. The addition of the vacuum manifold was able to far surpass the 200 mL that the syringe with no vacuum was capable of filtering.

3.4 Discussion

Eliminating the transportation of large amounts of sampled water by filtering onsite with the enclosed sterile Sterivex filter unit would be the optimal method to decrease any potential contamination during travel. The USGS recommends that immediate DNA isolation would be the ideal situation to produce the most accurate results to identify the bacterial community within a sample. Additionally, an in-field filtration system is a necessary component for creating an onsite microbial identification system using the Oxford Nanopore MinION. The small, portable design of the Sterivex filter unit makes it an ideal choice to filter the sampled water directly; however, even though the filter unit is capable of filtering 1,000 mL to 2,000 mL of water, this is only possible when coupled with a vacuum to allow the Sterivex to be depressurized.

Kirshstein, and colleagues were able to pump 50 to 2,300 mL of pond water by using a peristaltic pump and pumped the water through the Sterivex, but did not provide further details about whether the pump they used was battery operated (Kirshstein, J.D. et al 2007). Park et al. (2006) also utilized a peristaltic pump with a Sterivex filter unit and they were able to pump 5,000 to 10,000 mL of surface water through the units, yet with

no indication whether this pump was battery operated (Park, H. et al 2006). Other studies did not specify whether collected water samples were first transported to the laboratory and then pumped with an electric-powered vacuum. Bruce et al. (2012) and Hunt et al. (2013) collected seawater using the Sterivex filter units; however, unlike the studies by Kirshstein and Park, these researchers neglected to provide a more comprehensive filtration protocol. Hunt et al. (2013) noted that 11,000 to 19,000 mL of seawater were filtered through the units and Bruce et al. (2012) noted that 2,000 to 4,000 mL of seawater was filtered through the unit. It would be interesting to acquire the filtration protocol used for the Hunt study given the large amount of seawater the researchers were able to filter (Kirshstein, J.D. et al 2007, Park, H. et al 2006).

The portable vacuum manifold developed for this study would allow researchers to filter straight from the sample, although the Milwaukee 18V cordless pump could be replaced with a more powerful peristaltic pump. However, the previous studies used cleaner water than samples collected from Lake Warner, which likely accounts for the larger volume of water filtered in those studies.

CHAPTER 4

ESTABLISHING A PRESERVATION METHOD FOR SAMPLE TRANSPORTATION FOR METAGENOMIC ANALYSIS OF RIVER WATER FROM REMOTE AREAS

4.1 Background

Transporting water samples from remote locations to processing facilities requires preserving the samples to prevent DNA degradation. Changes in temperature of unpreserved samples during transport can result in microbial overgrowth of samples, resulting in inaccurate assessments of microbial diversity. The preferred preservation method is cold storage; however, under certain circumstances this method may not be readily available. Studies have found that liquid preservatives such as DESS have been effective for transportation of environmental samples for subsequent metagenomic analysis (Beknazarova, M, et al 2017, Gray, M.A. et al 2013). Though other liquid-based preservatives perform equally as well as DESS, the other options are expensive in comparison; therefore, DESS is the preferred option. The USGS recommends storing water samples in -20ºC for up to 6 hours before the DNA starts to degrade (U.S. Geological Survey. 1997). If cold storage is not a viable option in remote locations where samples must be transported over 6 hours, the addition of liquid preservatives may be a suitable method. One study using parasitic nematode samples evaluated ratios of 1:1 and 1:3 sample to DESS for time points of 0, 3, 7, 28, and 56 days. The researchers found that both ratios of DESS significantly prevented DNA degradation compared to the controls (Gray, M.A. et al 2013). This study aims to evaluate whether DESS preservation of

samples for metagenomic analysis, collected in Millipore Sterivex filters units, will have the same microbial diversity as the samples that are immediately processed or have been maintained in cold storage.

4.2 Methods

4.2.1 Chemicals

According to the DESS protocol, 46.53 g EDTA disodium salt and 100 mL of deionized water are mixed together. While heating the solution, 1 M NaOH is added until the EDTA disodium salt is dissolved and the pH reaches 7.5. 100 mL of DMSO is then added to a 400 mL aliquot to produce a 20% DMSO concentration and a 0.25 M of EDTA disodium salt concentration. The final step is the addition of 50 g NaCl (Gray, M.A. et al 2013). Phosphate buffer saline (PBS) was purchased from Fisher chemical.

4.2.1.1 Filtration

Primary wastewater effluent was collected at the Amherst Wastewater Treatment Plant located in Hadley, MA. The primary wastewater effluent was placed into a sterile flask and stirred to ensure homogeneity before filtration through Sterivex filter units. Approximately 50 mL of wastewater was filtered through each filter unit (N=57) using an electric vacuum pump and syringe system in the lab.

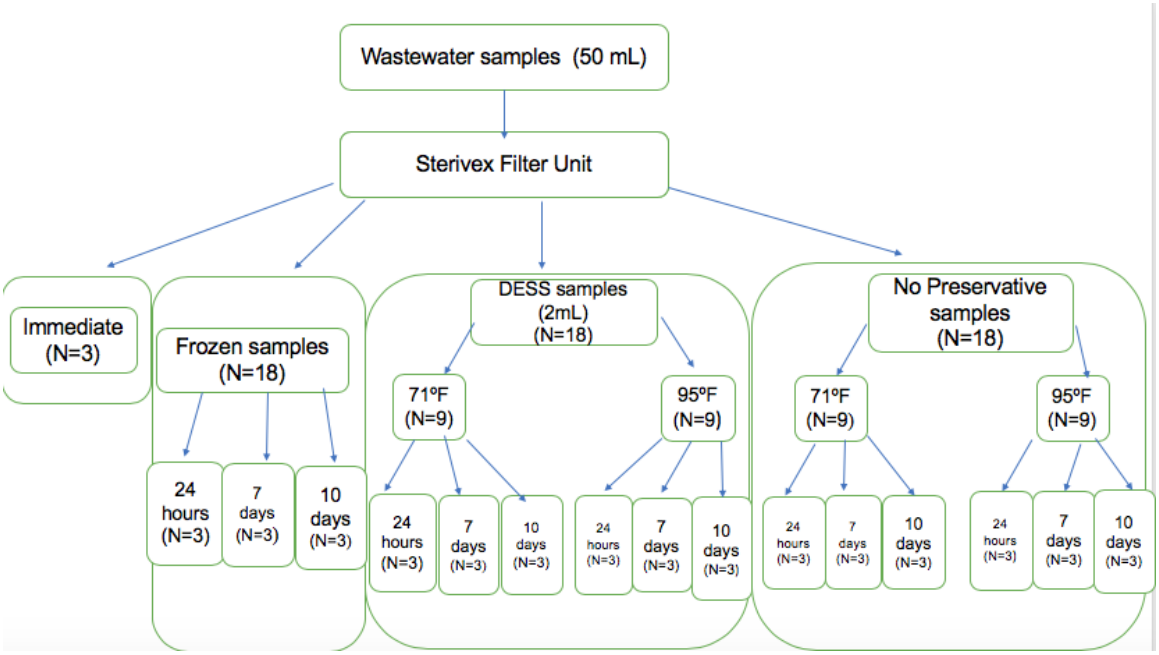

Figure 4.1. The experimental workflow for preserving wastewater samples by filtering through Sterivex filter unit with two positive controls (immediate and frozen at -20ºC), DESS preservation kept at two different temperatures, and without preservation kept at two different temperatures.

4.2.1.1.1 Preservation

Triplicate samples were immediately isolated to serve as the positive control and triplicate samples for each time period were frozen at -20ºC without preservation for a second positive control. The experimental design for the triplicate samples for each time period (24 hours, 7 days, and 10 days), temperature (71ºF and 95ºF), and preservation condition (with or without DESS) are depicted in Figure 4.1. The Sterivex samples preserved with DESS contained 2 mL of the preservative to completely submerge the unit while storing. Before the DNA isolation, the DESS was removed from the filter unit

using a syringe, and the filter unit was washed with PBS due to the high concentration of salt in DESS that could affect downstream applications.

4.2.1.1.1.1 DNA isolation

DNA isolation from the material storied on the Sterivex filter followed the PowerWater Sterivex kit protocol, described in Chapter 2.

4.2.1.1.1.1.1 DNA library preparation

Library prep for the isolated DNA followed the NEBNext Ultra FS II Library Prep Kit for Illumina protocol with Multiplex Oligos (#E7335L). Quality control was run on the BioAnalyzer for the library DNA and dilutions for pooling library DNA were performed (by the UMass Amherst Genomic Sequencing Laboratory) before sequencing.

4.2.2 Sequencing

The library DNA was sequenced on the MiSeq Illumina using the MiSeq Reagent Kit v2 Nano protocol.

4.2.3 Bioinformatics

The FASTQ files from MiSeq Illumina were uploaded to CosmosID, a microbial genomics platform that identifies virulence genes, pathogens, AMR, etc., using high performance data mining algorithms and GenBook, a database of 150,000 microbial genomes (CosmosID. n.d).

4.2.4 Statistical Analysis

The resultant fastq files from the MiSeq Illumina were processed and analyzed by the CosmosID bioinformatic platform that uses GenBook, a genomic database with hundreds of millions of marker sequences and 150,000 microbial genomes. The database is organized in phylogenetic trees that accurately and precisely identify not only the microbial diversity within the sample, but the fungi, viruses, bacteria, protists, and antibiotic resistance and virulence genes. The alpha diversity indices, beta diversity distance matrices, and species richness were calculated from the taxa abundance tables. The Principal Component Analysis (PCA) was generated by clustering the samples by the abundance to quantify the similarity using the covariance matrix of normalized data.

4.3 Results

The results provided by CosmosID show the bacterial diversity in relative abundance (the number of abundance of each species), frequency (the number of reads that hit the species), and total percent match (the shared matches between species that accounts for the unique matches) (CosmosID, n.d.). The PCA plot depicted in Figure 4.2 shows the similarity between the samples in terms of total percent matches. There are various outliers present with the 10 day frozen sample being the most dissimilar from the group, followed by the 10 day no DESS at 95ºF sample, 7 day with DESS at 95ºF sample, and 10 day with DESS at 95ºF sample. The remaining samples are considered similar with 10 day with DESS sample, 24 hour no DESS sample, and immediate sample slightly removed from the cluster.

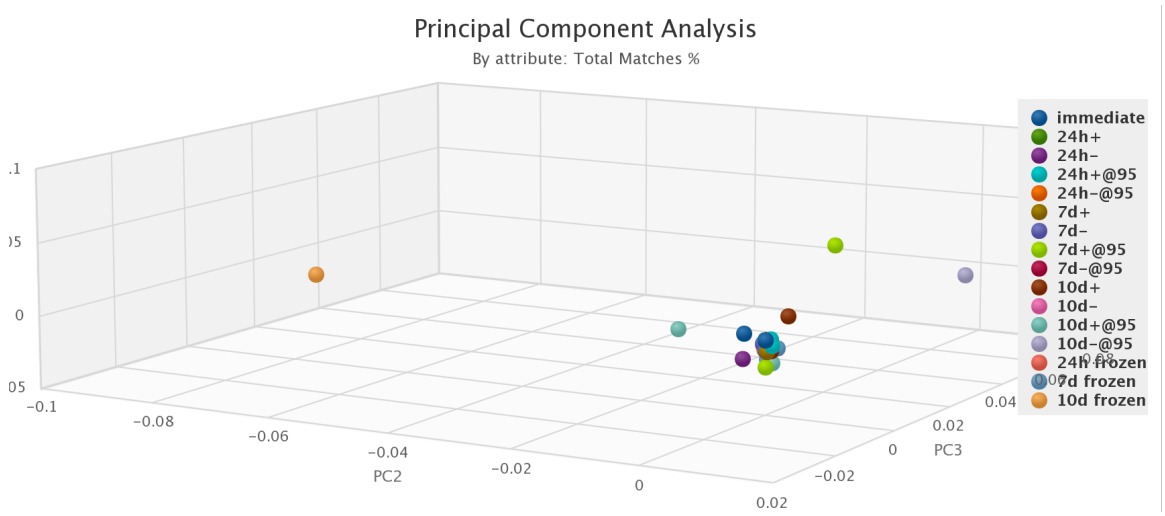

Figure 4.2. Principal Coordinate Analysis of the bacterial total percent matches for the preservation methods for 24 hour (24h), 7 days (7d), and 10 days (10d) with the immediate and frozen samples as the controls. The +/- indicates whether the filter was preserved with DESS or not.

The samples shown in Figure 4.3 are the 24 hour preservation DNA yields after the DNA isolation. The bar graph depicts the sum of the triplicates and only one of the positive controls (immediate) was used to compare the samples. The 24 hour with DESS, 24 hour without DESS, and 24 hour without DESS stored in 95ºF were all not significantly different than the control. The 24 hour with DESS stored at 95ºF was significantly different than the control, which is indicated by the asterisks.

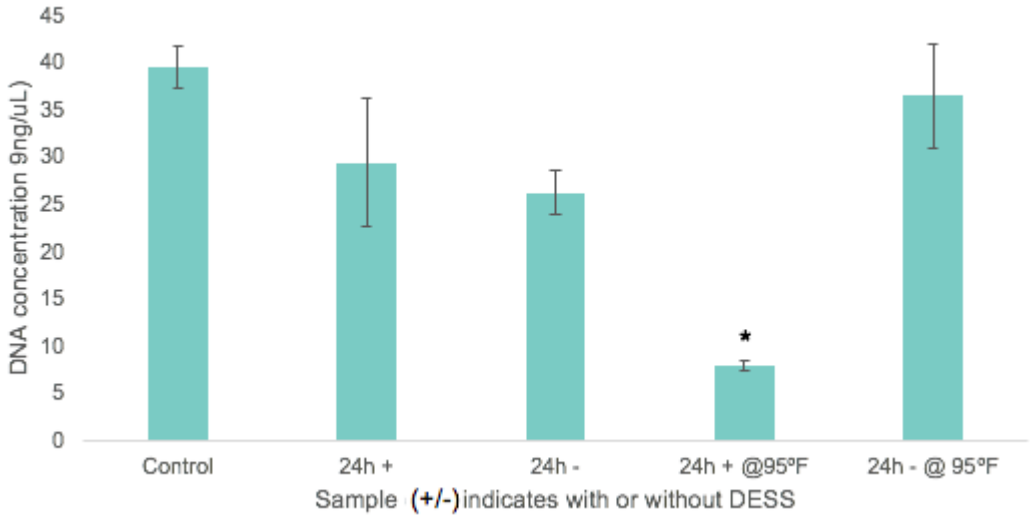

Figure 4.3. The DNA yield for 24 hour preservation after DNA isolation generated by JMP using T-test where * indicates a significant difference from the control

T The bar graph in Figure 4.4 shows the DNA yield of the 7 day preservation after the DNA isolation. The graph depicts the sum of the triplicates and only one of the positive controls (immediate) were used to compare to the other samples. The 7 day with DESS, 7 day without DESS, and 7 day without DESS stored at 95ºF were not significantly different than the control. Similarly, to the 24 hour bar graph, the 7 day with DESS stored at 95ºF was significantly different than the control, indicated by the asterisks.

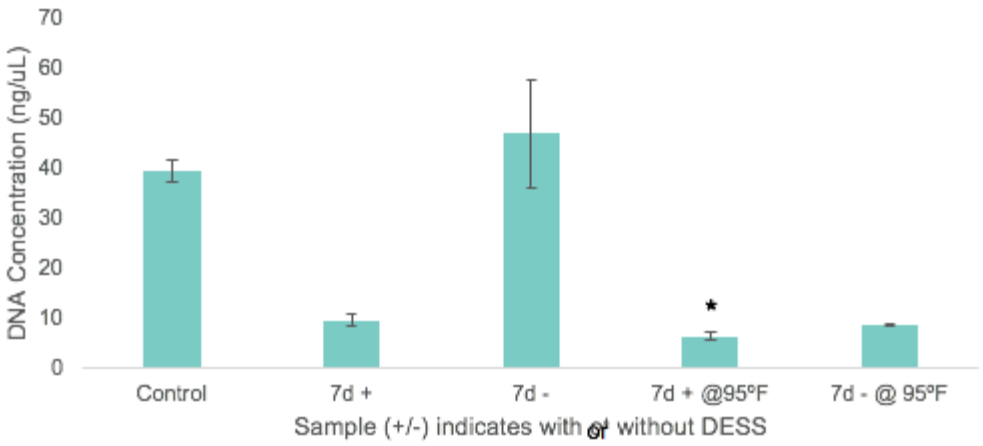

Figure 4.4. The DNA yield for 7 day preservation after DNA isolation generated by JMP using T-test where * indicates a significant difference from the control

The bar graph in Figure 4.5 shows the DNA yield of the 10 day preservation after the DNA isolation. The graph depicts the sum of the triplicates and only one of the positive controls (immediate) were used to compare to the other samples. Unlike the previous 24 hour and 7 day time periods, the 10 day preservation experiment shows that the 10 day with DESS, 10 day without DESS, 10 day with DESS stored at 95ºF, and the 10 day without DESS stored at 95ºF were all significantly different than the control, indicated by the asterisks.

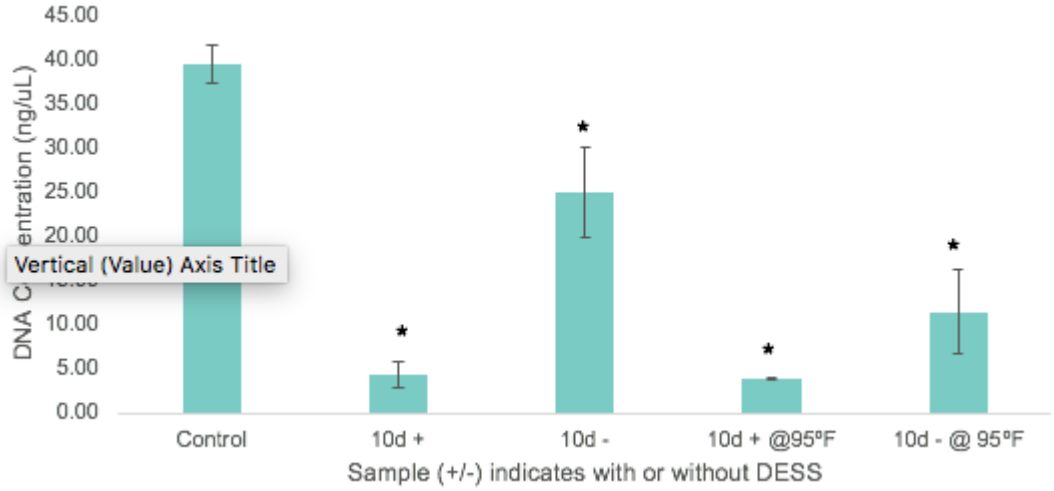

Figure 4.5. The DNA yield for 10 day preservation after DNA isolation generated by JMP using T-test where * indicates a significant difference from the control

4.4 Discussion

This study investigated a novel approach to preserve DNA from a filtered water source to provide insight and potential methodology for other investigators interested in preserving water samples for transportation or storage. Collecting water from sources with high human activity, whether used for bathing and/or washing clothing in rural, isolated areas could be used for monitoring pathogen occurrence, as described earlier in this thesis.

The ideal method would be immediate isolation after collection; however, this would require in field equipment to be employed out in remote areas. The most common method for water transportation is an ice box, but the total time samples are stored on ice

should not exceed 24 hours with the optimal time being ≤6 hours. If transportation to facilities for sequencing requires over 24 hours of travel, the DNA will degrade, and the microbial community will no longer represent the diversity of that body of water, because selected species will either grow or be inhibited.

Alternative preservation methods were explored by Gray and colleagues. These researchers investigated the ability of DNAgard, DESS, FTA cards, and FTA elute cards ability to preserve a known microbial composition that consisted of a mix of eight bacterial strains that were selected due to their capability to hinder preservation methods. This experiment spanned from 1-week intervals up to 3 months and the researchers found that the liquid-based preservations (DESS, RNAlater, DNAgard) identified more bacterial strains than the card-based preservation methods. The cluster analysis showed that all the liquid-based preservations performed similarly; however, the researchers found that the results from the best-case scenario presence/absence showed that over the average time periods, DNAgard outperformed the other liquid preservations with DESS following in second. Overall, the data from the DESS experiment showed that it would perform well in preserving mixed bacterial strains; however, the number of bacterial strains recovered for the 3-month time period dropped to 38%, suggesting that DESS is more effective when used as a preservation for less than 3 months (Gray, M.A. et al 2013).

This thesis study examined wastewater filtered through a Sterivex filter unit which was and then preserved with DESS over 3-time periods (24 hours, 7 days, and 10 days) at different temperatures (room temperature and 95ºF). The high temperature was included because, to our knowledge, no research studies have previously examined the

effect of temperature effects on DESS preservation, especially considering that many areas that experience extreme weather events have tropical climates. The PCA plot for the total percent matches showed some outliers with the 10-day frozen sample exhibiting a significant difference compared to the other outliers. These results could be due to the use of primary wastewater that contained small sediment pieces containing bacteria; although the wastewater was pooled and homogenized prior to filtration, some samples could have contained more sediment than others. The 10-day frozen sample did contain higher DNA yields after DNA isolation. The primary wastewater could have served as a limitation because of the variability in the pooled samples. In addition, studies such as Gray, M.A. et (2013) conducted experiments with known bacterial composition while this study investigated natural communities (Gray, M.A. et al 2013). Further research could be conduct on the performances of natural communities versus fabricated bacterial communities.

The DNA yield depicted for the 24 hour, 7 day and 10 day timeframes showed that only the sample with DESS stored at 95ºF for the 24 hour and 7 day periods were significantly different than the control, this data suggests that in terms of DNA yield, DESS is inhibiting the sample when exposed to higher temperatures; however, the bacterial composition was not affected. After the 7 day period, all the samples were significantly different than the control in terms of DNA yield, suggesting that if sequencers that required higher DNA yields for sequencing such as the Oxford Nanopore MinION then other preservation methods would need to be utilized. In addition, this study found that the bacterial diversity and the DNA yield are both affected when the sample is kept for more than 7 days at higher temperatures with or without a preservative.

Further investigation into the chemistry and biological mechanisms behind the interaction of DESS and increased temperatures would need to be conducted to better understand the decreased DNA yield and the difference in bacterial diversity.

This study suggests that DESS is not a good solution as a preservative for water samples filtered through the Sterivex filter, in contrast to the findings of Gray et al. (2013) with a defined bacterial community. High concentrations of a few known organisms may be very different from the microbial community of wastewater; therefore, more studies are necessary to determine an effective preservation method. Another potential investigation could explore whether other liquid-based and even card-based preservations perform similarly under higher temperatures as DESS. This knowledge is crucial when using alternative preservation methods to ensure that scientists are accurately identifying the bacterial community within an ecosystem.

The methodology used in this study will provide information to further establish this preservation technique or new techniques in situations that require alternative preservatives. In the 2018 study by Roy and colleagues, the investigators had difficulties with transporting the Sterivex filter units to the processing facility. Without proper preservation by ice pack storage due to rejection by the Haitian shipping company, these samples likely experienced DNA degradation (Roy, M.A. et al 2018). If alternative preservation methods were established, then researchers would have a longer timeframe to extract the DNA before DNA degradation started.

CHAPTER 5

DISCUSSION AND CONCLUSIONS

5.1 Sequencing methods

It is important to consider potential limitations in any data generated by new technology. The MiSeq Illumina and Oxford Nanopore MinION not only differ in size, but the mechanism used to sequence samples. The Illumina uses DNA polymerase with fluorescently labeled deoxyribonucleotide triphosphates (dNTPs). A DNA strand is produced after each cycle and is identified through fluorophore excitation which enables the system to identify the nucleotides. The Illumina can process millions of fragments and is marketed as having a high read, error free yield (Illumina, n.d.); however, in Alkan et al.'s (2010) assessment, the investigators found that the Illumina produces short read lengths ranging from 75-100 base pairs with insert sizes ranging from 200-500 base pairs. Due to these short-read lengths, the assemblage of longer duplications and repeats will be hindered. Furthermore, the fundamental algorithm approaches used by Illumina are the de Bruijn graph and Eulerian path which have been demonstrated to have difficulty with assembling genomes with great complexity (Alkan, C. et al 2010).

Alternatively, the newer developed Oxford Nanopore MinION sequences DNA by emitting an ionic current that is altered when a molecule passes through the nanopore, which enables identification of the molecule through the change in the current (Oxford Nanopore Technologies, n.d). Tyler and colleagues (2018) explored the capability of the Oxford Nanopore MinION to assess the yield, quality, and accuracy of the sequencer when utilized for metagenomic and bacterial genomics studies with 1D, 2D, and 1D rapid

chemistry flow cells. The study concluded that the recent updates have produced higher sequence yields due to the software updates and simplified protocols; however, the researchers observed inconsistencies with the DNA sequence yields that ultimately skewed the data. In addition, the limited accuracy in base calling for homopolymeric sequences can result in deletion and insertion of one base or more. However, miscalling is also observed in other technologies with complex sequences. Despite these limitations the Oxford Nanopore MinION was able to accurately sequence 96% for the 2D and 94% for the 1D chemistry of all the runs (Tyler, A.D. et al 2018) and future software updates as well as advances in the chemistry could decrease the miscalls and DNA sequence yield inconsistencies.

Future research to identify the limitations of these sequencing methods will help advance the accuracy of these technologies. Different sequencing processes enable researchers to explore an array of possible applications, such as the Oxford Nanopore MinION's capability to read long strands of DNA making it ideal for experiments using a metagenomic analyses. Furthermore, a database assessment of lower concentration pathogens correlation to human population risk needs to be developed to use these DNA-based methods for health risk assessments (Ramirez-Castillo, F.Y. et al 2015).

5.2 Preservations

Alternative sample preservation methods beyond use of the "cold chain" (Gray, M.A. et al 2013) have been researched, yet an ideal solution has yet to be found. Gray et al (2013) examined the effectiveness of DNAgard, RNAlater, DESS, FTA cards, and FTA Elute cards effectiveness in preserving the DNA integrity of a known microbial

community; however, these tests were only conducted at room temperature. Despite the success of DESS as a preservative for less than 3 months in the Gray 2013 study (Gray, M.A. et al 2013), the results described in our work suggest that DESS may be less effective at higher temperatures, potentially allowing rapid DNA degradation to occur. However, these results require further analysis into the chemical and biological mechanisms involved in this inhibition process. Additionally, the Gray 2013 study suggests that the most effective preservative, considering both the cluster analysis and best-case scenario presence/absence, was DNAgard (Gray, M.A. et 2013), which could replace DESS in our future experiments so that we can evaluate whether it might be a better option for complex DNA samples.

The components of DESS allow the preservative to prevent DNA precipitation by inhibiting DNA polymerase and other potential cleavage enzymes due to the high salt concentration and divalent metal chelator EDTA. Tatangelo et al. (2013) investigated the performance of DESS and Lifegard performance on soil and water aliquots in temperatures below and above room temperature (39ºF and 86ºF) over a 15, 30, and 46 day time period. Interestingly, the researchers found that the aliquots with no preservative did not show any significant impact on bacterial structure or relative abundance of the major bacterial taxa compared to the immediate and frozen samples. These results were consistent with findings from Lauber et al. (2010) which concluded that duration of storage and change in temperatures did not significantly affect the bacterial diversity and phylogenetic structure of their samples, suggesting that environmental samples collected and stored with no preservation method could still be utilized in analyzes for microbial diversity (Lauber, C.L. et al 2010). Importantly, this experiment was conducted on soil,

human fecal matter, and human skin making it difficult to extrapolate to water. In terms of DNA yield, our study demonstrated a significant decrease in DESS preserved samples kept at 95ºF. However, with the exception of a few outliers, the PCA plot indicated that the samples were similar in total percent matches for bacterial species, even in the absence of DESS compared to the positive controls. The results from the PCA of bacterial species were similar to the findings from Lauber et al. (2010) and Tatangelo et al. (2013) that demonstrated the similarity between the microbial community regardless of preservation, temperature, and duration. However, our study explored preservation after filtration, while these other studies preserved the water aliquots before filtration.

The preferred preservation method for samples collected for bacterial community assessment is storage at -20ºC (U.S. Geological Survey. 1997); however, when transporting samples this may not be a viable option. Understanding the optimal conditions for alternative preservations methods may be critical to accurately identify the bacterial community within a sample. However, Lauber et al. (2010) suggest that the samples with no preservative under various temperature conditions for 2 weeks would show similar bacterial taxa compared to immediate samples and that molecular techniques and environmental attributes are the components that hinder the results not the lack of preservation. This was confirmed by Tatangelo et al. (2013) in aliquoted water samples; however, this does not necessarily mean that unpreserved filtered water will show similar results, this study observed significantly low DNA yield for the unpreserved samples kept at 95ºF.

5.3 Monitoring and Surveillance of Waterborne Diseases

Developing a disease is dependent on particular factors: pathogenicity, host susceptibility, minimal infectious dose, and environmental characteristics. Waterborne diseases cause the death of 2.2 million people globally a year, this statistic does not include people hospitalized for waterborne illnesses with symptoms such as diarrhea. Reducing sources of infectious waterborne diseases can be difficult especially from drinking water, where some organisms have become resistant to chlorination, heat, and inactivation by UV light, the common methods for disinfectants for drinking water (Ramirez-Castillo, F.Y. et al 2015). Additionally, roughly 2.5 billion people live in areas with poor sanitation conditions and 780 million people lack access to a treated water source. The decrease in the global disease burden of infectious waterborne diseases has been significant, yet the number of outbreaks since the 1990s have been frequent with 64% caused by parasites, 21% by bacteria, and 2.8% by viruses. Given the historical data and persistence of these waterborne diseases there is an urgent need to reduce these incidences.

Currently, no methodology (collection and analysis) exists that can account for all pathogenic microorganisms due to many obstacles: low concentration of pathogens would require enrichment, major pathogen groups physical differences, and inhibitors present among other factors. Ramirez-Castillo et al (2015) states there are 1,407 species of pathogens that are infectious to humans and this encompasses 538 species of bacteria, 208 types of viruses, 57 species of parasitic protozoa, along with helminths and fungi species (Ramirez-Castillo, F.Y. et al 2015).

Surveillance of waterborne diseases involves tracking the illness and provides important information that can prevent the spread of disease and future outbreaks. The U.S. has the Waterborne Disease and Outbreak Surveillance System (WBDOSS) that was created in 1971 and uses data about waterborne outbreaks and diseases provided by the state health departments, territories, and Freely associated states. This data includes epidemiological and environmental health investigations, water sample testing, clinical specimen testing, and the characteristics of the outbreak such as the timing, location, and number of cases reported. An outbreak is determined following the strength-of -evidence, if the epidemiological data shows a relative risk of >2 or multiple cases with identical exposures and pathogen with the same molecular characterization. Also, if the environmental data show that at least one clinical specimen and molecular characterization of pathogens are the same with historical data then it would be considered an outbreak (CDC, n.d).

Many factors can cause contamination of drinking water with subsequent outbreaks of waterborne diseases such as water treatment deficiencies (improper filtration of surface water), poor drinking water infrastructure promoting the growth of microbial communities, and weather that can cause flooding leading to runoff of pollutants into drinking water. An essential part of surveillance is monitoring, which is a routine and ongoing observation of the health (in this case) of a body of water (WHO, n.d.)

The water quality of Lake Warner is assessed using the Massachusetts Department of Environmental Protection (MassDEP) protocol that follows the Clean Water Act using the Water Quality Monitoring: Quality Management Program. This program includes the objectives and goals, data quality objectives, sampling logistics, equipment used, quality

control sampling, sampling design, data validation and management, data reporting, training, and corrective actions outlined in the EPA-approved Quality Assurance Program plan (QAPP). These monitoring programs include tiers with tier I monitoring involving the surface water quality assessment using the Section 305 (b) of the Clean Water Act, which is the calculation of the Total Maximum Daily Load (TMDL) of the pollutant. Massachusetts' fecal coliform assessment has a TMDL of 235 colonies of *E.coli* for 100 mL of water (Johnson, J. 2015, Mass.gov. n.d). Tier II is quantifying the contaminant loads from the major rivers, tier III is to identify the "hot spots" for the pollutant, tier IV develops the TMDLs for that body of water, and finally tier V is monitoring to comply with the regulatory and permit limits. Future steps are to decrease the non-point source pollution and point source pollution (MassDEP. 2015). The primary bacteria monitored is *E.coli*, in the MassDEP Quality Assurance Program Plan, no other bacteria are listed and in the State of the Lake Report for Lake Warner only *E. coli* levels are assessed (MassDEP. 2015, Johnson, J. 2015). During culture-dependent methods, false negatives could occur due to the wide range of environmental spread of pathogens that can survive in a viable but non-culturable condition. Another drawback is that culturing cannot detect, it can only provide high likelihood of fecal contamination. Furthermore, using the fecal indicator bacteria method has been under question because the absence of *E. coli* does not necessarily mean the absence of all pathogens given that research has suggested that culturing cannot be used to determine the likelihood of *Cryptosporidium* and *Giardia* in a cultured sample (Meals, D.W. et al 2013).

When designing a monitoring program for a particular body of water, typically the sampling locations are upstream, downstream and paired watersheds with a trend

monitoring for these locations. The collection type (grab sampling is the most common for pathogen source assessment), timing (historical data on fecal indicator bacteria, seasonal patterns, or dry vs. wet weather sampling will be used) and frequency are determined. During sampling, sterile collection is required for fecal indicator bacteria and depending on the microbial pathogen of interest, high volumes of water may be needed for sampling such as protozoa analysis, which would require 10 L or more of water. The timing and frequency of collection will depend on the historical data of the known microbial pathogen present in the body of water. There are pathogens that are known to occur sporadically and are associated with livestock such as *E. coli* O157:H7 and *Cryptosporidium*. If bacterial counts are extremely varied and seem to be influenced by seasonal patterns and agricultural management then high frequency water collection would be required to account for this variability and to provide the appropriate trends. Water quality standards for fecal indicators produce a geometric mean that are taken from wet vs dry weather as well as a particular number of samples over a certain timeframe (dependent on the body of water of interest). Additionally, source tracking for fecal contamination using the culturing method of Total Maximum Daily Load (TMDL) is not quantitative enough to accurately evaluate whether a particular source is causing pathogen contamination because the method uses indicator organisms instead of the actual pathogens of concern. However, culturing can be used for spatial and temporal trends that could be associated with a source of fecal contamination. The molecular based microbial source tracking can provide the direct organism needed to provide evidence to the source of the specific pathogens of interest (Meals, D.W. et al 2013).

The results from this thesis show that other clinically significant bacteria are present in Lake Warner that could pose a threat to the public's health. Sequencing using a metagenomics approach provides valuable information in monitoring these clinically significant bacteria. However, this DNA – based method (like MiSeq Illumina and Oxford Nanopore MinION) lack data to provide the infectious risk to the population when pathogens are at low levels (Ramirez-Castillo, F.Y. et al 2015). The sequenced data from Lake Warner had overall low levels of the clinically significant bacteria and provides no further information on the level of human threat these bacteria possess, unlike the culturing method that has established risk-based guidelines for indicator organisms. Culturing is the primary method for pathogen detection due to the low cost, but there are many limitations such as low sensitivity, time involved, and the chance of false negatives. Furthermore, molecular methods such as sequencing may be ideal for health risk assessment because the host – origin libraries can be utilized for pathogen source tracking. Source tracking the microbes of interest under the Clean Water Act indicates that only multiple sites over a period of time are required for assessment (Meals, D.W. et al 2013). This thesis demonstrated that completely different clinically significant bacteria were present at opposite ends of the lake. More investigation into the number of sampling sites needed to accurately assess the overall clinically significant pathogens present would provide insight into the conditions and more accurate public health risk assessments. Lake Warner experiences nonpoint source pollution and internal loading, though historically high levels of *E. coli* have been observed in the inlet of Mill River, other pathogens could be present in other un-sampled locations throughout the

Lake. There is also no known sequencing data other than this thesis and no molecular-based seasonality data exists.

In the U.S., the monitoring and surveillance programs are well-established and funded compared to developing countries. This thesis sought to optimize an affordable molecular-based technique for water quality assessments for rapid, portable surveillance of waterborne pathogens in countries like Haiti. Epidemiological and environmental studies are important to assess water quality and to track the source of an outbreak. Surveillance programs are too costly for developing countries to maintain, which can lead to future outbreaks. The cholera outbreak required an epidemiological study to determine the origin of the disease. This distinguished the epicenter of the cholera outbreak, and environmental studies confirmed the source. Depending on the pathogen of interest (whether other pathogens will be assessed besides *V. cholerae*), designing the monitoring program for Haiti would need historical molecular-based data starting at the epicenter of the outbreak. This data would indicate the frequency and time required to source track the pathogen(s). There are no protocol standards for molecular-based water quality assessment, when culturing for a particular pathogen, a known amount of water is filtered allowing quantification; however, using a sequencing approach, no standards are in place to allow for the appropriate assessment (Ramirez-Castillo, F.Y. et al 2015). A standard collection and analysis method that is inexpensive, rapid, and serves as an accurate representation of the water source of interest would need to be designed to ensure that all the pathogens are identified and monitored before another natural disaster occurs. Future development of these molecular-based standards along with a database for determining

the human risk levels of the identified pathogens, would provide a powerful surveillance tool.

5.4 Future directions

Many limitations exist when establishing a completely in-field approach to sequence samples for metagenomic analysis. The Oxford Nanopore MinION is designed for portability, with its small size and capability of sequencing with only the addition of a laptop. However, DNA isolation and library preparation create numerous challenges, such as maintaining a sterile environment, the required centrifugation steps, and ensuring reagents are kept at their optimal condition.

New technology is emerging to help address these challenges. The Bento lab created a mobile PCR, centrifuge, and gel visualization kit for DNA analysis kit that is the size of a laptop computer (Nature Technology 2016), albeit still requiring a main power supply. This technology could provide the means for scientists and water quality managers to conduct analyses out in the field and would eliminate the need for preservation techniques. A methodology that incorporated the Sterivex filter unit coupled with a portable vacuum manifold, Bento lab mobile DNA analysis kit (adapted for solar power), and the Oxford Nanopore MinION could produce quick, reliable information about the microbial diversity without ever leaving the sampling site. This powerful method could have the capacity to monitor water quality in rural, low resource settings. If simplified, it could be taught to local volunteers as a potential surveillance prevention tool, following, for example, severe weather events.

Limitations of the Oxford Nanopore MinION have been assessed, in the recent work by Tyler and colleagues (2018). These researchers evaluated the Oxford Nanopore MinION's quality, accuracy, and yield from bacterial genome and complex metagenomic sequences. The researchers found that the DNA sequence yields were inconsistent producing low and high yields as well as miscalling but noted that the recent software updates and simplified protocols had enabled the device to provide higher DNA sequence yield than the previous versions. Even with these limitations, the Oxford Nanopore MinION's ability to identify 96% of the DNA for the 2D chemistry flow cell and 94% for the 1D chemistry flow cell is impressive (Tyler, A.D. et al 2018). Technology, especially newly developed, will experience limitations; it is the responsibility of the researchers to consider these disadvantages when interpreting data. The Oxford Nanopore MinION's portability and rapid sequencing as well as real-time bacterial analysis capabilities could change the future for sequencing with many potential applications in environmental and clinical diagnostics.

The Oxford Nanopore MinION was capable of detecting lower levels of bacterial species opposed to the MiSeq Illumina; however, in terms of surveillance, the question of whether the Oxford Nanopore MinION is capable of detecting these clinically significant bacteria using a metagenomics approach is unknown. The Oxford Nanopore MinION should be evaluated using various surface water qualities and designed known bacterial samples with more complex bacterial species and levels. The Oxford Nanopore MinION was unable to assess the surface water quality of Lake Warner given the low DNA yield that was insufficient to run on the portable sequencer.

New preservation techniques could be utilized until reliable in-field sequencing is established. DESS was selected in this study because storage was assessed under 3 months and it was the most cost effective liquid preservation option. Many different variables could be added to re-evaluate DESS as a preservative for filtered water samples. DESS could be assessed using a known bacterial community in an aqueous solution filtered through the Sterivex unit to identify the DNA yield from the bacterial strains, as well establishing the temperature range that inhibits the effectiveness of DESS. Another filter unit could be used to compare whether DESS is more effective when paired with certain filters. The Sterivex filter unit was immersed in DESS and then the preservative was extracted and washed with PBS to ensure DESS was completely removed. It is possible that bacteria collected in the filter could have been washed out with the preservative, which could have contributed to loss of DNA yield. Isolating DNA from the DESS preservative and the PBS washes could have indicated whether any bacteria escaped from the filter.

Gray et al.'s (2013) study suggested that DNAgard was the most effective liquid-based preservation. A future study could investigate the liquid-based and card-based preservation performances using the Sterivex filter for the microbial community within a water source. The differences in procedures from our study and the Gray study could be responsible for the different results observed given that our study did not use a defined mixture of bacteria. The metagenomic approach as opposed to analyzing defined bacterial cultures could account for the differences observed between the two studies.

No seasonality data was collected on Lake Warner; thus lack of spring and summer collections may have contributed to the low DNA yield. If water collection had

been conducted during the spring and summer months after rain events, the DNA yield could have been high enough to sequence on the Oxford Nanopore MinION. Further investigation into the seasonal changes in bacterial diversity of Lake Warner could provide information for water quality management, especially for times of the year when the lake is most heavily used for recreational activities. We anticipate that bodies of water that are utilized for daily human activities and receive untreated human and livestock wastes, such as the Artbonite River in Haiti, will yield higher concentrations of DNA, as was observed by Roy et al (2018).

The WHO estimated that waterborne diarrheal diseases are responsible for the death of 1.5 million people every year worldwide; 58% of those deaths are due to unsafe drinking water consumption (WHO 2012). In 2017, the United States had 6,939 deaths from 13 diseases associated with water-related pathogens. Around 7% were due to oral-fecal contamination while 91% were from pathogens that grow in water system biofilms. In addition to these deaths, roughly 477,000 waterborne diseases were reported from 13 water-related pathogens with 21% requiring immediate hospitalization (CDC 2017). Even developed countries such as the United States, experience waterborne-related diseases, although the death toll is lower than in underdeveloped countries because of effective monitoring programs and other medical resources accessible in these countries. Developing a methodology for detecting waterborne pathogens requires high sensitivity, specificity, reproducibility, rapid, and low cost. This method of collection and analysis is challenging given the differences between the pathogens affecting humans. Though molecular techniques such as sequencing using a metagenomics approach could replace the less sensitive and accurate culturing method, many challenges still remain such as the

sample processing (Ramirez-Castillo, F.Y. et al 2015). Future advances in technology and research to optimize monitoring and surveillance for waterborne pathogens may be crucial to reduce the global burden of morbidity and mortality from waterborne diseases, especially for vulnerable populations.

5.5 Acknowledgements

Qubit Fluorometric and bioanalyzer performed by the UMass Amherst Genomic Sequencing Laboratory.

REFERENCES

1. Alast, Maarteen K. Van. "The Impacts of Climate Change on the Risk of Natural Disasters." Disasters, (2006). https://doi.org/10.1111/j.1467-9523.2006.00303.x.
2. Albert, M. John, A. S. G. Faruque, and D. Mahalanabis. "Association of Providencia Alcalifaciens with Diarrhea in Children." *Journal of Clinical Microbiology* 36, no. 5 (1998): 1433–35.
3. Alkan, Can, Saba Sajjadian, and Evan E Eichler. "Limitations of Next-Generation Genome Sequence Assembly." *Nature Methods* 8, no. 1 (2011): 61–65. https://doi.org/10.1038/nmeth.1527.

4. Applications, National Research Council (US) Committee on Metagenomics: Challenges and Functional. *Why Metagenomics?* National Academies Press (US), (2007). https://www.ncbi.nlm.nih.gov/books/NBK54011/.

5. "Artibonite River | River, Hispaniola | Britannica.Com." Accessed July 19, 2019. https://www.britannica.com/place/Artibonite-River.

6. Beknazarova, Meruyert, Shelby Millsteed, Gemma Robertson, Harriet Whiley, and Kirstin Ross. "Validation of DESS as a DNA Preservation Method for the Detection of Strongyloides Spp. in Canine Feces." *International Journal of Environmental Research and Public Health* 14, no. 6 (09 2017). https://doi.org/10.3390/ijerph14060624.

7. Bentley, R., and R. Meganathan. "Biosynthesis of Vitamin K (Menaquinone) in Bacteria." *Microbiology and Molecular Biology Reviews* 46, no. 3 (September 1, 1982): 241–80.
8. "Bento Lab | Nature Biotechnology." (2016). https://www.nature.com/articles/nbt0516-455.
9. Chen, K & Pachter, L"Bioinformatics for Whole-Genome Shotgun Sequencing of Microbial Communities." (2005). https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.0010024.

10. Brabb, Thea, Denise Newsome, Andrew Burich, and Martha Hanes. "Chapter 23 - Infectious Diseases." In *The Laboratory Rabbit, Guinea Pig, Hamster, and Other Rodents*, edited by Mark A. Suckow, Karla A. Stevens, and Ronald P. Wilson, 637–83. American College of Laboratory Animal Medicine. Boston: Academic Press, 2012. https://doi.org/10.1016/B978-0-12-380920-9.00023-7.

11. Bruce, Thiago, Pedro M. Meirelles, Gizele Garcia, Rodolfo Paranhos, Carlos E. Rezende, Rodrigo L. de Moura, Ronaldo-Francini Filho, et al. "Abrolhos Bank Reef Health Evaluated by Means of Water Quality, Microbial Diversity, Benthic Cover, and Fish Biomass Data." *PLOS ONE* 7, no. 6 (June 5, 2012): e36687. https://doi.org/10.1371/journal.pone.0036687.

12. "Can Bacterial Interference Prevent Infection?: Trends in Microbiology." Accessed July 19, 2019. https://www.cell.com/trends/microbiology/fulltext/S0966-842X(01)02132-1?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS0966842X01021321%3Fshowall%3Dtrue.

13. Cao, Minh Duc, Devika Ganesamoorthy, Alysha G. Elliott, Huihui Zhang, Matthew A. Cooper, and Lachlan J. M. Coin. "Streaming Algorithms for Identification of Pathogens and Antibiotic Resistance Potential from Real-Time MinION(TM) Sequencing." *GigaScience* 5, no. 1 (26 2016): 32. https://doi.org/10.1186/s13742-016-0137-2.

14. Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S.M., Betley, J., Fraser, L., Bauer, M., Gormley, N., Gilbert., J.A. "Ultra-High-Throughput Microbial Community Analysis on the Illumina HiSeq and MiSeq Platforms | The ISME Journal." Accessed July 19, 2019. https://www.nature.com/articles/ismej20128.

15. CDC. "Current Waterborne Disease Burden Data & Gaps | Healthy Water" (2018). https://www.cdc.gov/healthywater/burden/current-data.html.

16. "CDC Novel H1N1 Flu | The 2009 H1N1 Pandemic: Summary Highlights, April 2009-April 2010." (2010). https://www.cdc.gov/h1n1flu/cdcresponse.htm.

17. Charbeck, E. "Providencia Infections: Background, Pathophysiology, Epidemiology." (2019). https://emedicine.medscape.com/article/226541-overview.

18. Chen, Baowei, Ying Yang, Ximei Liang, Ke Yu, Tong Zhang, and Xiangdong Li. "Metagenomic Profiles of Antibiotic Resistance Genes (ARGs) between Human Impacted Estuary and Deep Ocean Sediments." *Environmental Science & Technology* 47, no. 22 (November 19, 2013): 12753–60. https://doi.org/10.1021/es403818e.

19. Clark, David P., and Nanette J. Pazdernik. "Chapter E9 - Genomics and Systems Biology." In *Molecular Biology (Second Edition)*, edited by David P. Clark and Nanette J.

20. "CosmosID - Home." (n.d) https://www.cosmosid.com/.

21. D'Agata, E., Douglas, M., & Bennett's Principles of Practice of Infectious Disease. (2015). P.aeruginosa and the Human Microbiota. https://www.sciencedirect.com/book/9781455748013/mandell-douglas-and-bennetts-principles-and-practice-of-infectious-diseases

22. Piarroux, Renaud, Robert Barrais, Benoît Faucher, Rachel Haus, Martine

Piarroux, Jean Gaudart, Roc Magloire, and Didier Raoult. “Understanding the Cholera Epidemic, Haiti.” *Emerging Infectious Diseases* 17, no. 7 (July 2011): 1161–68. https://doi.org/10.3201/eid1707.110059.

23. Pazdernik, e110–17. Boston: Academic Press, 2013. https://doi.org/10.1016/B978-0-12-378594-7.00043-3.

24. “Climate Data Online (CDO) - The National Climatic Data Center’s (NCDC) Climate Data Online (CDO) Provides Free Access to NCDC’s Archive of Historical Weather and Climate Data in Addition to Station History Information. | National Climatic Data Center (NCDC).” Accessed July 19, 2019. https://www.ncdc.noaa.gov/cdo-web/.

25. “Complete Ecological Isolation and Cryptic Diversity in Polynucleobacter Bacteria Not Resolved by 16S RRNA Gene Sequences | The ISME Journal.” Accessed July 19, 2019. https://www.nature.com/articles/ismej2015237

26. Dowell, S.F., & Tappero, J.W. “Public Health in Haiti — Challenges and Progress | NEJM.” Accessed July 19, 2019. https://www.nejm.org/doi/full/10.1056/NEJMp1100118.

27. Eckburg, Paul B., Elisabeth M. Bik, Charles N. Bernstein, Elizabeth Purdom, Les Dethlefsen, Michael Sargent, Steven R. Gill, Karen E. Nelson, and David A. Relman. “Diversity of the Human Intestinal Microbial Flora.” *Science (New York, N.Y.)* 308, no. 5728 (June 10, 2005): 1635–38. https://doi.org/10.1126/science.1110591.

28. Finkelstein, Richard A. “Cholera, Vibrio Cholerae O1 and O139, and Other Pathogenic Vibrios.” In *Medical Microbiology*, edited by Samuel Baron, 4th ed. Galveston (TX): University of Texas Medical Branch at Galveston, 1996. http://www.ncbi.nlm.nih.gov/books/NBK8407/.

29. Funari, Enzo, Maura Manganelli, and Luciana Sinisi. “Impact of Climate Change on Waterborne Diseases.” *Annali Dell’Istituto Superiore Di Sanita* 48, no. 4 (2012): 473–87. https://doi.org/10.4415/ANN_12_04_13.

30. “General Information | Cholera | CDC,” (2018). https://www.cdc.gov/cholera/genera/index.html.

31. Gray, Michael A., Zoe A. Pratte, and Christina A. Kellogg. “Comparison of DNA Preservation Methods for Environmental Bacterial Community Samples.” *FEMS Microbiology Ecology* 83, no. 2 (2013): 468477. https://doi.org/10.1111/1574-6941.12008.

32. Hahn, Martin W., Jitka Jezberová, Ulrike Koll, Tanja Saueressig-Beck, and Johanna Schmidt. “Complete Ecological Isolation and Cryptic Diversity in *Polynucleobacter* Bacteria Not Resolved by 16S RRNA Gene Sequences.” *The ISME*

Journal 10, no. 7 (July 2016): 1642–55. https://doi.org/10.1038/ismej.2015.237.

33. Hamner, Steve, Susan C. Broadaway, Ethan Berg, Sean Stettner, Barry H. Pyle, Nita Big Man, Joseph Old Elk, et al. "Detection and Source Tracking of Escherichia Coli, Harboring Intimin and Shiga Toxin Genes, Isolated from the Little Bighorn River, Montana." *International Journal of Environmental Health Research* 24, no. 4 (August 2014): 341–62. https://doi.org/10.1080/09603123.2013.835030.

34. Hartland EL, Leong JM. Enteropathogenic and enterohemorrhagic E. coli: ecology, pathogenesis, and evolution. Front Cell Infect Microbiol. 2013;3:15. Published 2013 Apr 30. doi:10.3389/fcimb.2013.00015.

35. Hunt, Dana E., Yajuan Lin, Matthew J. Church, David M. Karl, Susannah G. Tringe, Lisa K. Izzo, and Zackary I. Johnson. "Relationship between Abundance and Specific Activity of Bacterioplankton in Open Ocean Surface Waters." *Applied and Environmental Microbiology* 79, no. 1 (January 1, 2013): 177–84. https://doi.org/10.1128/AEM.02155-12.

36. "Hurricanes and Climate Change." Center for Climate and Energy Solutions, October 11, 2017. https://www.c2es.org/content/hurricanes-and-climate-change/.

37. "Illumina | Sequencing and Array-Based Solutions for Genetic Research." Accessed July 19, 2019. https://www.illumina.com/.

38. Johnson, J. "HADLEY — Lake Warner Is Approximately a Quarter Mile from the Connecticut River in Hadley. It Is a 68-Acre, Freshwater." (2015). https://www.gazettenet.com/Archives/2015/07/INCLOSEPROXIMITY-hg-072815.

39. Johnson, Jason L. "State of the LAKE Report 2015," n.d., 27. (2015). http://www.ecs.umass.edu/eve/facilities/MillRiver/SOL%20Final%20draft%202-16-16.pdf

40. Karkman, Antti, Thi Thuy Do, Fiona Walsh, and Marko P. J. Virta. "Antibiotic-Resistance Genes in Waste Water." *Trends in Microbiology* 26, no. 3 (March 1, 2018): 220–28. https://doi.org/10.1016/j.tim.2017.09.005.

41. Kirshstein, J.D., Anderson, C.W., Wood, J.S., Longcore, J.C., & Voytek, V.A. Quantitative PCR detection of Batrachochytrium dendrobatidis DNA from sediments and water. (2007). https://www.int-res.com/articles/dao_oa/d077p011.pdf

42. Kothari, Ramesh K., Neelam M. Nathani, Chandrashekar Mootapally, Jalpa K. Rank, Haren B. Gosai, Bharti P. Dave, and Chaitanya G. Joshi. "Chapter 11 - Comprehensive Exploration of the Rumen Microbial Ecosystem With Advancements in Metagenomics." In *Metagenomics*, edited by Muniyandi Nagarajan, 215–29. Academic Press, 2018. https://doi.org/10.1016/B978-0-08-102268-9.00011-2.

43. Kratz, J.M. “In the Time of Cholera – Foreign Policy.” (2013). https://foreignpolicy.com/2013/01/10/in-the-time-of-cholera/.

44. Lauber, Christian L., Nicholas Zhou, Jeffrey I. Gordon, Rob Knight, and Noah Fierer. “Effect of Storage Conditions on the Assessment of Bacterial Community Structure in Soil and Human-Associated Samples.” *FEMS Microbiology Letters* 307, no. 1 (June 2010): 80–86. https://doi.org/10.1111/j.1574-6968.2010.01965.x.

45. Lim, Ji Youn, Jangwon Yoon, and Carolyn J. Hovde. “A Brief Overview of Escherichia Coli O157:H7 and Its Plasmid O157.” *Journal of Microbiology and Biotechnology* 20, no. 1 (January 2010): 5–14.

46. Liu, Z., Wang, Y., Deng, T., & Chen, Q. “Solid-State Nanopore-Based DNA Sequencing Technology.” (2016). https://www.hindawi.com/journals/jnm/2016/5284786/.

47. Lu, Hengyun, Francesca Giordano, and Zemin Ning. “Oxford Nanopore MinION Sequencing and Genome Assembly.” *Genomics, Proteomics & Bioinformatics* 14, no. 5 (October 2016): 265–79. https://doi.org/10.1016/j.gpb.2016.05.004.

48. Massachusetts Department of Environmental Protection. “Quality Assurance Program: Surface Water Monitoring and Assessment.” (2019). https://www.mass.gov/files/documents/2016/08/nx/qapp-2015.pdf

49. Meals, D.W., Richards, R.P., & Dressing S.A. “Pollutant Load Estimation for Water Quality Monitoring Projects. *Natural Nonpoint Source Monitoring Program.* (2013). https://www.epa.gov/sites/production/files/2016-05/documents/tech_notes_8_dec_2013_load.pdf

50. “MinION.” (n.d). https://nanoporetech.com/products/minion.

51. Mitsuhashi, Satomi, Kirill Kryukov, So Nakagawa, Junko S. Takeuchi, Yoshiki Shiraishi, Koichiro Asano, and Tadashi Imanishi. “A Portable System for Rapid Bacterial

52. “NOAA. Climate Data Online”. (n.d) https://www7.ncdc.noaa.gov/CDO/cdoselect.cmd?datasetabbv=GSOD&countryabbv&georegionabbv

53. Composition Analysis Using a Nanopore-Based Sequencer and Laptop Computer.” *Scientific Reports* 7, no. 1 (18 2017): 5657. https://doi.org/10.1038/s41598-017-05772-5.

54. “Abrolhos Bank Reef Health Evaluated by Means of Water Quality, Microbial Diversity, Benthic Cover, and Fish Biomass Data.” Accessed July 19, 2019. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0036687.

55. O'Connor, M.R. "Public Health in Haiti — Challenges and Progress | NEJM." Accessed July 19, 2019. https://www.nejm.org/doi/full/10.1056/NEJMp1100118.

56. Pandya, Shirali, Kavitha Ravi, Vijaya Srinivas, Smitha Jadhav, Anisa Khan, Anjali Arun, Lee W Riley, and Purnima Madhivanan. "Comparison of Culture-Dependent and Culture-Independent Molecular Methods for Characterization of Vaginal Microflora." *Journal of Medical Microbiology* 66, no. 2 (2017): 149–53. https://doi.org/10.1099/jmm.0.000407.

57. Park, Haewon, Bongkeun Song, and François M. M. Morel. "Diversity of the Cadmium-Containing Carbonic Anhydrase in Marine Diatoms and Natural Waters." *Environmental Microbiology* 9, no. 2 (2007): 403–13. https://doi.org/10.1111/j.1462-2920.2006.01151.x.

58. Pierce, N. F. "DIFFERENTIAL INHIBITORY EFFECTS OF CHOLERA TOXOIDS AND GANGLIOSIDE ON THE ENTEROTOXINS OF VIBRIO CHOLERAE AND ESCHERICHIA COLI." *Journal of Experimental Medicine* 137, no. 4 (April 1, 1973): 1009–23. https://doi.org/10.1084/jem.137.4.1009.

59. Ramírez-Castillo, Flor Yazmín, Abraham Loera-Muro, Mario Jacques, Philippe Garneau, Francisco Javier Avelar-González, Josée Harel, and Alma Lilián Guerrero-Barrera. "Waterborne Pathogens: Detection Methods and Challenges." *Pathogens* 4, no. 2 (2015): 307–34. https://doi.org/10.3390/pathogens4020307.

60. Ranjan, Ravi, Asha Rani, Ahmed Metwally, Halvor S. McGee, and David L. Perkins. "Analysis of the Microbiome: Advantages of Whole Genome Shotgun versus 16S Amplicon Sequencing." *Biochemical and Biophysical Research Communications* 469, no. 4 (January 22, 2016): 967–77. https://doi.org/10.1016/j.bbrc.2015.12.083.

61. Reid, G., Howard, J., Gan, B.S. "Can Bacterial Interference Prevent Infection?: Trends in Microbiology". (2001) https://www.cell.com/trends/microbiology/fulltext/S0966-842X(01)02132-1?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS0966842X01021321%3Fshowall%3Dtrue.

62. Rio, Carlos del, and Jeannette Guarner. "The 2009 Influenza A (H1N1) Pandemic: What Have We Learned in the Past 6 Months." *Transactions of the American Clinical and Climatological Association* 121 (2010): 128–40.

63. Roberts, M. "Haiti Cholera 'Far Worse than Expected', Experts Fear - BBC News." (2011). https://www.bbc.com/news/health-12744929.

64. Roy, Monika A., Jean M. Arnaud, Paul M. Jasmin, Steve Hamner, Nur A. Hasan, Rita R. Colwell, and Timothy E. Ford. "A Metagenomic Approach to Evaluating Surface Water Quality in Haiti." *International Journal of Environmental Research and*

Public Health 15, no. 10 (10 2018). https://doi.org/10.3390/ijerph15102211

65. Sandle, Tim. "8 - Specified and Objectionable Microorganisms." In *Pharmaceutical Microbiology*, edited by Tim Sandle, 93–101. Oxford: Woodhead Publishing, 2016. https://doi.org/10.1016/B978-0-08-100022-9.00008-6.

66. "Shotgun Metagenomic Sequencing." (n.d). https://www.illumina.com/areas-of-interest/microbiology/microbial-sequencing-methods/shotgun-metagenomic-sequencing.html.

67. Singleton, P. "Bacteria in Biology, Biotechnology and Medicine, 6th Edition." Wiley.com. (1999). https://www.wiley.com/en-ai/Bacteria+in+Biology%2C+Biotechnology+and+Medicine%2C+6th+Edition-p-9780470090275.

68. Song, Se Jin, Amnon Amir, Jessica L. Metcalf, Katherine R. Amato, Zhenjiang Zech Xu, Greg Humphrey, and Rob Knight. "Preservation Methods Differ in Fecal Microbiome Stability, Affecting Suitability for Field Studies." *MSystems* 1, no. 3 (May 3, 2016). https://doi.org/10.1128/mSystems.00021-16.

69. "Sterivex™ Filter Units - Sterile Filtration." (n.d). http://www.emdmillipore.com/US/en/product/Sterivex-Filter-Units,MM_NF-C3235.

70. Stoddart, David, Andrew J. Heron, Ellina Mikhailova, Giovanni Maglia, and Hagan Bayley. "Single-Nucleotide Discrimination in Immobilized DNA Oligonucleotides with a Biological Nanopore." *Proceedings of the National Academy of Sciences of the United States of America* 106, no. 19 (May 12, 2009): 7702–7. https://doi.org/10.1073/pnas.0901054106.

71. Tatangelo, Valeria, Andrea Franzetti, Isabella Gandolfi, Giuseppina Bestetti, and Roberto Ambrosini. "Effect of Preservation Method on the Assessment of Bacterial Community Structure in Soil and Water Samples." *FEMS Microbiology Letters* 356, no. 1 (July 1, 2014): 32–38. https://doi.org/10.1111/1574-6968.12475.

72. Thorp, James H., and D. Christopher Rogers. *Thorp and Covich's Freshwater Invertebrates: Ecology and General Biology*. Elsevier, 2014.

73. Thiagarajah, Jay R., and A. S. Verkman. "New Drug Targets for Cholera Therapy." *Trends in Pharmacological Sciences* 26, no. 4 (April 2005): 172–75. https://doi.org/10.1016/j.tips.2005.02.003.

74. Todar, K. "Pathogenic E. Coli." (2012) http://www.textbookofbacteriology.net/e.coli.html.

75. "Total Maximum Daily Loads (TMDLs)." Mass.gov. (n.d). https://www.mass.gov/total-maximum-daily-loads-tmdls.

76. Trevelyan, L. “Cholera Reported in Haiti Capital.” BBC News. Accessed July 19, 2019. https://www.bbc.com/news/av/world-latin-america-11614883/haiti-cholera-cases-reported-in-port-au-prince.

77. “Tropical Cyclone Climatology.” (n.d). https://www.nhc.noaa.gov/climo/.

78. Tyler, A.D., Mataseje, L., Urfano, C.J., Schmidt, L., Antonation, K.S., Mulvey, M.R., & Corbett, C. R. “Evaluation of Oxford Nanopore’s MinION Sequencing Device for Microbial Whole Genome Sequencing Applications | Scientific Reports.” (2018). https://www.nature.com/articles/s41598-018-29334-5#Sec14.

79. US EPA, OA. “Summary of the Clean Water Act.” Overviews and Factsheets. US EPA, (2013). https://www.epa.gov/laws-regulations/summary-clean-water-act.

80. “USGS Field Manual for Collection of Water-Quality Data.” Accessed July 19, 2019. https://water.usgs.gov/owq/FieldManual/.

81. Vogt, Richard L., and Laura Dippold. “Escherichia Coli O157:H7 Outbreak Associated with Consumption of Ground Beef, June-July 2002.” *Public Health Reports* 120, no. 2 (2005): 174–78.

82. Wang, Xin, I. King Jordan, and Leonard W. Mayer. “Chapter 29 - A Phylogenetic Perspective on Molecular Epidemiology.” In *Molecular Medical Microbiology (Second Edition)*, edited by Yi-Wei Tang, Max Sussman, Dongyou Liu, Ian Poxton, and Joseph Schwartzman, 517–36. Boston: Academic Press, 2015. https://doi.org/10.1016/B978-0-12-397169-2.00029-9.

83. “Waterborne Diseases.” National Institute of Environmental Health Sciences. Accessed July 19, 2019. https://www.niehs.nih.gov/research/programs/geh/climatechange/health_impacts/waterborne_diseases/index.cfm.

84. “WHO | Diseases and Risks.” WHO. (n.d). http://www.who.int/water_sanitation_health/diseases-risks/en/.

85. “WHO | Water Quality Monitoring.” WHO. (n.d). http://www.who.int/water_sanitation_health/publications/wqmonitor/en/.

86. Yang, Ying, Bing Li, Feng Ju, and Tong Zhang. “Exploring Variation of Antibiotic Resistance Genes in Activated Sludge over a Four-Year Period through a Metagenomic Approach.” *Environmental Science & Technology* 47, no. 18 (2013): 10197–205. https://doi.org/10.1021/es4017365.

87. Zhou, Y., Wylie, KM., El Feghaly, R.E., Mihindukulasuriya, K.A., Elward, A., Haslam, D.B., Storch, G.A., & Weinstock., G.M. “Metagenomic Approach for Identification of the Pathogens Associated with Diarrhea in Stool Specimens | Journal of Clinical Microbiology.” Accessed July 19, 2019. https://jcm.asm.org/content/54/2/368.