8-10-2018

Next Generation Sequencing Methods for Coastal Zone Water Quality Monitoring

Catherine Iskrenko
Nova Southeastern University, cg1318@mynsu.nova.edu

Capstone of
Catherine Iskrenko

Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Science

M.S. Coastal Zone Management

Nova Southeastern University
Halmos College of Natural Sciences and Oceanography

August 2018

Approved:
Capstone Committee

Major Professor: Donald McCorquodale

Committee Member: George Duncan

HALMOS COLLEGE OF NATURAL SCIENCES AND OCEANOGRAPHY

NEXT GENERATION SEQUENCING METHODS FOR COASTAL ZONE WATER QUALITY MONITORING

By

Catherine Emily Iskrenko

Submitted to the Faculty of

Halmos College of Natural Sciences and Oceanography

in partial fulfillment of the requirements for

the degree of Master of Science with a specialty in:

Coastal Zone Management

Nova Southeastern University

10 August 2018

A Capstone Paper

Submitted in Partial Fulfillment of the Requirements for the Degree of

Masters of Science:
Coastal Zone Management

CATHERINE EMILY ISKRENKO

Nova Southeastern University

Halmos College of Natural Sciences and Oceanography

August 2018

Capstone Committee Approval

_____

Dr. Donald McCorquodale, Major Professor

_____

Dr. George Duncan, Committee Member

Acknowledgements

I would like to thank, first of all, my capstone advising committee for their constant oversight and guidance as I completed this capstone project. Dr. Donald McCorquodale and Dr. George Duncan helped me devise a project tying coastal zone management to human health applications, connecting the two exciting and dynamic fields. They offered a constant source of inspiration and always came prepared to conversations with fascinating insights on a topic of which I had little to no background. I will forever remain appreciative for their guidance and encouragement.

I would also like to thank representatives from ThermoFisher, Illumina, and Qiagen for the wealth of information they provided about their products. Finally, I would like to thank my husband, Yuriy, for the enjoyable deadly pathogen conversations that arose in our household as a result of this project, and my 2-year-old daughter Lily who graciously provided me with 30-minute "nap" windows in which to work and who provided the most interesting typed contributions which later (sadly) had to be edited out.

Table of Contents

List of Tables

List of Figures

Abstract

When analyzing the water quality of the coastal zone, culture-based techniques have been utilized most often to identify Fecal Indicator Bacteria in samples. Since the advent of the Sanger Method for DNA sequencing, other techniques have arisen that provide significantly more information on the microorganisms in sample, but they are still not the mainstream for water quality analysis. This capstone reviews and compares culture-based techniques, DNA sequencing, RNA sequencing, qPCR for biomarker, and 16S rDNA sequencing to highlight their merits and shortcomings for analyzing environmental water samples. The technique presented that provides the broadest range of information (including the identification of bacteria, viruses, fungi, pathogens, virulence factors, and antibiotic resistance genes) is whole genome shotgun sequencing paired with k-mer based microbial identification. This technique allows researchers and managers not only to identify all microorganisms present in a given sample, but to identify sources of these microorganisms and infection potential to humans as well. This has huge implications for the future of water quality management and provides invaluable information that recreational water managers can use to determine risk to human health. As modern methods drop in price, they are becoming more accessible to user groups. This capstone is designed to help users determine the best method for their individual needs.

Keywords: whole genome sequencing, shotgun sequencing, water quality, RNA sequencing, qPCR, culture-based techniques, 16S sequencing, virulence factors, antibiotic resistance gene, CosmosID, coastal zone management

Introduction

The microbiome of coastal sands and waters has historically been difficult to assess due to technological limitations. Scientists and government officials have long had the ability to recognize fecal indicator bacteria such as *E. coli* in water samples using culture-based methods as an indicator of the risk to human health. The makeup of just what organisms (bacteria, viruses, fungi) are present in coastal waters and beach sands requires significantly more refined procedures and can tell officials much more about the overall risk to human and environmental health in a given area. With the advent of the Sanger Method, it became possible to sequence DNA effectively and the massive undertaking of uncovering the human genome and creating DNA libraries took off. While this is considered by many the "gold standard" method, it has its shortcomings. It is costly, time consuming, and labor-intensive. With next-generation sequencing methods, thousands to millions of DNA molecules can be sequenced in a single run, and the number of applications continues to grow (Vincent et al. 2015).

With the advent of these next generation sequencing methods, the task of understanding the microbiome of the coastal zone has made its way to the forefront with the objective to identify new microorganisms, identify distribution patterns, and ultimately create predictive models to aid in identifying areas of concern.

While there are a few next generation sequencing methods, "whole-genome shotgun sequencing" (WGS) carries a lot of promise as a technique to identify large quantities of microbes quickly. If more widely accessible, it could prove a valuable asset to researchers and government officials. WGS is known as a "high-throughput" DNA sequencing method (Falkowski and Vargas 2004). It sequences massive data sets of DNA fragments in parallel, making it faster, more efficient, and more all-encompassing than traditional methods. Other techniques, such as 16S rDNA sequencing, also show improvement on more traditional methods, and have been employed successfully in coastal studies, with applications expanding (Ranjan et al. 2015). Five methods in particular will be detailed, reviewed, and compared.

DNA Sequencing: Whole genome shotgun sequencing of metagenomic DNA and bio-informatic analysis. Generates data on Bacteria, Protozoa, Fungi, Viruses, virulence genes, and antibiotic resistant genes.
RNA Sequencing: Whole genome shotgun sequencing of metagenomic cDNA. Includes RNA extraction and cDNA generation and bio-informatics analysis (for viable cells only).
qPCR for Biomarker: Screening for host-specific bio-marker genes by qPCR.
16S rDNA Sequencing: Single gene sequencing of 16S rDNA of metagenomic DNA in a given sample. Generates data for taxonomic identification of bacteria in sample including indicator bacteria (Molecular Testing).
Traditional culture-based methods: Bacteria are cultured on plates and colonies identified.

Purpose of Applying NGS Methods to the Coastal Zone

As can be seen from genomic sequencing of marine waters and beach sands, there is an immense microbial world that has yet to be discovered (Ranjan et al. 2015; Poretsky et al. 2014), the magnitude of which researchers can only speculate. These methods can be used to uncover the unknown diversity that is characteristic of the marine environment. As the methods' applications expand and the technology becomes more widely accessible, libraries will grow and more users will benefit.

In the coastal zone management context, those who monitor and manage water quality, as well as those who manage beaches will benefit greatly. Widespread use of whole genome shotgun sequencing and other genomic sequencing testing methods will result in the creation of the first ever comprehensive catalogue of the beach and coastal microbiome, forming the basis for future work on beach and coastal water quality management. With a better understanding of the microbiome, managers can make more informed decisions on management practices.

With the additional information on the microbiome that genomic sequencing testing provides, detailed analyses of beach sand, the swash zone, and coastal waters can aid managers in developing the best possible approach to preserving human health and

the coastal environment. The immense amount of data provided in the coastal zone will provide valuable tools for a rigorous review of current regulatory policies for beach safety. This review will allow managers to develop the most appropriate, comprehensive monitoring program encompassing areas posing significant risk. With recent data sets supporting the existence of human pathogens in beach sand (Cui et al. 2013; Solo-Gabriele et al. 2015), it is likely that current protocols would then be expanded to include sediments and the swash zone in many areas.

Another key benefit of the expanded use of genomic sequencing testing to monitor coastal zones is that it will lead to the adoption of a new risk assessment protocol and improve efficiency in monitoring. Instead of time-consuming culture-based methods that do not work for the detection of many microbes or traditional sequencing methods that take much more time to produce results, WGS methods can quickly identify the microbes present in an area. This will save costs and labor while providing detailed assessments of health risks to beach users. With improved risk assessment protocols using efficient and effective WGS methods, tourists will be more confident than ever in the management of their coastal recreation areas. Tourism will flourish as trips to the beach will not be slowed by health concerns and lack of faith in management practices.

Overview of Microbiome Analysis: Metagenomics, Metatranscriptomics, Metabolomics

Aguiar-Pulido et al. (2016) provides an exceptional overview of the types of analyses for uncovering information on microbiomes. Microbiome analysis is a type of analysis carried out across multiple sectors and fields, and the tools that can be used to conduct this type of analysis are ever expanding. These tools fit into three broad categories: metagenomics, metatranscriptomics, and metabolomics. Each of these approaches provides its own unique insights into the microbiome of a given sample, and when combined with the other two, provides a much more comprehensive picture that just one method alone. The first two approaches are the focus of this capstone.

The definition of microbiome has expanded from the more traditional definition (the microbial community within a reasonably well-defined habitat) (Whipps et al. 1988) to including information on environmental and host factors (Aguiar-Pulido et al. 2016)

There are a couple of environmental microbiome discovery initiatives of note. The Earth Microbiome Project (EMP) started in 2010 and aims to characterize microbiomes across the planet. At present, they have more than 30,000 samples from diverse ecosystems, including humans, animals, oceans, sediment, air, etc. (Gilbert et al. 2014). The J. Craig Venter Institute's (JCVI) Global Oceanic Sampling (GOS) expeditions and the European Tara Oceans initiatives (Venter et al. 2004; Nealson and Venter 2007; Lima-Mendez et al. 2015; Karsenti et al. 2011; Sunagawa et al. 2015) have focused on marine microbiome diversity across the globe.

Metagenomic studies provide researchers with the composition of the microbial community. These studies involve next-generation sequencing (NGS) after DNA has been extracted from samples, producing large quantities of data in short reads which can be pieced together to give longer reads. There are recent recommendations that metagenomics (broad and comprehensive genomic approach) be separated from metataxonomics (use of amplicons from a targeted marker gene to infer taxonomic classifications—for example 16S rDNA markers) (Barberan et al. 2011; Chaffron et al. 2010; Gonzalez and Knight 2012; Freilich et al. 2010; Kuczynski et al. 2010; Faust et al. 2012). There are a large number of databases to aid in taxonomic classification when the 16S region is targeted. Studies have, however, shifted toward a shotgun approach resulting in several databases with complete reference genomes that can be used to construct taxonomic profiles and for inferring potential functional profiles for the microbial community based on genes that are present (Sharpton, T.J. 2104; Nelson et al. 2010; Frias-Lopez et al. 2008; Chain et al., 2009).

After the pre-processing stage where reads are filtered by quality and length, contaminants are removed, chimeric sequences generated during PCR amplification are identified, and data is prepared for further analysis, classification of each read based on the taxa with the highest probability of containing the read can occur. With metataxonomics, reads are clustered prior to assigning a label. These clusters are inferred to have a common origin and are called operational taxonomic units (OTUs).

Metataxonomics helps to compute the taxonomic profile of the community, while metagenomics computes the functional profile by focusing on gene content and using the available functional annotations of the corresponding proteins.

There are shortcomings to these methods. These methods (metagenomics and metataxonomics) underestimate the number of microbial species in samples, in large part because reference databases are so limited. Reads are often discarded from undocumented microbes or grouped with the most similar database microbe. Metagenomics can also not reveal dynamic properties such as spatiotemporal activity of a given community and the environmental impact on these activities. WGS produces lower coverage and may identify thousands of strains per sample, however, targeted approaches have reads that come from small regions of the genome. The additional clustering step lowers errors in classification.

The classification and labeling steps are either taxonomy-dependent or independent. Dependent methods use a database of genomes adding some bias toward data with pathogenic or commercial applications. Independent methods do not require a-priori knowledge and segregate reads based on distance, k-mers, abundance levels, and frequencies. If there is a good likelihood that there will be non-documented microbes in the samples, this is often a good approach. Accurate classification and labeling are a challenge because of several factors. Sequencing technologies produce short reads and economic pressure leads researchers to obtain low coverage datasets. To add to difficulties, some technologies have higher error rates, while reference genome databases are not as comprehensive as sometimes desirable leading to inaccurate taxonomic context due to lateral gene transfers between microbial taxa.

Metatranscriptomics focuses on what genes are expressed by the entire microbial community, shedding light on the active functional profile of a community (Moran 2009). The metatranscriptome captures the total mRNA and gives a snapshot of gene expression in a sample at a given moment under specific conditions. It is now possible to conduct whole metatranscriptomics shotgun sequencing, providing the expression and functional profile of the microbiome (Frias-Lopez 2008; Carvalhais et al. 2012; Gilbert et al. 2008). From there, reads are either mapped to a reference genome, or a de novo assembly of reads into contigs and supercontigs is performed. These two strategies share the same

shortcomings as the metagenomic studies—constrained to database data and limited by software for assembly. Metatranscriptomics is still a relatively uncommon method for gaining information on microbiomes. While promising, these methods have their shortcomings and limitations to be addressed before being applied on the large scale. To start, much of the harvested RNA comes from rRNA, and its abundance reduces the coverage of mRNA (the main focus of these studies). It is helpful to remove as much rRNA as possible (Peano et al. 2013). mRNA is also quite unstable, which means that upon sampling, the integrity of the sample decreases. It is challenging to differentiate between host and microbial RNA, so it is helpful to have a reference genome for any hosts.

WGS approaches provide information on the taxonomic profile of a microbial community and its functional profile, whereas whole metatranscriptome sequencing describes the active functional profile. This method is most useful when studying the dynamics of functional profiles under varying conditions.

Metabolomics is the comprehensive analysis whereby all metabolites (small molecules released by the organism into the environment) are identified and quantified (Fiehn et al. 2002). The metabolome is an incredibly useful indicator of environmental health or of a deviation from homeostasis (Bernini et al. 2009). When a variation in the production of a signature metabolite is noted, this indicates a change in activity of metabolic routes, offering a means of pathway analysis (Krumsiek et al. 2015). This method provides information about more than just the characteristics of the microbiome. It gives information on the interactions of the community with the host environment (Xu et al. 2007; Manor et al. 2014; Wu et al. 2014). The study of the metabolome can help in the development of predictive markers for environmental stressors (Lankadurai et al. 2013). The microbiome reacts to environmental stressors in such a way that can be either helpful or harmful to the environment. Bioremediation response to pollution is an example of this (Kimes et al. 2014). The study of metabolomics can help identify these responses.

The metabolome can also illustrate signaling processes between bacteria (like quorum sensing) which can relate gene expression to cell population density changes (Bassler et al. 1997; Miller & Bassler 2001; Bassler 2002; Henke & Bassler 2004; Waters

& Bassler 2005; Camilli & Bassler 2006). This could revolutionize infectious disease control and help with environmental conservation.

This type of analysis describes not just the systems themselves, but their internal and external interactions. Sequencing is not used. Identification and quantification of metabolites is carried out using chromatographic techniques (such as liquid and gas chromatography) and detection methods (such as mass spectrometry and nuclear magnetic resonance). From these techniques, spectra are produced consisting of peaks that allow for the quantification and identification of metabolites. These patterns are stored in spectral databases. This is still considered to be a high-throughput analytical method.

A key challenge of this method is the difficulty in determining whether a metabolite comes from the microbiome or the host. Also, this data must necessarily be combined with other omics data to make conclusions about which genes, enzymes, or pathways are associated with a given metabolite.

For optimal results, it is helpful to combine techniques. The best way to do this is to perform separate omics analyses, and to integrate the data in a downstream analysis. This downstream analysis, as integration is mastered, will allow researchers to build and test more robust models of microbial activity and interactions with their environment (Reigstad & Kashyap 2013; Aw & Fukuda 2015). To provide an example, metagenomics and metatranscriptomics can combine to reveal over-or-under expression of functions and the activities of organisms (Duran-Pinedo et al. 2014; Mason et al. 2012; McNulty et al. 2011; Maurice et al. 2013). Metabolomics can be added to show the outcome of these changes in gene expression (Verberkmoes et al. 2009; Weir et al. 2013; Wang et al. 2011; Koeth et al. 2013; Kaddurah-Daouk et al. 2011). With more analyses in agreement, a more confident conclusion can be reached (Aguiar-Pulido et al. 2016).

Overview of 5 Methods

DNA Sequencing: Whole genome shotgun sequencing of metagenomic DNA and bio-informatic analysis. Generates data on Bacteria, Protozoa, Fungi, Viruses, virulence genes, and antibiotic resistant genes.

Craig Venter's team was the first to apply whole-genome shotgun sequencing methods in an oceanographic context in 2004, where he sought to identify the microbiome of the Sargasso Sea (Venter et al. 2004). From this early study, his team identified a staggering 1800 genomic species, to include 148 previously unknown bacterial phylotypes, 1.2 million previously unknown genes, and 782 new rhodopsin-like photoreceptors. The WGS approach had to date been used primarily for identifying the genome sequences from one organism, so this was a huge success for environmental exploration. It was not, however, without its shortcomings. Regardless of method, the true number of distinct species would be higher than that determine by finite sequence sampling, particularly low abundance species (Venter et al. 2004).

A more recent study from 2017 sequenced microbial communities in Great Lakes beach sand, considering the possibility that recreational waters were impacted by the adjacent beach sands that harbor large quantities of viruses and bacteria. Other studies have shown that bacterial concentrations in beach sands are frequently 10-100 times higher than the neighboring water bodies (Alm et at. 2003; Cui et al. 2013). Fecal indicator bacteria and other pathogenic species are also harbored in beach sand (Yamahara et al. 2012), which can pose public health concerns. Fecal indicator bacteria (FIB) (*Escherichia coli*, fecal coliforms, and enterococci) and pathogenic bacterial are introduced to beach environments through point-source (wastewater) and non-point source pollution, directly from birds, humans, and other animals, and transfer from water to sand (Whitman et al. 2015). It is important to accurately characterize these populations to determine their potential impacts to the environment and to human health. Samples were collected, and following library preparation, each library was quantified using qPCR and sequenced using a shotgun metagenomic sequencing approach on the HiSeq 2000 Illumina platform. Analysis found that there was greater taxonomic richness in sand vice water. The Shannon index showed significantly greater taxonomic diversity at the species level and the phylum level in sand vice water. A total of 34 pathogenic and indicator bacterial species were detected in sand and water. The most abundant was *E. coli*, whose abundance did not differ greatly between sand and water. *Pseudomonas mendocina* and *Pseudomonas aeruginosa* were abundant in both environments but were elevated in sand. There were other low abundance pathogens such as those from the

Clostridium genus that occurred most commonly in water, along with those from the *Vibrio* spp. The findings of this study support the creation of monitoring programs for pathogens on recreational beaches. The shotgun metagenomic approach could be used to augment traditional methods. However, standardization procedures for methods must be developed prior to adopting these new monitoring strategies (Mohiuddin et al. 2017).

There are some potential weaknesses of whole genome shotgun sequencing for sand studies, a primary one being that the methods are limited by sample preparation and collection. Bias could be introduced when capturing the microbial communities of one environment that is dependent on the other (sand and water). This could lead to underestimation of the differences between environments due to their interactions. Also, the inability of water flow to mobilize biofilms that are strongly attached to sand could lead to an underestimation of the diversity in sand (Mohiuddin et al. 2017).

In another study focusing on ssDNA and dsDNA sequencing methods, it was found that bias can be introduced if certain steps are not optimally performed using the most appropriate methods at each stage in the process. The amplification of viral DNA using different amplification methods can produce markedly different results in taxonomic classifications, functional assignments, and assembly patterns for the same samples depending on which method was used. 454 pyrosequencing was used to read the metagenomic sequences prepared from the linker amplified shotgun library (LASL) and the multiple displacement amplification (MDA) methods. Only dsDNA was identified from LASL, and primarily ssDNA was identified from the MDA library. Depending on the aims of a given study, the amplification method must be carefully chosen (Kim et al. 2011).

RNA Sequencing: Whole genome shotgun sequencing of metagenomic cDNA. Includes RNA extraction and cDNA generation and bio-informatics analysis (for viable cells only).

One application of RNA Sequencing is the discovery and identification of RNA viruses. As recently as 2006, the vast diversity of coastal RNA communities was studied using reverse-transcribed whole-genome shotgun sequencing. Culley et al. (2006) presented one of the earliest studies to take this approach, and the outcome was the

determination that much remains to be discovered. At that time, RNA viruses in the ocean were “essentially unknown”. RNA was reverse transcribed into cDNA and used to construct libraries that represented natural RNA viral communities (Culley et al. 2006). Reverse-transcribed WGS was then used to characterize the diversity of uncultivated marine RNA assemblages. Cultivation independent methods do not require prior assumptions of the composition of the target communities and produce data that can be used to estimate community structure. In the end, their analysis showed that a diverse group of RNA viruses existed in coastal areas that were related to viruses known to infect marine protists. RNA sequencing through WGS was a good technique for this study, as the genomes of RNA viruses are relatively small, making the approach realistic and effective (Culley et al. 2006).

Other studies have arisen since then that suggest reverse transcription may not be necessary for RNA sequencing studies. Ozsolak et al. (2009) presented one such study on “Direct RNA Sequencing”. Previously, work with transcriptomes was limited because of the limited knowledge of the dynamic state of transcription. Most of the work in this area involved indirect methods, as RNA had to be converted to complementary DNA (cDNA) before measurements could be made even though the cDNA synthesis step introduces many biases and artifacts that interfere with the proper characterization and quantification of transcripts. cDNA analysis is not suitable for analysis of short, degraded, or small quantity RNA samples, for instance. Ozsolak et al. (2009) focused on direct single molecule RNA sequencing without prior conversion of RNA to cDNA. Sequencing-by-synthesis reaction was performed using a modified polymerase and proprietary fluorescent nucleotide analogues that allowed step-wise sequencing. Their direct RNA Sequencing (DRS) methods introduce a simplicity as previously-required steps are not required. In these methods, only femtomole quantities of RNA are required and the biases associated with cDNA synthesis, end repair, ligation, and amplification are eliminated. This makes these methods promising for applications that require minute RNA quantities and/or short RNA species that are challenging for cDNA-based methods (Ozsolak et al. 2009). There are a good number of recent studies applying microarray and sequencing technologies to transcriptomics, and knowledge of transcription now shows that a large fraction of transcripts originates from unannotated regions of genomes (Denoeud et al.

2008; Kapranov et al. 2007; Marioni et al. 2008; Mortazavi et al. 2008; Nagalakshmi et al. 2008; Sultan et al. 2008; Wilhelm et al. 2008; Ozsolak et al. 2009). With this detail and the addition of conversion biases, important regions of the genome may go overlooked if direct RNA sequencing techniques are not utilized.

Other weaknesses associated with cDNA-based transcriptome analysis include a "tendency of various reverse transcriptases to generate spurious second-strand DNA due to their DNA dependent polymerase activities", "generation of artefactual cDNAs due to template switching", and "the error prone and inefficient nature of RTs yielding low quantities of cDNA" (Ozsolak et al. 2009).

qPCR for Biomarker: Screening for host-specific bio-marker genes by qPCR.

qPCR can be used for human health applications, such as identifying and quantifying Aeromonads in water and sand. These are gram-negative, non-spore forming, rod shaped waterborne bacteria that are found in soil, freshwater, brackish water, sewage, wastewater, and drinking water (Altwegg 1996; Brandi et al. 1996; Janda & Abbot 1998). Several species within this genus are associated with human infections, including acute gastroenteritis, septicaemia, wound infections, endocarditis, meningitis, and respiratory infections (Hanninin & Siitonen 1995; Janda & Abbot 1998; Isonhood & Drake 2002).

In 2009, Khan et al. applied a "novel, rapid, direct DNA-based protocol to enumerate aeromonads in recreational water" using an *Aeromonas*-specific biomarker, resulting in a genus-specific real-time qPCR protocol. Culture-based methods have commonly been used for identifying *Aeromonas* cells, however this qPCR protocol quantified the desired cells in less than 3 hours without culturing. Quantitative culture-independent DNA-based methods (real-time qPCR) provide a measure for quantification and monitoring of pathogens and microbial water quality indicators and use short fragments that can be amplified efficiently. With the protocols used by Khan et al. (2009), the qPCR protocol detected 16% more cells than culture-based techniques. One theory to this is that qPCR picks up on viable but non-culturable *Aeromonas* spp (Khan et al. 2009).

Many studies have compared traditional methods for identifying fecal indicator bacteria with qPCR techniques (molecular analysis). Sinigalliano et al. (2009) took

samples at intervals from bathers at a beach. Various techniques were employed to analyze the samples. Membrane filtration method was used to analyze *S. aureus* and enterococci. Enterococci were also analyzed using chromogenic substrate (CS) and qPCR methods. Bacteroidales markers were analyzed (2 human and 1 dog marker) by qPCR. 1 seagull marker was analyzed by qPCR as well. These methods were used for epidemiological purposes, showing that bathers had a significantly higher risk of reporting GI, respiratory, and skin illnesses when exposed to non-point source subtropical recreational marine waters that those who did not bathe (Sinigalliano et al. 2009).

Though not in EPA regulations for water quality monitoring, EPA has considered what it calls "rapid microbiological methods for ambient water" and has revised methods 1609 and 1611 to incorporate these methods for the detection of enterococci in water. The use of genetic qPCR is the primary method included, and modifications have been made to protocols to aid in greater standardization for water quality monitoring purposes. Despite this, EPA "Clean Water Analytical Methods" still rely heavily on culture-based methods ("Other Clean Water Act Test Methods").

While a promising method in many instances, qPCR still has its weaknesses. Khan et al. (2009) stated that real-time qPCR can overestimate the number of viable cells due to detection of dead cells. To counter this, the live and dead cells in a sample can be quantified before applying qPCR assays. A study by Nocker et al (2006) suggested the use of propidium monoazide (PMA) combined with qPCR to quantify viable and dead cells over a wide range of bacterial pathogens.

16S rDNA Sequencing: Single gene sequencing of 16S rDNA of metagenomic DNA in a given sample. Generates data for taxonomic identification of bacteria in sample including indicator bacteria (Molecular Testing).

16S rDNA sequencing is an effective means to determine spatial patterns of bacterial community composition in the coastal zone and oceanographic contexts (Bouzat et al. 2013). This method was applied in the Great Lakes region on Lake Erie where 435 environmental clones were sequenced from 11 sediment samples throughout the basins of Lake Erie. The purpose was to characterize microbial diversity to gain insight into the factors causing differential functional diversity (nutrients, contaminants, ecological

conditions). To assess spatial patterns of microbial community composition among locations, Fast UniFrac was used, which compares phylogenetic distances of DNA sequences collected from multiple locations (Hamandy et al. 2010). This revealed significant spatial structuring of microbial community composition (Bouzat et al. 2013).

Another question researchers have sought to answer with 16S amplicon sequencing is how microorganisms are transported along the coast. Large volumes of water are flushed through the beach on a daily basis, but just what bacteria are transported and how is a question to be considered, and one that was considered by Boehm et al. (2014). Using massively parallel sequencing to characterize microbial communities present at 49 beaches in California, they identified extensive diversity including 1000 unique taxa from 10 beaches. This indicated the presence of what they called “cosmopolitan” sand microorganisms. There were similarities in microbial communities on beaches with similar grain size, organic carbon content, similar wave climate, and anthropogenic influence. Microbes also enter the water column from the sand (Boehm et al. 2014).

A useful feature of 16S rDNA sequencing is that unique, specific primers can be used for very specific purposes. Lee et all. (2017) explored “novel primer sets for next generation sequencing-based analysis of water quality”. They presented novel new 16S rDNA primer sets that are compatible with NGS approaches and can be used for water quality studies. These new primers show increased specificity for Cyanobacteria and Proteobacteria phyla, meaning increased sensitivity for detection, identification, and relative quantification of toxic bloom-forming microalgae, microbial water quality bioindicators, and common pathogens. With these primers, these taxa accounted for 95% of sequences obtained compared with 50% for standard NGS primers, which provided higher sensitivity and greater phylogenetic resolution of water quality microbial groups. The result is that the increased sensitivity allows parallel sequencing of a greater number of samples through reduced sequence retrieval levels, reducing NGS costs by 50% and still guaranteeing optimal coverage and discrimination of important species (Lee et al., 2017).

The use of these primers is of great benefit, as much research has focused on developing “universal primers” to amplify all taxa with equal efficiency (Klindworth et

al. 2013; Caporaso et al. 2012; Takahashi et al. 2014) in effort to provide broad taxonomic representation and preserve community proportions (Milani et al. 2014; Caporaso et al. 2012; Whiteley et al. 2012; Sinclair et al. 2015; Menchaca et al. 2013). There are shortcomings to these "universal primers", as when inadequate sequencing depth is employed and where some species are strongly dominant, rarer taxa including pathogens, bioindicators, and target groups, may remain undetected (Gofton et al. 2015). This highlights a weakness of 16S amplicon sequencing—outcome depends heavily on use of appropriate primer. The method has its strengths, though. Lee et al. (2017) also stated that "Compared with other loci, the hypervariability and multi-copy nature of the small ribosomal subunit (16S rRNA) gene, coupled with the availability of growing sequence information, confer higher detectability and allow taxonomic classification of bacteria and archaea, potentially to species level (Kermarrec et al. 2013)."

For all of the merits of 16S amplicon sequencing, WGS makes much more information available to researchers. According to Mohuiddin et al. (2017), "shotgun metagenomic sequencing, compared to amplicon sequencing, provides better resolution at the species level and analysis of such sequences helps in understanding the structure and diversity of microbial communities as well as their metabolic potential…The shotgun metagenomic approach may be more sensitive in capturing bacterial diversity than 16S rDNA analysis" (Mohuiddin et al. 2017).

Traditional culture-based methods: Bacteria are cultured on plates and colonies identified.

As discussed above, the US Environmental Protection Agency relies heavily on culture-based methods for coastal water quality monitoring. They use membrane filtration techniques frequently and analyze the colonies that grow from the membrane. ("Approved CWO Microbiological Test Methods")

When Khan et al. (2009) carried out their qPCR analysis for *Aeromonas* in recreational waters, they identified several shortcomings for culture-based methods that were eliminated or reduced through use of qPCR methods. One such weakness in certain applications was is that they only enumerate cells that are culturable. Viable cells that are injured or stressed, or viable but for a given reason not culturable will not be detected

(Pommepuy et al 1996; Santo Domingo et al 2003; Khan et al 2007). Culture-based methods are also not ideal for rapid analysis of large numbers of samples in the context of surveillance or outbreak investigation. To provide an example, these methods when used to isolate and identify *Aeromonas* are time-consuming and labor intensive and can often not identify cultures to a genus and species level. Culture-based methods may also underestimate the concentration of *Aeromonas* because only green centered with opaque margin colonies on selective plates are generally considered *Aeromonas.* Some strains, however, may be Ampicillin-sensitive, which is a problem when Ampicillin is used to suppress background bacterial growth (Havelaar et al. 1987; Holmes & Sartory 1994; EPA 2001). Culture-based methods can underestimate the number of cells, as after prolonged incubation colonies can diffuse and merge (Havelaar et al. 1987; Khan et al. 2009). While culture-based methods effectively identify some culturable indicator bacteria in water samples, they do not provide as comprehensive a picture as other more recent methods.

Sinigalliano et al. (2009) also commented on why no current common methods for regulatory monitoring purposes are completely optimal: "Historically, fecal indicator bacteria (including total and fecal coliforms and enterococci) have been used as indicators for the presence of bacterial, viral and protozoan pathogens (Savichtcheva et al. 2005). These microorganisms are of fecal origin from mammals and birds, and their presence in water may indicate fecal pollution and possible association with enteric pathogens. However, there are major problems with these bacterial indicators, including: short survival in water bodies (McFeters et al. 1974; McFeters 1990); non-fecal source (Scott et al. 2002; Simpson et al. 2002); ability to multiply after release into the water column (Desmarais et al. 2002; Solo- Gabriele et al. 2000); susceptibility to disinfection processes (Hurst et al. 2002); an inability to be used to identify the source of fecal contamination (Field et al. 2003); low levels of correlation with the presence of pathogens (Deetz et al. 1984; Gerba and Rose 1990; Jiang et al. 2001; Griffin et al. 2003; Jiang and Chu 2004; Noble and Fuhrman 2005); and a low sensitivity of detection methods (Horman et al. 2004; Winfield and Groisman 2003). As a result, none of the bacterial indicators currently used meet all the criteria for an ideal indicator. Furthermore, the only detection methodology currently accepted for regulatory purposes (i.e.

enterococci) depends upon culture-based growth and enrichment of the target organisms for an incubation period of at least 18-24 h, so current regulatory methods lack the ability to assess the same-day water quality status of tested water bodies."

Table 1: Summary of Methods

Summary of Methods		
Method	**Strengths**	**Shortcomings**
Whole Genome Sequencing	• When compared with 16S rRNA data, has been shown to detect significantly more phyla and genera in a given sample (Poretsky et al. 2014). • When compared with 16S, eliminates bias associated with the PCR amplification of a single gene (Poretsky et al. 2014). • Faster and more efficient than other methods due to its ability to sequence large quantities of short DNA fragments in parallel (Aguiar-Pulido et al. 2016). • Can be coupled with CosmosID to identify the microbiome of a sample for rapid identification of bacteria, viruses, fungi. • Provides better resolution at the species level than 16S rRNA sequencing data (Mohuiddin et al. 2017).	• Underestimates # of microbial species in samples. Limited by reference databases (Aguiar-Pulido et al. 2016). • Amplification methods can introduce bias and cause differences in taxonomic classifications, so methods must be carefully considered (Kim et al. 2011). • For water quality monitoring applications, WGS methods and procedures must be standardized to provide a baseline for analysis (Mohiuddin et al. 2017).
RNA Sequencing	• Very helpful when combined with WGS data to determine the genes expressed by the	• Constrained by databases. • It is often difficult to determine between host

	microbial community and its functional profile (Aguiar-Pulido et al. 2016; Frias-Lopez et al. 2008; Carvalhais et al. 2012; Gilbert et al. 2008). • Can be carried out using shotgun sequencing methods and reads compared to a reference genome (Aguiar-Pulido et al. 2016). • Direct RNA sequencing methods have shown promise in recent studies, reducing bias introduced by reverse transcription (Ozsolak et al. 2009).	and microbial RNA (if relevant), so having a reference genome for hosts is important (Aguiar-Pulido et al. 2016). • The common practice of reverse transcription introduces biases that interfere with proper characterization of transcripts (Ozsalak et al. 2009).
qPCR for Biomarker	• Allows users to identify the presence and quantity of a given species in a sample, such as pathogens in drinking water (Khan et al. 2009), or indicator species for fecal pollution (Sinigalliano et al. 2009). • Recently introduced by the EPA as a possibility for identification of enterococci in water quality monitoring settings (EPA).	• Limited in scope—only identifies the queried species. • Can overestimate the number of viable cells due to detection of dead cells. To counter this, users can quantify viable and dead cells prior to applying the qPCR assay (Khan et al. 2009; Nocker et al. 2006).
16S rDNA Sequencing	• There are well-established 16S rDNA databases for the identification of microbes (GreenGenes, MicroSeq ID). • Can be used with unique primers to	• Limited by short read lengths obtained (Poretsky et al 2014; Quince 2009 and 2011). • Among closely related species, the resolution of 16S rRNA gene is limited, making proper

	increase the sensitivity of detection for certain phyla, providing greater sensitivity and resolution of the target species (Lee et al. 2017).	identification a challenge (Poretsky et al. 2014). • Constrained by databases.
Culture-based methods	• Commonly accepted method for water quality analysis by the EPA. Highly standardized. (EPA) • Fecal indicator bacteria can be used to identify potential fecal pollution in water bodies. (EPA)	• Limited in scope to viable and culturable organisms. Injured or stressed cells that will not grow on culture will not be detected, regardless of presence or absence in sample. (Pummepuy et al. 1996; Santo Domingo et al 2003; Khan et al. 2007) • Slow process--labor intensive and time-consuming. Cultures often cannot be identified to a genus and species level. (Khan et al. 2007)

Available Products for Whole Genome Sequencing and Whole Transcriptome Sequencing

As Whole Genome Sequencing and Whole Transcriptome sequencing show incredible promise for the future of water quality monitoring in the coastal zone, available products from the top companies will be presented and analyzed below.

Illumina offers a broad range of products from nucleic acid extraction all the way through bioinformatic analysis. For whole genome and whole transcriptome sequencing, different library preparation kits must be used but the libraries can be run on the same sequencing platforms. For whole genome shotgun sequencing, DNA libraries can be prepared using the Nextera XT DNA Library Prep Kit. This offers Illumina's fastest assay time of 90 minutes. Fragmentation is enzymatic, eliminating the need for mechanical shearing. The prep kit also allows for low-quality DNA input. This kit is

good for “small genomes, PCR amplicons greater than 300 bp, plasmids, microbial genomes, concatenated amplicons, and double-stranded cDNA” (Nextera), so there is a wide range of flexibility. Samples are normalized using bead normalization during the kit procedures, eliminating the need for library quantification before sequencing. DNA is tagged with sequencing adapters as it is fragmented in a single-tube enzymatic reaction. This kit accepts extremely low inputs of DNA—as low as 1 ng. For high throughput studies, researchers can multiplex up to 384 samples per library (Nextera).

For RNA library preparation, the TruSeq RNA Library Prep Kit v2 provides users with a streamlined and quick library preparation experience. The total hands-on time for this kit is 4.5 hours, with a total time start to finish of 12 hours as opposed to 16 hours for other methods. The number of steps is reduced from ~49 to 16, with the majority of the pipetting steps replaced with the use of master mixed reagents. With automation, users can process up to 96 samples in parallel, making the kit economical and offering substantial time savings for high-throughput studies. This kit will process libraries with as little as .1 µg of RNA (TruSeq). [see graphic on next page for comparison to other kits]

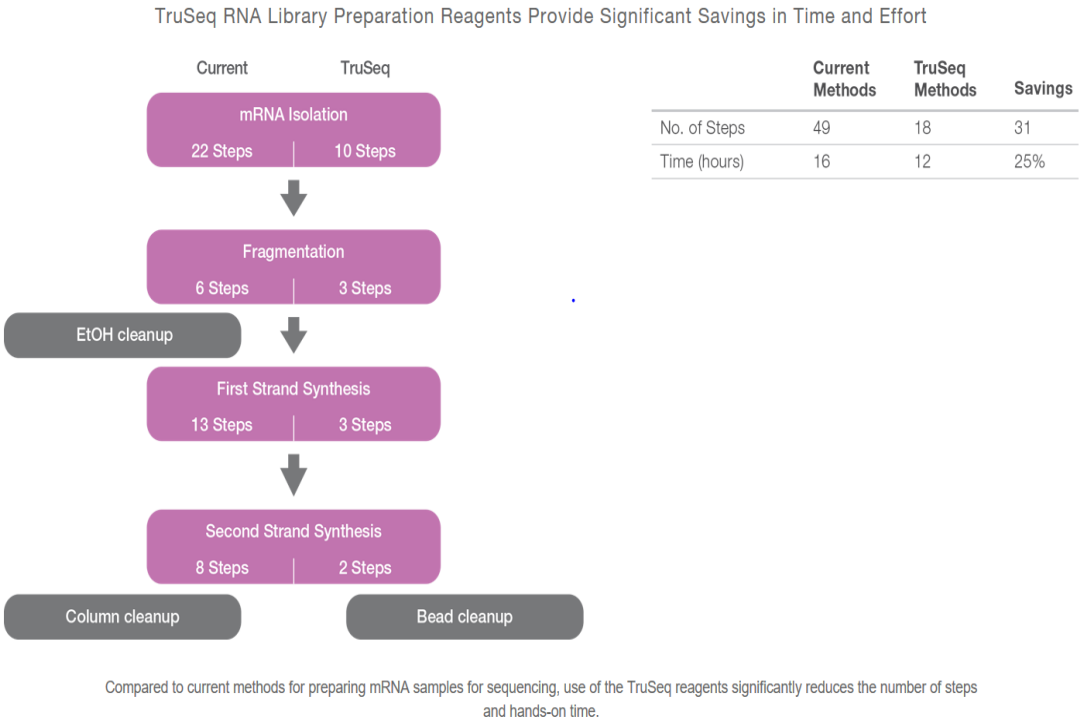


Figure 1: TruSeq RNA Library Preparation Kit Workflow and Time Savings (Ref: TruSeq)

For sequencing, Illumina makes two sequencing platforms optimal for coastal zone management purposes: the NextSeq550 and the NovaSeq6000. The NextSeq550 is ideal for whole genome metagenomics, and a very powerful sequencer for small genome sequencing applications. It boasts mid-output ad high-output run modes. This sequencer is used for high-throughput applications and enables transcriptome sequencing as well. It enables users to tune read length and configure output to meet needs. It offers quick turnaround for samples, from preparation steps to data in just over a day. After loading into NextSeq550, data can be generated in as little as 12 hours for a 75-cycle sequencing run, and approximately 30 hours for paired 150-cycle runs. It also scans microarrays and can switch to lower throughput sequencing as needed. Since it does both microarray scanning and integrated sequencing, it eliminates the need for multiple additional instruments. This platform is designed to be very user-friendly, touting a “load-and-go” design, minimizing the need for extensive user training or instrument set-up time. This sequencer can sequence between 1-20 transcriptomes per run, and between 9-96 targeted

panels. Paired-end read lengths are user-defined, allowing users to set read lengths of up to 2x150 bp. The platform uses Illumina SBS (sequencing by synthesis) chemistry. “This proprietary, reversible, terminator-based method enables the parallel sequencing of millions of DNA fragments, detecting single bases as they are incorporated into growing DNA strands.” For data analysis, there are integrated instrument computers that perform “base calling and quality scoring”. Sequencing data can be analyzed using a wide range of products, to be discussed later. Microarray scanning can be used as a complementary method to sequencing and can enable further “exploration or confirmation of copy number variants detected through sequencing (NextSeq 550). See Figure 2 below for details on the NextSeq550 workflow.

Figure 2: NextSeq 550 System Workflow—The NextSeq 550 System offers a simple, integrated workflow from sample preparation to data analysis. Workflow times will vary by experiment and assay type. Details shown are for an mRNA expression profiling experiment assuming 2 × 75 bp on instrument. Analysis results include differential expression and identification of alternative transcripts.

Figure 2: NextSeq 550 System Workflow (Ref: NextSeq 550)

The NovaSeq6000 is a newer platform that offers many of the same features as the NextSeq 550 system, with several additions to streamline and add flexibility to the sequencing process. This also uses Illumina SBS Chemistry and can be used for a wide variety of sequencing applications. The platform offers great flexibility and time saving options, such as the ability to individually load each flow cell lane, allowing for the inclusion of samples from different libraries if desired. The platform provides the option to use one of four unique flow cells depending on sequencing needs and desired output,

and any two flow cells can run concurrently. A chart showing output range of the flow cells when compared with Illumina's HiSeq systems is shown below in Figure 3. Figure 4 below also shows specifications for the flow cells, with options to select between 2x50, 2x100, or 2x150 bp read lengths. 384 samples can be run in one flow cell each time if the optional NovaSeqXp Workflow is selected (NovaSeq 6000).

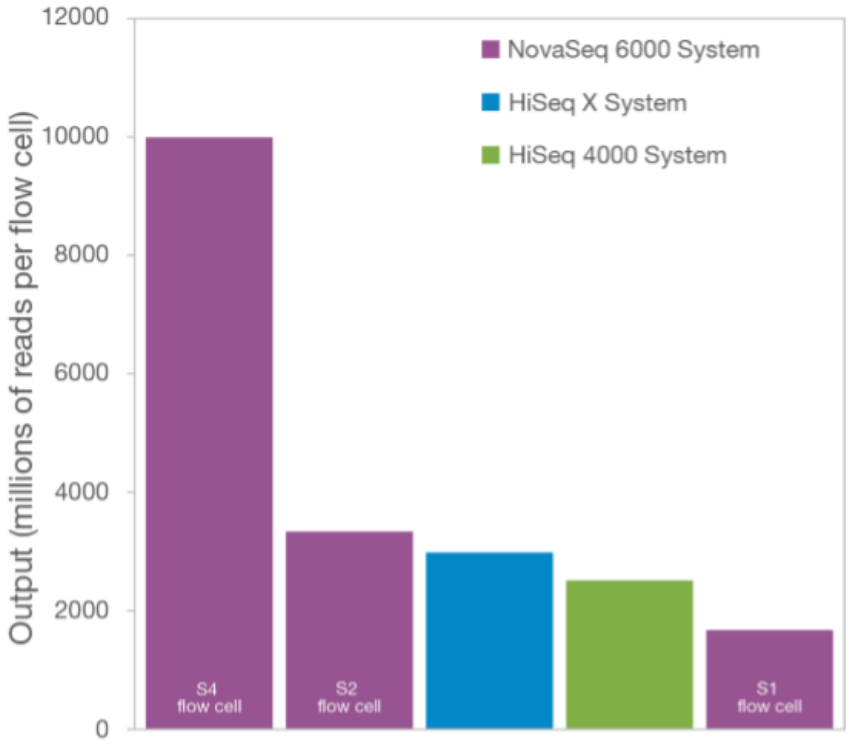


Figure 2: The NovaSeq 6000 System Offers the Broadest Output Range—The NovaSeq 6000 System generates from 134 Gb and 1.3 B reads to 3 Tb and 10 B reads of data in single flow cell mode. In dual flow cell mode, output can be up to 6 Tb and 20 B reads. The tunable output makes the NovaSeq 6000 System accessible for a wide range of applications.

Figure 3: NoveSeq 6000 System Output Range Comparison for Available Flow Cells (Ref: NovaSeq 6000)

Table 1: NovaSeq 6000 System Flow Cell Specifications

Flow Cell Type	S1	S2	S4
Lanes per Flow Cell	2	2	4
Output per Flow Cell[a]			
2 x 50 bp	134-167 Gb	333-417 Gb	N/A[b]
2 x 100 bp	266-333 Gb	667-833 Gb	N/A[b]
2 x 150 bp	400-500 Gb	1000-1250 Gb	2400-3000 Gb
Single Reads (Clusters Passing Filter)	1.3-1.6 B	3.3-4.1 B	8-10 B
Paired-End Reads (Clusters Passing Filter)	2.6-3.2 B	6.6-8.2 B	16-20 B
Quality Scores[c]			
2 x 50 bp		≥ 85%	
2 x 100 bp		≥ 80%	
2 x 150 bp		≥ 75%	
Run Time[d]			
2 x 50 bp	~13 hr	~16 hr	N/A[b]
2 x 100 bp	~19 hr	~25 hr	N/A[b]
2 x 150 bp	~25 hr	~36 hr	~44 hr

a. Output and read number specifications based on a single flow cell using Illumina PhiX control library at supported cluster densities. The NovaSeq 6000 System can run one or two flow cells simultaneously.
b. N/A: not applicable.
c. Quality scores are based on NovaSeq S2 Reagent Kits run on the NovaSeq 6000 System using an Illumina PhiX control library. Performance may vary based on library type and quality, insert size, loading concentration, and other experimental factors.
d. Run times are based on running two flow cells of the same type. Starting two different flow cells will impact run time.

Figure 4: NovaSeq 6000 System Flow Cell Specifications (Ref: NovaSeq 6000)

After sequencing on one of the available platforms, data can be uploaded immediately into Illumina's BaseSpace Sequence Hub for data management, data analysis, and cloud storage. The BaseSpace Sequence Hub Professional Account is free for cloud-based storage and offers a wide array of applications for users to analyze their data. For metagenomic analysis, there are three main applications that can analyze DNA and identify microorganisms. Kraken Metagenomics assigns taxonomic labels to short DNA sequences with high sensitivity and speed using exact alignment of k-mers and a novel classification algorithm (Kraken Metagenomics). MetPhlAn profiles microbial community composition from metagenomic shotgun sequence data. It uses unique clade-specific marker genes identified from reference genomes to determine taxonomic assignments (MetaPhlAn). CosmosID, the newest and most versatile of the three, uses k-mer based assembly of reads and an extensive database to identify bacteria, viruses, and fungi, and parasites, often even down to the strain level in just minutes. This process requires no prior assumptions regarding what is present in the sample. Output includes name of organism, taxonomic hierarchy, frequency of hits, and estimated relative abundance of the organism in the sample (GENIUS). For RNA sequencing, many of the Base Space Applications are limited in application, as they contain reference databases for exclusively human and a very limited number of other species. The National Institute of Health's National Center for Biotechnology Information, however, hosts the Basic Local Alignment Search Tool (BLAST), which is a program that locates similar regions between biological sequences and compares nucleotide or protein sequences to databases. The program is used to infer functional and evolutionary relationships between sequences or to identify members of gene families and calculated statistical significance on all matches. This is not, however, an application that can be used through BaseSpace. Data would have to be uploaded to BLAST and analyzed outside of BaseSpace. Culley et al. (2006) used BLAST to identify RNA viruses in coastal samples. BLAST contains human, mouse, rat, and microbial genome reference databases (Basic Local Alignment Search Tool).

ThermoFisher is another large company providing products for the entire sequencing workflow. For shotgun metagemonics, users can start with the Ion Xpress Plus library prep kit. This library prep kit fragments either enzymatically or

mechanically, then end repairs the fragments. Adapters are attached, and users can select between barcoded and non-barcoded. Size selection is then carried out, followed by amplification steps. This library prep kit requires low input of DNA (50-100 ng if more reactions out of the kit are desired, or 1 µg otherwise) (Ion Xpress Plus). The below graphic shows the IonXpress Plus Library prep kit workflow.

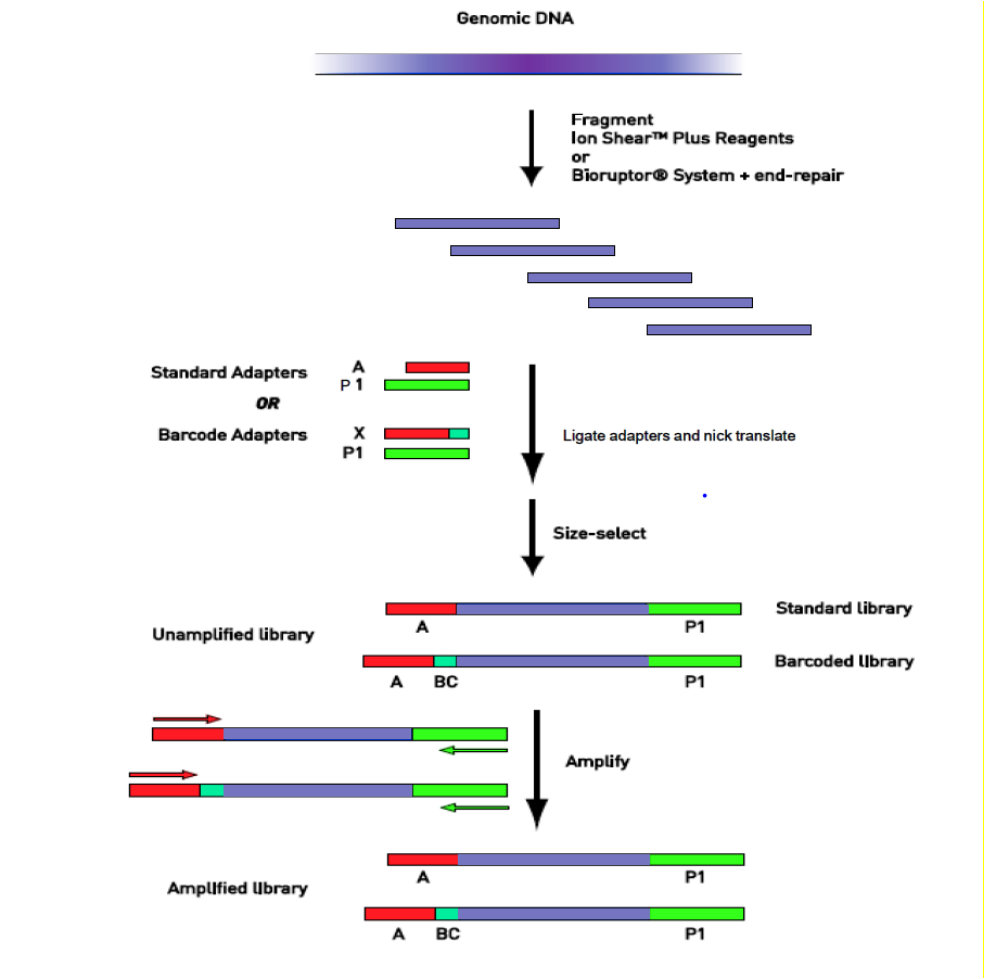


Figure 5: IonXpress Library Prep Kit Workflow (Ref: Ion Xpress User Guide)

For metatranscriptomics, users can start with the Ion Total RNA-Seq v2 library prep kit. This kit creates whole transcriptome libraries while preserving strand information. 100 ng total RNA is required. This kit fragments the RNA and attaches a 3' and 5' adapter. The sequences undergo ligation and reverse transcription, followed by

amplification of cDNA and the addition of barcoded adaptors. Beads are used for cleanup and size selection (Ion Total RNA-Seq).

Following library preparation, template preparation is required, and the Ion Chef system can be used. During amplification, amplified products bind to B' capture primer. Polymerase extends from B'. A complementary sequence is extended, and the amplified product dissociates from the bead (ThermoFisher 2015). See Figure 6 for a visual representation of this process.

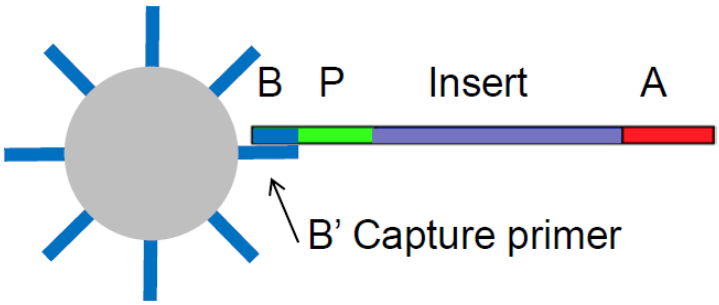


Figure 6: Ion Chef Template Preparation (ThermoFisher 2015)

The products are then placed on a chip to be run and sequenced the one of the Ion S5 system platforms (see Figure 7 for a summary of the three S5 options). The principle behind the flexibility of these sequencing platforms are the 5 different types of chips. They use Ion Torrent semiconductor sequencing technology to deliver fast sequencing results. The sequencer floods the chip with one nucleotide after another. When a nucleotide is incorporated into a strand of DNA on the chip, a hydrogen ion is released, changing the pH of the solution surrounding the chip. If a nucleotide is not incorporated, there will be no pH change detected. Readings are recorded instantly as nucleotides are incorporated into DNA strands (Ion Torrent NGS Technology).

If deciding between the Ion S5 or the S5 Plus/ Prime, it is helpful to know that the S5 takes 510-540 chips, whereas the S5 Plus/ Prime can take the 550 chip. Specifications should be carefully reviewed prior to making this determination. See Figures 8 and 9 for chip specifications and for guidance on chip selection by application (Ion GeneStudio Spec Sheet).

Figure 7: Ion GeneStudio Series Comparison (Ref: Ion GeneStudio Spec Sheet)

Chip type	Number of reads	Read length (output*)	Ion GeneStudio™ S5 System	Ion GeneStudio™ S5 Plus System	Ion GeneStudio™ S5 Prime System
			Turnaround time (sequencing run** plus analysis time)		
Ion 510™ Chip	2–3 million	200 bp (0.3–0.5 Gb)	4.5 hr	3 hr	3 hr
		400 bp (0.6–1 Gb)	10.5 hr	5 hr	5 hr
Ion 520™ Chip	4–6 million	200 bp (0.6–1 Gb)	7.5 hr	3.5 hr	3 hr
		400 bp (1.2–2 Gb)	12 hr	5.5 hr	5.5 hr
	3–4 million	600 bp (0.5–1.5 Gb)	12 hr	5.5 hr	5.5 hr
Ion 530™ Chip	15–20 million	200 bp (3–4 Gb)	10.5 hr	5 hr	4 hr
		400 bp (6–8 Gb)	21.5 hr	8 hr	6.5 hr
	9–12 million	600 bp (1.5–4.5 Gb)	21 hr	8 hr	7 hr
Ion 540™ Chip	60–80 million	200 bp (10–15 Gb)	19 hr	10 hr	6.5 hr
		200 bp (20–30 Gb) 2 runs in 1 day	NA	20 hr	10 hr†
Ion 550™ Chip	100–130 million	200 bp (20–25 Gb)	NA	11.5 hr	8.5 hr
		200 bp (40–50 Gb) 2 runs in 1 day	NA	NA	12 hr†

* Expected output with >99% aligned or measured accuracy. Output dependent on read length and application.
** Sequencing run times are between 2.5 and 4 hr.
† Analysis of first run occurs concurrently with the second sequencing run.

Figure 8: Ion Chip Comparison (Ref: Ion GeneStudio Spec Sheet)

	Ion 510™ Chip	Ion 520™ Chip	Ion 530™ Chip	Ion 540™ Chip	Ion 550™ Chip
Max. output (reads)	3 M	6 M	20 M	80 M	130 M
Targeted DNA sequencing ** e.g., Ion Torrent™ Oncomine™ Focus Assay	•	•	•	•	•
Small genome sequencing† e.g., Bacterial typing using Ion Xpress™ Plus Fragment Library Kit		•	•		
16S metagenomics sequencing†† e.g., Ion 16S™ Metagenomics Kit		•	•		
Exome sequencing e.g., Ion AmpliSeq™ Exome Panel				•	•
Targeted RNA sequencing e.g., Ion AmpliSeq™ made-to-order RNA panels	•	•	•	•	•
miRNA/small RNA profiling e.g., Ion Total RNA-Seq v2 Kit	•	•	•		
Targeted transcriptome sequencing e.g., Ion AmpliSeq™ Transcriptome Human Gene Expression Kit				•	•
Whole transcriptome sequencing e.g., Ion Total RNA-Seq v2 Kit				•	•
Low-pass whole genome sequencing (PGS) e.g., Ion ReproSeq™ PGS Kit	•	•	•		

Five Ion Torrent™ sequencing chips achieve 2–130 M reads per run (or 2–260 M reads per day) to enable a broad range of sequencing applications.

Figure 9: Ion Chip Selection Guide (Ref: Introducing GeneStudio)

From a bioinformatics perspective, the company offers access to the MicroSeq ID database and the GreenGenes database. The MicroSeq ID database is a database created for use with 16S sequencing but can also be used for WGS applications. This will provide data on bacteria in the sample but will not reflect presence of viruses and fungi. This database has data on around 9000 organisms (MicroSEQ). Users can run sequences through this database, and if it there is no good alignment, then through the GreenGenes database which has data on 400,000 organisms (also designed for use with 16s methods)

(GreenGenes; Dutken 2018a). COSMOS ID also has a plug-in through Thermo Fisher's cloud, which would identify bacteria, viruses, and fungi (Dutken 2018a).

For whole transcriptome data analysis, users can make use of the Ion Torrent RNA-sequence workflow, which allows for efficient sequencing and analysis of data. BLAST can be used to analyze data once sequenced (RNA Sequencing).

Storage in the Ion Reporter cloud is free if users store under 1 TB of data. If network agreements cannot be reached through facilities to allow for cloud storage from government or university labs, then local Ion Reporter software can be purchased for $30K. Applications are built into or linked to the Ion Reporter software (Dutken 2018a).

Qiagen is another company with a great number of products for sequencing, however it is helpful more for sample and library preparation than anything else. They have a sequencer, but only for clinical applications. Their primary utility for water quality analysis purposes would be preparation for sequencing on Illumina or Ion Torrent sequencers. They have a homogenizer called the Powerlyser 24 that homogenizes samples (PowerLyzer 24). They offer the QIACube for automation of nucleic acid extraction and library preparation. They offer spectrophotometers for quantitation of DNA. They have the RotorGene Q for real-time PCR, and a Pathways Analysis option that maps data back to published data (RotorGene Q). For sample preparation, the QIACube is available, and automates the process. Some kits for use with the QIACube are specific for DNA, and others for RNA. The QIAcube purifies DNA, RNA and proteins, and automates the use of QIAGEN spin-column kits (purchased separately) and eliminates manual processing steps. This can purify and prepare up to 12 samples per run. DNA purified using QIACube performs well in sensitive PCR even with the use of large amounts of eluate. The platform offers ease of use by eliminating the need for standalone computers. A touch screen allows users to easily change settings on the machine. Kits are available for the purification of RNA, genomic DNA, and viral RNA (QIACube).

Qiagen offers very specific library prep kits, including the DNeasy Power Water Kit and the RNeasy Power Water Kit. These kits isolate DNA and RNA (respectively) from filtered water samples. High quality DNA or RNA can be extracted even from highly contaminated water samples. The extracted DNA can then be used for any downstream application, including Sanger, qPCR, and WGS. The DNA is ready to use in

a final 100 µl volume. The RNeasy Power Water Kit isolates RNA from bacteria (Gram + or -), algae, and fungi. Processing is complete in 40 minutes. The extracted RNA can be used in RT-PCR, qPCR, RNA-seq, and other downstream applications (DNeasy; RNeasy).

Like ThermoFisher and Illumina, Qiagen also has a relationship with CosmosID and has incorporated an optional Plugin for the CLC Genomics Workbench (Qiagen's cloud-based bioinformatics workbench). The CosmosID plugin uses novel techniques to detect pathogens, antimicrobial resistance genes, or track microbial community changes. This plugin is great for microbiome studies using shotgun metagenomic data. It can discriminate pathogens and near neighbors with no prior assumptions about the sample. CosmosID takes raw, unassembled reads and matches the sequences against their reference database (GenBook) of bacteria, fungi, parasites, and antibiotic resistance and virulence factors. GenBook draws from private and public databases of assembled genomes "and constitutes hundreds of millions of marker sequences representing both coding and non-coding sequences that are shared or uniquely identified across taxonomic or phylogenetic levels." Results are returned in minutes, and since CosmosID is incorporated into the CLC Workbenches and the QIAGEN Microbial Genomics Pro Suite, users have direct access to powerful statistical analyses, interactive visualizations, and other NGS tools for microbiology. The plugin completes phylogenetic placement for identification, community resistome and virulome characterization, and identification at sub-species and strain level (CosmosID Plugin).

Pricing for Sequencing

Table 2: Illumina Pricing Information (Ref: BaseSpace, Nextera, TruSeq, Kolas, 2018)

Illumina		
Step	Product	Pricing for Purchase
Library Preparation [DNA]	Nextera XT DNA Library Prep Kit	For 24 Samples: $798 For 96 Samples: 3030
		Index Kit 96 Samples: $258
Library Preparation [RNA]	TruSeq RNA Library Prep Kit v2	$3880 for 48 samples
Sequencing	NextSeq550	$275000
	NovaSeq 6000	$985000
Bioinformatics	BaseSpace Sequence Hub	Free for Cloud Storage/ Free Professional Account [Apps charge per use-1 iCredit=$1)
	CosmosID (metagenomics-bacterial ID)	3 iCredits per node/ hr No license cost
	Kraken Metagenomics (shotgun metagenomics-taxonomic classification)	3 iCredits per node/hr No license cost
	MetPhIAn (metagenomics-phylogenetic analysis)	3 iCredits per node/hr No license cost
	RNA-Seq Translator (RNA-seq/ generates protein sequences)	3 iCredits per node/hr No license cost

Table 3: Thermo Fisher Pricing Information (Ref: IonXpress Plus Fragment, Ion Total RNA, Ion 16S, Ion 530 Chip, Ion 530 Chip, Ion 540 Chip, GreenGenes, MicroSeqID, Dutken 2018a)

Thermo Fisher Whole Genome and Whole Transcriptome Sequencing		
Step	Product	Pricing for Purchase
Library Preparation [DNA]	Ion Xpress Plus	$906 for 10 reactions
Library Preparation [RNA]	Ion Total RNA Seq v2	$1538 for 12 reactions
Template Preparation [Ion Chef]	Ion Chef	$55000
Ion Chip	Ion 520 Chip	$3120/ 8 chips
	Ion 530 Chip	$4495/ 8 chips
	Ion 540 Chip	$5350/ 8 chips
Sequencing	Ion Gene Studio S5	$65000
	Ion Gene Studio S5 Plus	$120000

	Ion Gene Studio S5 Prime	$150000
Bioinformatics	Ion Reporter Software	(Free for cloud/ $30000 for software purchase)
	CosmosID	Recently Added Plug-In
	GreenGenes Database	Component of Ion Reporter Software
	MicroSeq ID Database	Component of Ion Reporter Software

Table 4: Qiagen Pricing Information (Ref: PowerLyzer 24, DNeasy, RNeasy, QIAseq FX, Wright 2018.)

Qiagen		
Step	Product	Pricing for Purchase
Sample Preparation	Powerlyzer 24	$10305
	DNeasy Power Water Kit	$501 for 50 samples
	RNeasy Power Water Kit	$501 for 50 samples
Library Preparation	QIACube (automation)	$20.7 K
	QIAseq FX DNA Library Kit	$828 for 24 samples $3280 for 96 samples

ThermoFisher and Illumina also provided cost-per-sample analyses for some of their most effective sequencing workflows.

Sample Illumina Sequencing Costs

Table 5: Illumina 16S Sequencing Costs (Schellhaas 2018)

SAMPLE ILLUMINA 16S SEQUENCING COSTS		
Estimated Costs Per Sample		
MiSeq ~ 250 spls/run		
Sequencing	$6.32	
Library Prep Indexes	$2.66	
	$8.98	**per sample**
NextSeq Mid-Output v2 Kit ~ 384 spls/run		
Sequencing	$4.66	
Library Prep Indexes	$2.66	
	$7.32	**per sample**

NovaSeq S1 Flow Cell (300 cycle) ~ 2x384 spls/run		
Sequencing	$11.72	
Library Prep Indexes	$2.66	
	$14.38	**per sample**

Table 6: Illumina Shotgun Sequencing Costs (Schellhaas 2018)

SAMPLE ILLUMINA SHOTGUN SEQUENCING COSTS			
Estimated Costs Per Sample			
NextSeq	**4 spls/run (100M reads)**	**8 spls/run (50M reads)**	
Sequencing	$1,170.00	$585.00	
Library Prep Indexes	$5.00	$5.00	
Library Prep	$42	$42	per sample
	$1,217.00	**$632.00**	
NovaSeq S4 Flow Cell ~ 96 spls/run			
Sequencing	$320.63		
Library Prep Indexes	$5.00		
Library Prep	$42.00	per sample	
	$367.63		

Sample ThermoFisher Sequencing Costs

Table 7: ThermoFisher 16S Sample Sequencing Costs (Dutken 2018b)

				List price	Cost per sample	Cost per chip	Cost per 100 samples
Ion 16S ™ Metagenomics	A26216	100 reactions per kit	The Ion 16S™ Metagenomics Kit is designed for rapid, comprehensive and broad-range analyses of mixed microbial populations using the Ion Torrent semiconducter sequencing workflow. The kit permits PCR amplification of hypervariable regions of the 16S rDNA gene from bacteria. In combination with the Ion Xpress(TM) Barcoding Kits,	$962.00	$9.62	$384.80	$962.00
Ion Plus Fragment Library Kit	A28950	1 kit -48 samples per kit	The Ion Plus Fragment Library Kit is an integral component of the Personal Genome Machine sequencing workflow. The Ion Fragment Library Kit contains sample preparation reagents for 48 DNA libraries for semiconductor sequencing	$2,080.00	$43.33	$1,733.33	$4,333.33
Ion Xpress Barcode Adaptors 1-96 Kit	4474517	160 reactions per 16 plex for gDNA, 640 for ampliseq	The Ion Xpress(TM) Barcode Adaptors 1-96 Kit contains reagents to conduct multiplexed sequencing analysis of 10 sets of 96 gDNA libraries or 40 sets of 96 Ion AmpliSeq(TM) libraries. When used in combination with the Ion Xpress™ Fragment Library Kit or the Ion Fragment Library Kit, the kit enables pooling of barcoded fragment libraries prior to template preparation, drastically reducing cost per sample.	$9,030.00	$56.44	$2,257.50	$5,643.75
Ampure XP	NC9933872	$400.00		$1,481.55	$3.70	$148.16	$370.39
Qubit hsDNA	Q32854	500 assays		$281.00	$0.56	$56.20	$112.40
Ion 520™ Chip Kit	A27762	8 chips	The Ion 520™ Chip Kit contains 8 barcoded chips for semiconductor-based sequencing with a high-density array of wells to perform massively parallel DNA sequencing and generate 3-5 M reads with support for up to 400 base read libraries.	$3,120.00	$7.80	$390.00	$780.00
Ion 510/520/530 Kit Chef v2 (launch May 2017)	A34019	8 rxns (1 run/init)	The Ion 510/520/530™ Kit enables robust and reproducible template preparation, chip loading, and sequencing of up to 200 base-read libraries on the Ion Chef™ system for the Ion S5/S5 XL system sequencing workflow. The Ion 520/530™ Kit contains reagents for template preparation and sequencing of 8 Ion 520™ or 530™ barcoded chips.	$5,690.00	$14.23	$711.25	$1,422.50
Total			Total	$22,644.55	$135.68	$5,681.24	$13,624.37

Table 8: Illumina Shotgun Sequencing Sample Cost Per Sample Analysis (2M reads)

						List Price	Cost per sample	Cost per chip	Cost per 100 samples	
Ion Xpress™ Plus Fragment Library Kit	4471269	1 kit up to 20 rxns			The Ion Xpress™ Plus Fragment Library Kit provides streamlined library construction methodology for the Personal Genome Machine sequencing workflow. The Ion Xpress™ Fragment Library Kit contains sample preparation reagents for enzymatic shearing and library construction for up to 20 DNA libraries (depending on input DNA type) for semiconductor sequencing with 100bp or 200 bp run configuration.	$906.00	$45.30	$453.00	$4,530.00
	4474517	160 reactions per 16 plex for gDNA, 640 for ampliseq			The Ion Xpress(TM) Barcode Adaptors 1-96 Kit contains reagents to conduct multiplexed sequencing analysis of 10 sets of 96 gDNA libraries or 40 sets of 96 Ion AmpliSeq(TM) libraries. When used in combination with the Ion Xpress™ Fragment Library Kit or the Ion Fragment Library Kit, the kit enables pooling of barcoded fragment libraries prior to template preparation, drastically reducing cost per sample.	$9,030.00	$56.44	$564.38	$5,643.75
	Ampure XP	NG9933872	400			$1,481.55	$3.70	$37.04	$370.39
	Qubit hsDNA	Q32854	500 assays			$281.00	$0.56	$5.62	$56.20
	Ion 530™ Chip Kit	A27764		8 chips	The Ion 530™ Chip Kit contains 8 barcoded chips for semiconductor-based sequencing with a high-density array of wells to perform massively parallel DNA sequencing and generate 15-20M reads with support for up to 400 base read libraries.	$4,495.00	$56.19	$561.88	$5,618.75
	Ion 510/520/530 Kit Chef v2 (launch May 2017)	A34019	8 rxns (1 run/init)		The Ion 510/520/530™ Kit enables robust and reproducible template preparation, chip loading, and sequencing of up to 200 base-read libraries on the Ion Chef™ system for the Ion S5/S5 XL system sequencing workflow. The Ion 520/530™ Kit contains reagents for template preparation and sequencing of 8 Ion 520™ or 530™ barcoded chips.	$5,690.00	$71.13	$711.25	$7,112.50
					Total	$21,883.55	$233.32	$2,333.16	$23,331.59

As can be observed from the cost differences between 16S sequencing and shotgun sequencing by Illumina and ThermoFisher, the latter costs significantly more. This cost difference, however, is worth the expenditure in many applications as the wealth of information gathered through WGS increases dramatically. As will be shown in the Quantitative Analysis portion of this capstone, shotgun sequencing methods in conjunction with k-mer based species identification allows for identification of bacterial, viral, and fungal microorganisms, as well as antibiotic resistance genes and virulence factors. 16S sequencing allows for bacterial identification exclusively.

Quantitative Analysis

Data from 3 separate storm water catchment areas and 1 municipal tap water source were provided by this capstone committee. COSMOS ID was used to analyze shotgun sequencing data from the samples. The same samples had previously been analyzed using traditional culture-based techniques to identify fecal coliform bacteria as well as enterococci bacteria (See Figure 10). Samples 1004, 1003, and 1002 were from stormwater catchments, and 1001 was from a municipal tap water source (McCorquodale and Duncan 2018).

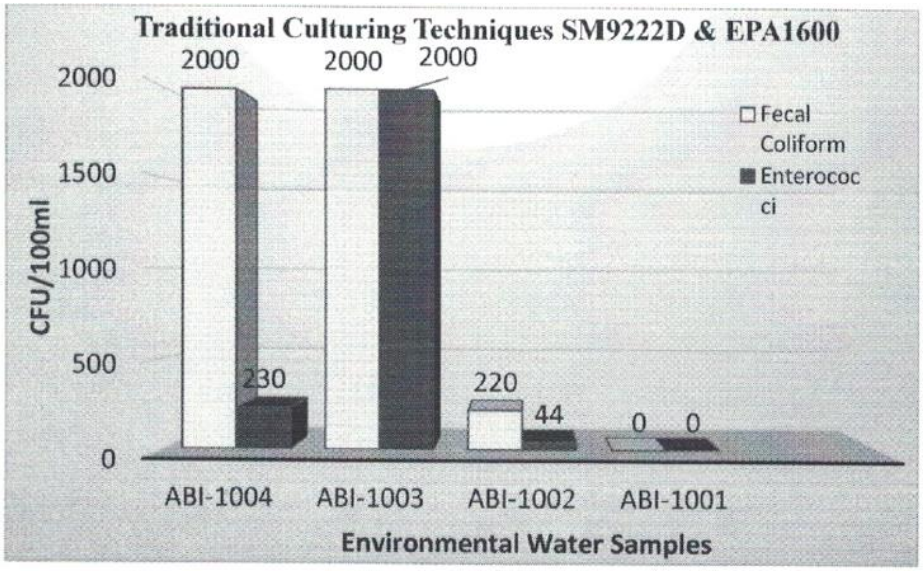
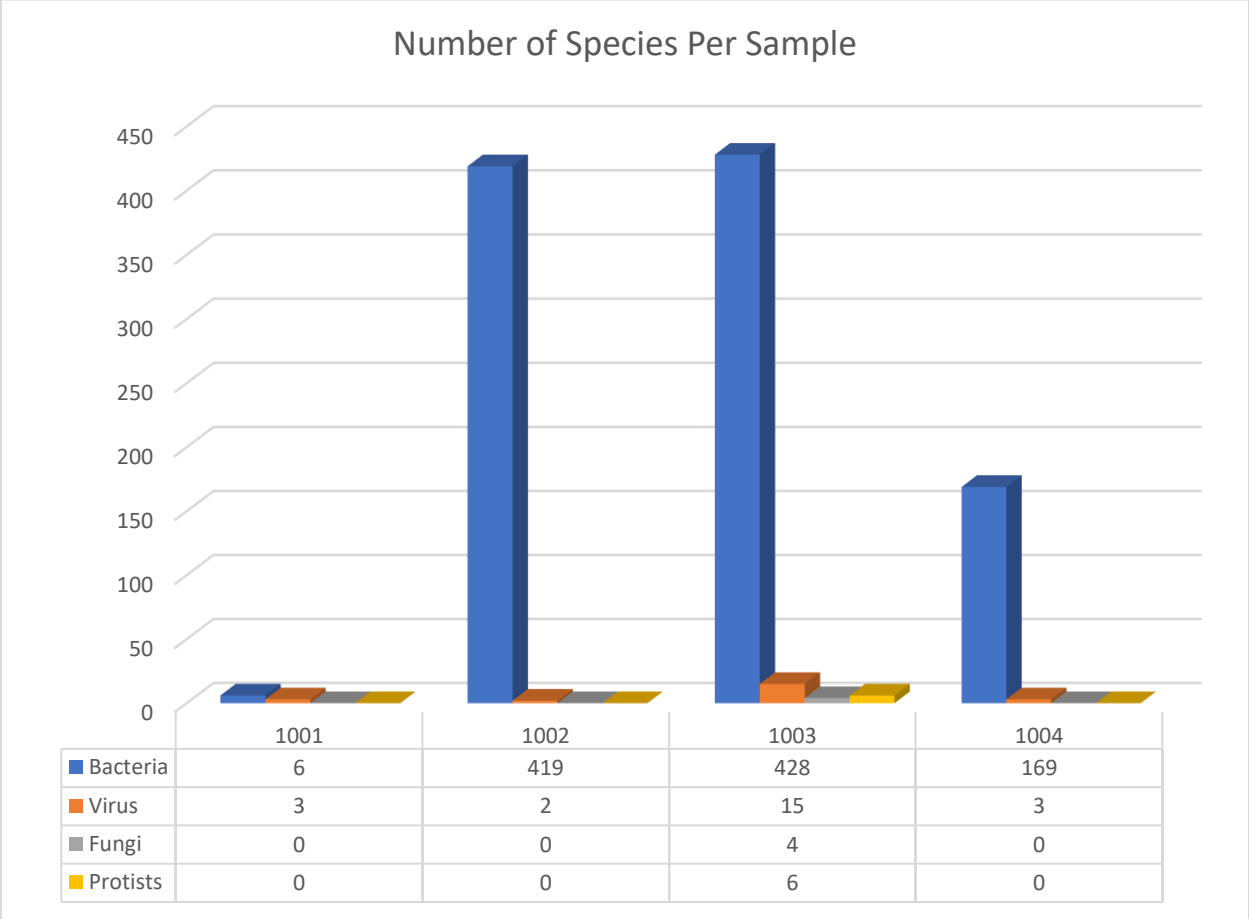


Figure 10: Capstone committee-provided culture results for samples

Figure 10 demonstrates the utility of utilizing traditional culture-based techniques to identify the presence or absence of indicator bacteria in water samples. Fecal coliforms as well as enterococci bacteria were detected in all three catchment samples. However, it is not apparent from this data whether the specific strains found originate from human or environmental sources, or whether they pose a risk to human health. Traditional culture techniques greatly limit scope as researchers and managers can only draw information on a small number of indicator bacteria. The utility is limited greatly by this. More recent techniques, in particular shotgun sequencing combined with k-mer based microbial identification procedures, provide researchers with a broad expanse of information that helps to determine source as well as antibiotic resistance and virulence. Figure 11 shows the number of species of bacteria, viruses, fungi, and protists detected by COSMOS ID. Hundreds of species were identified in the catchment samples, providing a much larger scope of information.



	1001	1002	1003	1004
Bacteria	6	419	428	169
Virus	3	2	15	3
Fungi	0	0	4	0
Protists	0	0	6	0

Figure 11: Number of Species Per Sample

Beyond simply identifying hundreds of species, COSMOS ID allows for the analysis of enteric pathogens in the samples. Nine common enteric pathogens were selected, and their relative abundance in each sample identified. Figure 12 shows these relative abundances. Each sample contains one or more of these enteric pathogens, with each sample containing widely varying abundances of these pathogens. Sample 1002 showed a dominance of *Bacteroidetes spp.* and *Enterobacter spp.*, while sample 1003 showed a split between *Enterobacter spp.*, *Pseudomonas aeruginosa*, and *Staphylococcus spp.* Within sample 1001, there were essentially no bacteria found; but one enteric species (*Pseudomonas aeruginosa*) appeared in low abundance. It's relative abundance within the sample was therefore very high, since only six bacterial species were found in total. These data show that within the samples, most of the bacteria found are not traditional enteric pathogens. However, these pathogens were found in high enough abundance to be of great interest. Further analysis of the virulence factors and antibiotic resistance genes provide more insight into whether the sources of these pathogens are environmental or have the potential to infect humans.

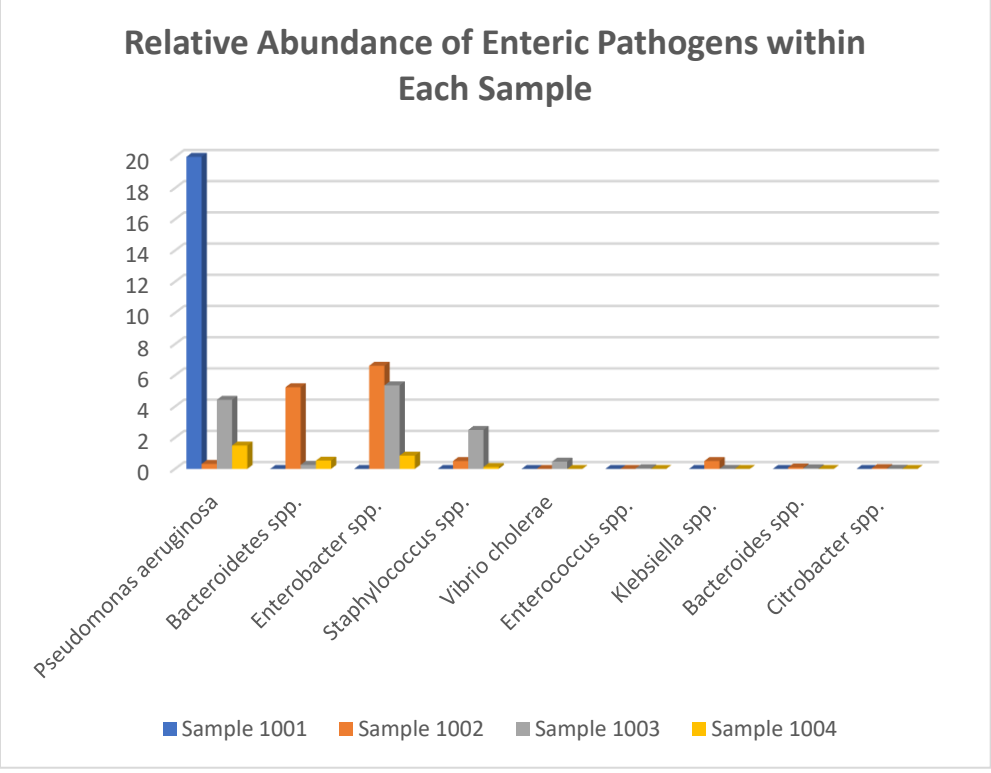


Figure 12: Relative Abundance of Enteric Pathogens

Observing the virulence factors and antibiotic resistance gene heat maps (Figures 13 and 14), it can be observed that all samples contain both. Samples 1003 and 1004 contain the highest numbers of virulence factors, and 1003 contain substantial antibiotic resistance genes. Antibiotic resistance genes do not exist if antibiotic treatments have not been utilized to treat infection, and virulence factors indicate the potential of a given microorganism to infect humans. This data therefore demonstrates that within all the samples, there exist pathogens with human sources that have the potential to infect humans. In summary, through next generation sequencing, we can discern the origin and infection potential of species found in samples, as opposed to traditional culture-based techniques that provide no such clarity.

To provide an example, a virulence factor for *Salmonella Typhimurium* (*S. Typhimurium*) was identified in sample 1003. According to Gart et al. (2016), *S. Typhimurium* is a foodborne pathogen that alters the gastrointestinal environment. It competes with existing microbes for nutrient needs and overcomes resistance to colonization, making it a potent infection risk. It also uses its own unique chemical signaling methods which can impact or even regulate host hormone metabolism (Gart et al. 2016). This data provides valuable information on the presence of an organism that is not commonly monitored, but that has the potential to alter the human gastrointestinal environment.

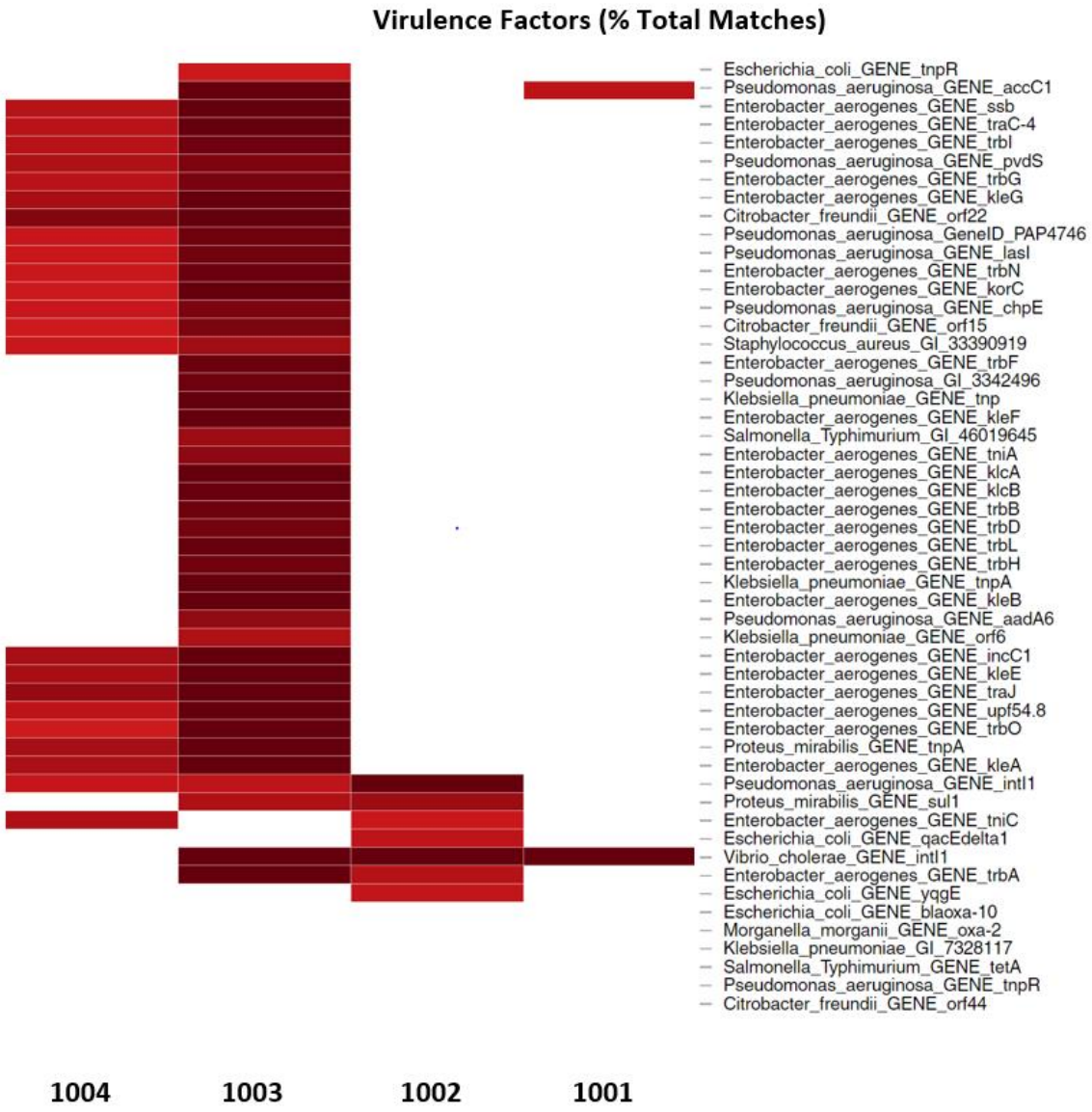


Figure 13: Virulence Factors by Sample

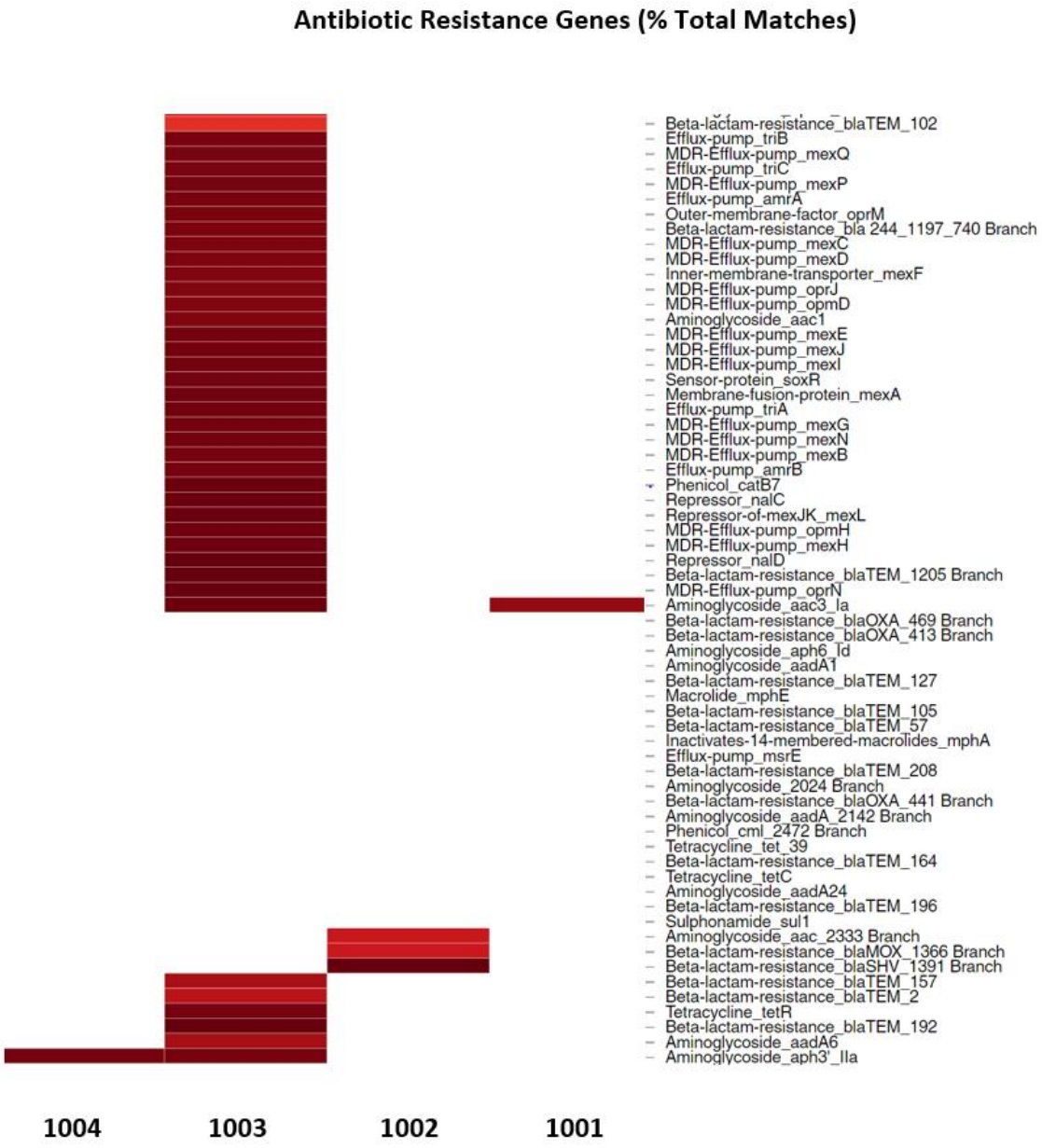


Figure 14: Antibiotic Resistance Genes by Sample

Next generation sequencing techniques also allow for valuable diversity calculations. Data from traditional culture-based techniques do not provide enough information to carry out the same types of calculations, as only a small number of bacterial species are cultured. Figure 15 shows the Alpha Diversity of each sample. Within all 4 samples, it is apparent that bacterial diversity is greatest, and that while viruses, fungi, and protists are present, they represent a much smaller contribution to overall diversity. Bacteria, therefore, are of greatest concern in these samples. Between samples, bacteria contribute to diversity in varying degrees; but regardless, in each sample they represent the greatest contributor to overall diversity.

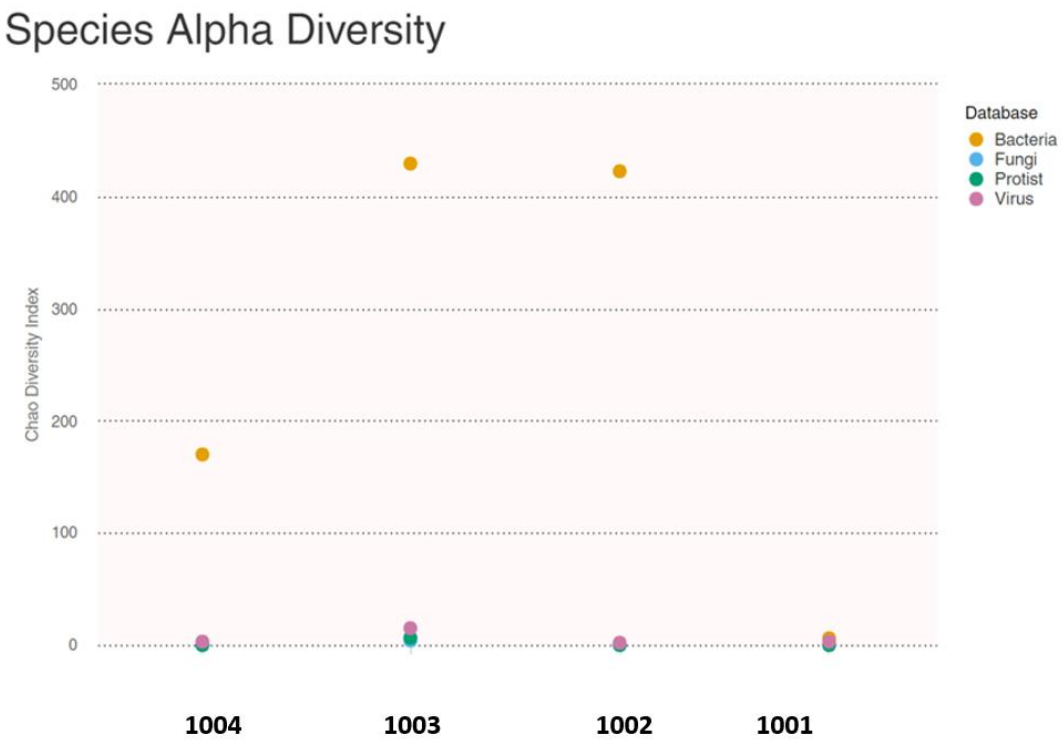


Figure 15: Species Alpha Diversity by Sample

Conclusions

The detection and monitoring of pathogenic species in coastal water samples is essential for the protection of human health. The same culture-based methods to detect fecal indicator bacteria have been in use for decades and provide researchers and water

quality managers with incredibly limited data. With these methods, the full spectrum of bacterial, viral, and fungal species cannot be identified, nor can sources be directly tracked to determine human versus environmental origin. More modern sequencing-based techniques are proving their worth and gaining the interest of researchers worldwide, and their uses are expanding rapidly. In the coastal zone management context, DNA sequencing, RNA sequencing, 16s rDNA sequencing, and qPCR for biomarker methods have all been used with great success. The method that seems to show the most promise due to the scope and wealth of information it provides is whole genome shotgun sequencing paired with k-mer based identification of species. Illumina and ThermoFisher both offer excellent sequencing platforms for shotgun sequencing, and sequences can be sent to COSMOS ID for species identification. This combination identifies bacteria, viruses, and fungi in samples, as well as virulence factors and antibiotic resistance genes, and is the only combination that will provide this wealth of data. In the past, managers had to rely on culture-based methods to detect the presence of fecal indicator bacteria and would base decisions from that data. If high levels of fecal indicator bacteria were present on a beach and the source was not readily known, the assumption was often made that the area posed a risk to human health whether or not this was the case. The bacteria could have originated from an environmental source and thus would not infect humans, but without this knowledge risks could not be taken. With modern techniques, these same managers can determine not only the whole scope of microorganisms present in a given sample, but also can infer source and risk to human health from antibiotic resistance genes and virulence factors. Quantitative analysis of data in this capstone showed a dramatic difference between traditional culture-based techniques and shotgun sequencing techniques, demonstrating the true value of considering modern methods. While the costs of sequencing techniques are significantly higher than culture-based techniques at present, costs continue to drop making them more accessible to would-be users. If users desire more information than can be obtained from culture-based techniques but cannot afford shotgun sequencing, an intermediate-cost method such as 16s rDNA sequencing can provide much more information for decision making Users must determine their organizational needs, scrutinize each method, analyze resources, and move forward with a procedure that best meets their end needs.

References

Aguiar-Pulido, V., W. R. Huang, V. Suarez-Ulloa, T. Cickovski, K. Mathee, and G. Narasimhan. 2016. "Metagenomics, Metatranscriptomics, and Metabolomics Approaches for Microbiome Analysis." *Evolutionary Bioinformatics* 12:5-16. doi: 10.4137/ebo.S36436.

Allen, L. Z., J. P. McCrow, K. Ininbergs, C. L. Dupont, J. H. Badger, J. M. Hoffman, M. Ekman, A. E. Allen, B. Bergman, and J. C. Venter. 2017. "The Baltic Sea Virome: Diversity and Transcriptional Activity of DNA and RNA Viruses." *Msystems* 2 (1):16. doi: 10.1128/mSystems.00125-16.

Alm, E. W., J. Burke, and A. Spain. 2003a. "Fecal indicator bacteria are abundant in wet sand at freshwater beaches." *Water Research* 37 (16):3978-3982. doi: 10.1016/s0043-1354(03)00301-4.

Alm, WA. , J. Burke, and A. Spain. 2003b. "Fecal indicator bacteria are abundant in wet sand at freshwater beaches." *Water Research* 37:3978-3982.

Altwegg, M. 1996. "Subtyping method for *Aeromonas* species." In *The Genus Aeromonas*, edited by B. Austin, M. Altwegg, P.J. Gosling and W.S. Joseph, 109-126. Chichester, US: John Wiley and Sons.

Aw W, Fukuda S. 2015. "Toward the comprehensive understanding of the gut ecosystem via metabolomics-based integrated omics approach." *Seminars in Immunopathology* 37 (1):5-16.

Barberan, A. , S.T. Bates, E.O. Casamayor, and N. Fierer. 2011. "Using network analysis to explore cooccurrence patterns in soil microbial communities." *The ISME Journal* 6 (2):343-351.

BaseSpace Sequence Hub Apps Quick Guide. 2017. edited by Inc. Illumina.

Basic Local Alignment Search Tool. National Institute of Health.

Bassler BL, Greenberg EP, Stevens AM. 1997. "Cross-species induction of luminescence in the quorum-sensing bacterium *Vibrio harveyi*." *Journal of Bacteriology* 179 (12):1943-5.

Bernini P, Bertini I, Luchinat C, et al. 2009. "Individual human phenotypes in metabolic space and time." *Journal of Proteome Research* 8 (9):4264-71.

BL, Bassler. 2002. "Small talk: cell-to-cell communication in bacteria." *Cell* 109 (4):421-4.

Boehm, A. B., K. M. Yamahara, and L. M. Sassoubre. 2014. "Diversity and Transport of Microorganisms in Intertidal Sands of the California Coast." *Applied and Environmental Microbiology* 80 (13):3943-3951. doi: 10.1128/aem.00513-14.

Bonilla, T. D., K. Nowosielski, M. Cuvelier, A. Hartz, M. Green, N. Esiobu, D. S. McCorquodale, J. M. Fleisher, and A. Rogerson. 2007. "Prevalence and distribution of fecal indicator organisms in South Florida beach sand and preliminary assessment of health effects associated with beach sand exposure." *Marine Pollution Bulletin* 54 (9):1472-1482. doi: 10.1016/j.marpolbul.2007.04.016.

Bonilla, T. D., K. Nowosielski, N. Esiobu, D. S. McCorquodale, and A. Rogerson. 2006. "Species assemblages of Enterococcus indicate potential sources of fecal bacteria at a south Florida recreational beach." *Marine Pollution Bulletin* 52 (7):807-10. doi: 10.1016/j.marpolbul.2006.03.004.

Bouzat, J. L., M. J. Hoostal, and T. Looft. 2013. "Spatial patterns of bacterial community composition within Lake Erie sediments." *Journal of Great Lakes Research* 39 (2):344-351. doi: 10.1016/j.jglr.2013.03.003.

Brandi, G., M. Sisti, G. Schiavano, L. Salvaggio, and A. Albano. 1996. "Survival of *Aeromonas hydrophila, Aeromonas caviae,* and *Aeromonas sobria* in soil." *Journal of Applied Bacteriology* 81:439-444.

Camilli A, Bassler BL. 2006. "Bacterial small-molecule signaling pathways." *Science* 311 (5764):1113-6.

Caporaso, J.G., C.L. Lauber, W.A. Walters, and et al. 2012. "Ultra-high throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms." *ISME Journal* 6 (8):1621-4.

Carvalhais LC, Dennis PG, Tyson GW, and Schenk PM. 2012. "Application of metatranscriptomics to soil environments." *Journal of Microbial Methods* 91 (2):246-51.

Chaffron, S., H. Rehrauer, J. Pernthaler, and C. von Mering. 2010. "A global network of coexisting microbes from environmental and whole-genome sequence data." *Genome Research* 20 (7):947-959. doi: 10.1101/gr.104521.109.

Chain, P.S.G., D.V. Grafham, R.S. Fulton, and et al. 2009. "Genome project standards in a new era of sequencing." *Science* 326 (326):236-237.

Chen, K., and L. Pachter. 2005. "Bioinformatics for whole-genome shotgun sequencing of microbial communities." *Plos Computational Biology* 1 (2):106-112. doi: 10.1371/journal.pcbi.0010024.

"CosmosID Plugin for CLC." QIAGEN. https://www.qiagenbioinformatics.com/plugins/cosmosid/.

Cui, H. L., K. Yang, E. Pagaling, and T. Yan. 2013. "Spatial and Temporal Variation in Enterococcal Abundance and Its Relationship to the Microbial Community in Hawaii Beach Sand and Water." *Applied and Environmental Microbiology* 79 (12):3601-3609. doi: 10.1128/aem.00135-13.

Culley, A. I., A.S. Lang, and C.A. Suttle. 2006. "Metagenomic Analysis of Coastal RNA Virus Communities." *Science* 312 (5781):1795--1798.

Deetz, T.R., E.M. Smith, S.M. Goyal, and et al. 1984. "Occurrence of rota-and enterovirus in drinking water and environmental water in a developing nation." *Water Research* 18:567-571.

Denoeud, F., and et al. 2008. "Annotating genomes with massive-scale RNA sequencing." *Genome Biology* 9 (R175).

Desmarais, T.R., H. M. Solo-Gabriele, and C.J. Palmer. 2002. "Influence of soil on fecal indicator organisms in a tidally influenced subtropical environment." *Applied and Environmental Microbiology* 68:1165-1172.

"DNeasy Power Water Kit." QIAGEN. https://www.qiagen.com/us/shop/sample-technologies/dna/genomic-dna/dneasy-powerwater-kit/#orderinginformation.

Duran-Pinedo AE, Chen T, Teles R, et al. 2014. "Community-wide transcriptome of the oral microbiome in subjects with and without periodontitis." *ISME Journal* 8 (8):1659-72.

Dutken, N.M. 2018a. "ThermoFisher Cost Per Sample Analysis Spreadsheet."
Dutken, N.M. 2018b. "Virtual Meeting to Discuss ThermoFisher Products for NGS." 4 April 2018.

Environmental Protection Agency (EPA). "Approved CWA Microbiological Test Methods." https://www.epa.gov/cwa-methods/other-clean-water-act-test-methods-microbiological.

Environmental Protection Agency (EPA). "Other Clean Water Act Test Methods: Microbiological." accessed 2018. https://www.epa.gov/cwa-methods/other-clean-water-act-test-methods-microbiological.

Environmental Protection Agency (EPA). 2001. Method 1605: Aeromonas in finished water by membrane filtration using Ampicillin-Dextrin agar with Vancomycin (ADA-V). edited by Office of Water United States Environmental Protection Agency. Washington, DC.

Falkowski, P.G., and C. de Vargas. 2004. "Shotgun Sequencing in the Sea: A Blast from the Past?" *Science* 304 (5667):58-60. doi: 10.1126/science.1097146.

Faust, K., J. F. Sathirapongsasuti, J. Izard, and et al. 2012. "Microbial co-occurrence relationships in the human microbiome." *Plos Computational Biology* 8 (7).

Fiehn, O. 2002. "Metabolomics-the link between genotypes and phenotypes." *Plant Molecular Biology* 48 (1):155-71.

Field, K.G., A.E. Bernhard, and T.J. Brodeur. 2003. "Molecular approaches to microbiological monitoring fecal source detection." *Environmental Monitoring and Assessment* 81:313-326.

Freilich, S., A. Kreimer, I. Meilijson, U. Gophna, R. Sharan, and E. Ruppin. 2010. "The large-scale organization of the bacterial network of ecological co-occurrence interactions." *Nucleic Acids Research* 38 (12):3857-3868. doi: 10.1093/nar/gkq118.

Frias-Lopez, J., Y. Shi, G.W. Tyson, and et al. 2008. "Microbial community gene expression in ocean surface waters." *PNAS* 105 (10):3805-3810.

Gart, Elena V., Jan S. Suchodolski, Thomas H. Welsh, Robert C. Alaniz, Ronald D. Randel, and Sara D. Lawhon. 2016. "Salmonella Typhimurium and Multidirectional Communication in the Gut." *Frontiers in Microbiology* 7 (1827). doi: 10.3389/fmicb.2016.01827.

"GENIUS Metagenomics: Know Now." 2018. Illumina. https://www.illumina.com/products/by-type/informatics-products/basespace-sequence-hub/apps/cosmosid-genius-metagenomics-know-now.html.

Gerba, C.P., and J.B. Rose. 1990. "Viruses in source and drinking water." In *Drinking Water Microbiology*, edited by G.A. McFeters, 380-395. New York: Springer-Verlag.

Gilbert, J. A, D Field, Y Huang, and et al. 2008. "Detection of large numbers of novel sequences in the metatranscriptomes of complex marine microbial communities." *PLoS One* 3 (8).

Gilbert, J. A., J. K. Jansson, and R. Knight. 2014. "The Earth Microbiome project: successes and aspirations." *Bmc Biology* 12:4. doi: 10.1186/s12915-014-0069-1.

Gofton, A.W., C.L. Oskam, N. Lo, and et al. 2015. "Inhibition of the endosymbiont "Candidatus Midichloria mitochondrii" during 16S rRNA gene profiling reveals potential pathogens in Ixodes ticks from Australia." *Parasites and Vectors* 8 (345).

Gonzalez, A., and R. Knight. 2012. "Advancing analytical algorithms and pipelines for billions of microbial sequences." *Current Opinion in Biotechnology* 23 (1):64-71. doi: 10.1016/j.copbio.2011.11.028.

"GreenGenes: The 16S rRNA Gene Database and Tools." http://greengenes.secondgenome.com/downloads.

Griffin, D.W., K.A. Donaldson, J.H. Paul, and J.B. Rose. 2003. "Pathogenic human viruses in coastal waters." *Clinical Microbiology Reviews* 16:129-143.

Hamandy, M., C. Lozupone, and R. Knight. 2010. "Fast UniFrac: facilitating high-throughput phylogenetic analyses of microbial communities including analysis of pyrosequencing and PhyloChip data." *ISME Journal* 4:17-27.

Hanninen, M.L., and A. Siitonen. 1995. "Distribution of *Aeromonas* phenospecies and genospecies among strains isolated from water, foods, or from human clinical samples." *Epidemiology and Infection* 115:39-50.

Havelaar, A.H., M. During, and J.F.M. Versteegh. 1987. "Ampicillin-dextrin agar medium for the enumeration of *Aeromonas* species in water by membrane filtration " *Journal of Applied Bacteriology* 62:279-287.

Henke JM, Bassler BL. 2004. "Three parallel quorum-sensing systems regulate gene expression in *Vibrio harveyi*." *Journal of Bacteriology* 186 (20):6902-14.

Holmes, P., and D.P. Sartory. 1993. "An evaluation of media for the membrane filtration enumeration of *Aeromonas* from drinking water." *Letters in Applied Microbiology* 17:15-68.

Horman, A., R. Rimhanen-Finne, L. Maunula, C. von Bonsdorff, and et al. 2004. "Campylobacter spp., Giardia spp., Cryptosporidium spp., Norovirus, and indicator organisms in surface water in southwestern Finland, 2000-2001." *Applied and Environmental Microbiology* 70:87-95.

Hurst, C.J., R.L. Crawford, G.R. Knudsen, M.J. McInerney, and L.D. Stenzenbach. 2002. *Manual of Environmental Microbiology, second ed.* Washington, DC: ASM Press.

Huson, D. H., A. F. Auch, J. Qi, and S. C. Schuster. 2007. "MEGAN analysis of metagenomic data." *Genome Research* 17 (3):377-386. doi: 10.1101/gr.5969107.

"Ion 16S Metagenomics Kit and Ion Reporter metaenomics workflow solution." ThermoFisher Scientific. https://www.thermofisher.com/us/en/home/life-science/sequencing/dna-sequencing/microbial-sequencing/microbial-identification-ion-torrent-next-generation-sequencing/ion-16s-metagenomics-solution.html.

"Ion 520 Chip Kit." ThermoFisher Scientific. https://www.thermofisher.com/order/catalog/product/A27762?SID=srch-srp-A27762.

"Ion 530 Chip Kit." ThermoFisher Scientific. https://www.thermofisher.com/order/catalog/product/A27764.

"Ion 540 Chip Kit." ThermoFisher Scientific. https://www.thermofisher.com/order/catalog/product/A27765?SID=srch-srp-A27765.

Ion Gene Studio S5 Series Specification Sheet. 2018. edited by ThermoFisher Scientific.

"Ion Torrent Next-Generation Sequencing Technology." Thermo Fisher Scientific. https://www.thermofisher.com/us/en/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-technology.html.

"Ion Total RNA-Seq Kit v2." ThermoFisher Scientific. https://www.thermofisher.com/order/catalog/product/4475936.

"Ion Xpress Plus Fragment Library Kit." ThermoFisher Scientific. https://www.thermofisher.com/order/catalog/product/4471269.

Ion Xpress Plus gDNA Fragment Library Preparation User Guide. edited by Thermo Fisher Scientific.

Introducing the Ion Torrent GeneStudio S5 series for next-generation sequencing. 2018. edited by ThermoFisher Scientific.

Isonhood, J.H., and M. Drake. 2002. "*Aeromonas* apecies in foods." *Journal of Food Protection* 65:575-582.

Janda, J.M., and S.L. Abbot. 1998. "Evolving concepts regarding the genus *Aeromonas*: an expanding panorama of species, disease presentations, and unanswered questions." *Clinical Infections Diseases* 27:332-344.

Jiang, S.C., and W. Chu. 2004. "PCR detection of pathogenic viruses in southern California urban rivers." *Applied and Environmental Microbiology* 97:17-28.

Jiang, S.C., R. Noble, and W.P. Chui. 2001. "Human adenovirus and coliphages in urban runoff-impacted coastal waters of Southern California." *Applied and Environmental Microbiology* 67:179-184.

Kaddurah-Daouk R, Baillie RA, Zhu H, et al. 2011. "Enteric microbiome metabolites correlate with response to simvastatin treatment." *PLoS One* 6 (10).

Kapronev, P., A.T. Willingham, and T.R. Gingeras. 2007. "Genome-wide transcription and the implications for genomic organization." *8* (413-323).

Karsenti, E., S. G. Acinas, P. Bork, C. Bowler, C. De Vargas, J. Raes, M. Sullivan, D. Arendt, F. Benzoni, J. M. Claverie, M. Follows, G. Gorsky, P. Hingamp, D. Iudicone, O. Jaillon, S. Kandels-Lewis, U. Krzic, F. Not, H. Ogata, S. Pesant, E. G. Reynaud, C. Sardet, M. E. Sieracki, S. Speich, D. Velayoudon, J. Weissenbach, P. Wincker, and Consortium Tara Oceans. 2011. "A Holistic Approach to Marine Eco-Systems Biology." *Plos Biology* 9 (10):5. doi: 10.1371/journal.pbio.1001177.

Kermarrec, L. , A. Franc, F. Rimet, and et al. 2013. "Next-generation sequencing to inventory taxonomic diversity in eukaryotic communities: a test for freshwater diatoms." *Molecular Ecology Resources* 2013 (13):4.

Khan, I., V. Gannon, R. Kent, W. Koning, and et al. 2007. "Development of a rapid quantitative PCR assay for direct detection and quantification of culturable and non-culturable *Escherichia coli* from agricultural watersheds." *Journal of Microbial Methods* 69:480-488.

Khan, I., A. Loughborough, and T.A. Edge. 2009. "DNA-based real-time detection and quantification of aeromonads from fresh water beaches on Lake Ontario." *Journal of Water and Health* 7 (2):312-323.

Kim, K., and J. Bae. 2011. "Amplification Methods Bias Metagenomic Libraries of Uncultured Single-Stranded and Double-Stranded DNA Viruses." *Applied and Environmental Microbiology* 77 (21):7663-7668.

Kimes, NE, AV Callahan, DF Aktas, and et al. 2013. "Metagenomic analysis and metabolite profiling of deep-sea sediments from the Gulf of Mexico following the deep-water horizon oil spill." *Frontiers of Microbiology* 4.

Klindworth, A., E. Pruesse, T. Schweer, and et al. 2013. "Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity species." *Nucleic Acids Research* 41 (1).

Koeth RA, Wang Z, Levison BS, et al. 2013. "Intestinal microbiota metabolism of l-carnitine, a nutrient in red meat, promotes atherosclerosis." *Nature Medicine* 19 (5):576-85.

Kollias, E. 2018. "Virtual Meeting to Discuss Illumina Options for NGS." 3 April 2018.

Krumsiek J, Mittelstrass K, Do KT, et al. 2015. "Gender-specific pathway differences in the human serum metabolome." *Metabolomics* 11 (6):1815-33.

"Kraken Metagenomics." 2018. Illumina, Inc. https://www.illumina.com/products/by-type/informatics-products/basespace-sequence-hub/apps/kraken-metagenomics.html.

Kuczynski, J., Z. Z. Liu, C. Lozupone, D. McDonald, N. Fierer, and R. Knight. 2010. "Microbial community resemblance methods differ in their ability to detect biologically relevant patterns." *Nature Methods* 7 (10):813-U67. doi: 10.1038/nmeth.1499.

Lankadurai BP, Nagato EG, Simpson MJ. 2013. "Environmental metabolomics: an emerging approach to study organism responses to environmental stressors." *Environmental Reviews* 21 (3):180-205.

Lee, E., M.S. Khurana, A.S. Whiteley, and et al. 2017. "Novel Primer Sets for Next Generation Sequencing-Based Analyses of Water Quality." *PLOS One* 12 (1).

Lima-Mendez, G., K. Faust, N. Henry, J. Decelle, S. Colin, F. Carcillo, S. Chaffron, J. C. Ignacio-Espinosa, S. Roux, F. Vincent, L. Bittner, Y. Darzi, J. Wang, S. Audic, L. Berline, G. Bontempi, A. M. Cabello, L. Coppola, F. M. Cornejo-Castillo, F. d'Ovidio, L. De Meester, I. Ferrera, M. J. Garet-Delmas, L. Guidi, E. Lara, S. Pesant, M. Royo-Llonch, G. Salazar, P. Sanchez, M. Sebastian, C. Souffreau, C. Dimier, M. Picheral, S. Searson, S. Kandels-Lewis, G. Gorsky, F. Not, H. Ogata, S. Speich, L. Stemmann, J. Weissenbach, P. Wincker, S. G. Acinas, S. Sunagawa, P. Bork, M. B. Sullivan, E. Karsenti, C. Bowler, C. de Vargas, J. Raes, S. G. Acinas, P. Bork, E. Boss, C. Bowler, C. De Vargas, M. Follows, G. Gorsky, N. Grimsley, P. Hingamp, D. Iudicone, O. Jaillon, S. Kandels-Lewis, L. Karp-Boss, E. Karsenti, U. Krzic, F. Not, H. Ogata, S. Pesant, J. Raes, E. G. Reynaud, C. Sardet, M. Sieracki, S. Speich, L. Stemmann, M. B. Sullivan, S. Sunagawa, D. Velayoudon, J. Weissenbach, P. Wincker, and Coordinators Tara Oceans. 2015. "Determinants of community structure in the global plankton interactome." *Science* 348 (6237):9. doi: 10.1126/science.1262073.

M.A., Moran. 2009. "Metatranscriptomics: eavesdropping on complex microbial communities." *Microbiome* 4 (7):329-34.

Manor, O, R Levy, and E Borenstein. 2014. "Mapping the inner workings of the microbiome: genomic-and metagenomic-based study of metabolism and metabolic interactions in the human microbiome." *Cell Metabolism* 20 (5):742-52.

Marioni, J.C., C.E. Mason, S.M. Mane, and et al. 2008. "RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays." *Genome Research* 18:1509-1517.

Mason OU, Hazen TC, Borglin S, et al. 2012. "Metagenome, metatranscriptome and single-cell sequencing reveal microbial response to deepwater horizon oil spill." *ISME Journal* 6 (9):1715-27.

Maurice CF, Haiser HJ, Turnbaugh PJ. 2013. "Xenobiotics shape the physiology and gene expression of the active human gut microbiome." *Cell* 152 (1):39-50.

McCorquodale, D. S., and G. Duncan. 2018. "COSMOS ID Data from 4 Samples."

McFeters, G.A. 1990. *Drinking Water Microbiology: Progress and Recent Developments*. Springer, New York.

McFeters, G.A., G.K. Bissonnette, J.J. Jezeski, C.A. Thompson, and D.G. Stuart. 1974. "Comparative survival of indicator bacteria and enteric pathogens in well water." *Applied and Environmental Microbiology* 27:823-829.

McNulty NP, Yatsunenko T, Hsiao A, et al. 2011. "The impact of a consortium of fermented milk strains on the gut microbiome of gnotobiotic mice and monozygotic twins." *Science of Translational Medicine* 3 (106).

Menchaca, A.C., D.K. Visi, O.F. Strey, and et al. 2013. "Preliminary Assessment of Microbiome Changes Following Blood-Feeding and Survivorship in the Amblyomma americanum Nymph-to-Adult Transition using Semiconductor Sequencing." *PLoS One* 8 (6).

"MetaPhlAn." 2018. Illumina. https://www.illumina.com/products/by-type/informatics-products/basespace-sequence-hub/apps/huttenhower-lab-hsph-metaphlan.html.

MicroSEQ Rapid Microbial Identification System. 2013. edited by Life Technologies Corporation.

Milani, C., G.A. Lugli, F. Turroni, and et al. 2014. "Evaluation of bifidobacterial community composition in the human gut by eans of a targeted amplicon sequencing (ITS) protocol." *FEMS Microbial Ecology* 90 (2):493-503.

Miller MB, Bassler BL. 2001. "Quorum sensing in bacteria." *Annual Review of Microbiology* 55 (1):165-99.

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

Molecular Testing to Monitor the Whole Microbial Community of Environmental Samples. edited by Inc. Florida-Spectrum Environmental Services. Fort Lauderdale, FL.

Mortazavi, A., B.A. Williams, K. McCue, L. Schaeffer, and B. Wold. 2008. "Mapping and quantifying mammalian transcriptomes by RNA-seq." *Nature Methods* 5 (621-628).

Nagalakshmi, U. , and et al. 2008. "The transcriptional landscape of the yeast genome defined by RNA sequencing." *Science* 320:1344-1349.

Nealson, K. H., and J. C. Venter. 2007. "Metagenomics and the global ocean survey: what's in it for us, and why should we care?" *Isme Journal* 1 (3):185-187. doi: 10.1038/ismej.2007.43.

Nelson, K.E., G.M. Weinstock, S.K. Highlander, and et al. 2010. "A catalog of references genomes from the human microbiome." *Science* 328:994-999.

"Nextera XT DNA Library Preparation Kit." Illumina. https://www.illumina.com/products/by-type/sequencing-kits/library-prep-kits/nextera-xt-dna.html.

NextSeq 550 Sequencing System. 2017. edited by Inc. Illumina.

Noble, R.T., and J.A. Fuhrman. 2005. "Enterovirus detected by reverse transcriptase polymerase chain reaction from the coastal waters of Santa Monica Bay, California: low correlation to bacterial indicator levels." *Hydrobiologia* 460:175-184.

Nocker, A., C.Y. Cheung, and A.K. Camper. 2006. "Comparison of propidium monoazide with ethidium monoazide for differentiation of live vs dead bacteria by selective removal of DNA from dead cells." *Journal of Microbial Methods* 67:310-320.

NovaSeq 6000 Sequencing System. 2018. edited by Inc. Illumina.

Ozsolak, F., A.R. Platt, D.R. Jones, and et al. 2009. "Direct RNA Sequencing." *Nature* 461:814-819.

Peano C, Pietrelli A, Consolandi C, et al. 2013. "An efficient rRNA removal method for RNA sequencing in GC-rich bacteria." *Microbial Informatics and Experimentation* 3 (1).

Pommepuy, M., M. Butin, A. Derrien, M. Gourmelon, R. Colwell, and M. Cormier. 1996. "Retention of enteropathogenicity by viable but nonculturable *Escherichia coli* exposed to seawater and sunlight." *Applied and Environmental Microbiology* 62:4621-4626.

Poretsky, R., L. M. Rodriguez-R, C. W. Luo, D. Tsementzi, and K. T. Konstantinidis. 2014. "Strengths and Limitations of 16S rRNA Gene Amplicon Sequencing in

Revealing Temporal Microbial Community Dynamics." *Plos One* 9 (4):12. doi: 10.1371/journal.pone.0093827.

"PowerLyser 24 Homogenizer." QIAGEN. https://www.qiagen.com/us/shop/automated-solutions/sample-disruption/powerlyzer-24-homogenizer/#orderinginformation.

"QIAcube." QIAGEN. https://www.qiagen.com/us/shop/automated-solutions/sample-preparation/qiacube/#orderinginformation.

"QIAseq FX DNA Library Kit." QIAGEN. https://www.qiagen.com/us/search/qiaseq-fx-dna-library-kit/?akamai-feo=off.

Quince, C., A. Lanzen, T.P. Curtis, R.J. Davenport, N. Hall, and et al. 2009. "Accurate determination of microbial diversity from 454 pyrosequencing data." *Nature Journal* 6:639-641.

Quince, C., A. Lanzen, R.J. Davenport, and P.J. Turnbaugh. 2011. "Removing Noise From Pyrosequenced Amplicons." *BMC Bioinformatics* 12.

Raes, J., K. U. Foerstner, and P. Bork. 2007. "Get the most out of your metagenome: computational analysis of environmental sequence data." *Current Opinion in Microbiology* 10 (5):490-498. doi: 10.1016/j.mib.2007.09.001.

Ranjan, R., A. Rani, A. Metwally, H. S. McGee, and D. L. Perkins. 2016. "Analysis of the microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing." *Biochemical and Biophysical Research Communications* 469 (4):967-977. doi: 10.1016/j.bbrc.2015.12.083.

Reigstad, CS, and PC Kashyap. 2013. "Beyond phylotyping: understanding the impact of gut microbiota on host biology." *Neurogastroenterology and Motility* 25 (5):358-72.

RNA Sequencing. 2016. In *ThermoFisher Scientific*.

"RNeasy PowerWater Kit." QIAGEN. https://www.qiagen.com/us/shop/sample-technologies/dna/genomic-dna/rneasy-powerwater-kit/#orderinginformation.

"Rotor-Gene Q." QIAGEN. https://www.qiagen.com/us/shop/automated-solutions/sample-disruption/powerlyzer-24-homogenizer/#orderinginformation.

Santo Domingo, J.W., S.C. Siefring, and R.A. Haugland. 2003. "Real-time PCR method to detect Enterococcus faecalis in water." *Biotechnology Letters* 25:261-265.

Savichtcheva, O., N. Okayama, T. Ito, and S. Okabe. 2005. "Application of a direct fluorescent-based live/dead staining combined with fluorescent in situ

hybridization for assessment of survival rate of *Bacteroides* spp. in oxygenated water." *Biotechnology and Bioengineering* 92:356-363.

Schellhaas, R. 2018. "Illumina Cost Per Sample Analysis Spreadsheet."

Scott, T., J.B. Rose, T.M. Jenkens, S.R. Farrah, and J. Lukasik. 2002. "Microbial source tracking: current methodology and future directions." *Applied and Environmental Microbiology* 68:5796-5803.

Sharpton, T. J. 2014. "An introduction to the analysis of shotgun metagenomic data." *Frontiers in Plant Science* 5:14. doi: 10.3389/fpls.2014.00209.

Simpson, J.M., J.W. Santo Domingo, and D.J. Reasoner. 2002. "Microbial Source Tracking: state of science." *Environmental Science and Technology* 24:5279-5288.

Sinclair, L. , O.A. Osman, S. Bertilsson, and A. Eiler. "Microbial Community Composition and Diversity via 16S rRNA Gene Amplicons: Evaluating the Illumina Platform." *PLoS One* 10 (2).

Sinigalliano, C.D., J.M. Fleisher, M.L. Gidley, and et al. 2010. "Traditional and molecular analyses for fecal indicator bacteria in non-point source subtropical recreational marine waters." *Water Research* 44:3763-3772.

Solo-Gabriele, H. M., V. J. Harwood, D. Kay, R. S. Fujioka, M. J. Sadowsky, R. L. Whitman, A. Wither, M. Canica, R. C. Da Fonseca, A. Duarte, T. A. Edge, M. J. Gargate, N. Gunde-Cimerman, F. Hagen, S. L. McLellan, A. N. Da Silva, M. N. Babic, S. Prada, R. Rodrigues, D. Romao, R. Sabino, R. A. Samson, E. Segal, C. Staley, H. D. Taylor, C. Verissimo, C. Viegas, H. Barroso, and J. C. Brandao. 2016. "Beach sand and the potential for infectious disease transmission: observations and recommendations." *Journal of the Marine Biological Association of the United Kingdom* 96 (1):101-120. doi: 10.1017/s0025315415000843.

Sultan, M., and et al. 2008. "A global view of gene activity and alternative spliciing by deep sequencing of the human transcriptome." *Science* 321:956-960.

Sunagawa, S., L. P. Coelho, S. Chaffron, J. R. Kultima, K. Labadie, G. Salazar, B. Djahanschiri, G. Zeller, D. R. Mende, A. Alberti, F. M. Cornejo-Castillo, P. I. Costea, C. Cruaud, F. d'Ovidio, S. Engelen, I. Ferrera, J. M. Gasol, L. Guidi, F. Hildebrand, F. Kokoszka, C. Lepoivre, G. Lima-Mendez, J. Poulain, B. T. Poulos, M. Royo-Llonch, H. Sarmento, S. Vieira-Silva, C. Dimier, M. Picheral, S. Searson, S. Kandels-Lewis, C. Bowler, C. de Vargas, G. Gorsky, N. Grimsley, P. Hingamp, D. Iudicone, O. Jaillon, F. Not, H. Ogata, S. Pesant, S. Speich, L. Stemmann, M. B. Sullivan, J. Weissenbach, P. Wincker, E. Karsenti, J. Raes, S. G. Acinas, P. Bork, and Coordinators Tara Oceans. 2015. "Structure and function

of the global ocean microbiome." *Science* 348 (6237):9. doi: 10.1126/science.1261359.

Takahashi, S., J. Tomita, K. Nishioka, and et al. 2014. "Development of a Prokaryotic Universal Primer for Simultaneous Analysis of Bacteria and Archaea Using Next-Generation Sequencing." *PLoS ONE* 9 (8).

Thermo Fisher Scientific. 2015. GeneStudio Metagenomics Introduction

Thermo Fisher Scientific. "MicroSEQ ID Microbial Identification Software." https://www.thermofisher.com/us/en/home/life-science/bioproduction/microbial-identification-and-detection/microseq--id-microbial.html.

TruSeq RNA and DNA Library Preparation Kits v2. 2014. edited by Illumina.

Venter, J. C., K. Remington, J.F. Heidelberg, and et al. 2004. "Environmental Genome Shotgun Sequencing of the Sargasso Sea." *Science* 304 (5667):66-74.

Verberkmoes NC, Russell AL, Shah M, et al. 2009. "Shotgun metaproteomics of the human distal gut microbiota." *ISME Journal* 3 (2):179-89.

Vincent, A. T., N. Derome, B. Boyle, A. I. Culley, and S. J. Charette. 2017. "Next-generation sequencing (NGS) in the microbiological world: How to make the most of your money." *Journal of Microbiological Methods* 138:60-71. doi: 10.1016/j.mimet.2016.02.016.

Wang Z, Klipfell E, Bennett BJ, et al. 2011. "Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease." *Nature* 472 (7341):57-63.

Waters CM, Bassler BL. 2005. "Quorum sensing: cell-to-cell communication in bacteria." *Annual Review of Cell Development Biology* 21:319-46.

Weir TL, Manter DK, Sheflin AM, Barnett BA, Heuberger AL, Ryan EP. 2013. "Stool microbiome and metabolome differences between colorectal cancer patients and healthy adults." *PLoS One* 8 (8).

Whipps, J.M., K. Lewis, and R.C. Cooke. 1988. *Mycoparasitism and Plant Disease Control*. Manchester, UK: Manchester University Press.

Whiteley, A.S., S. Jenkens, I. Waite, N. Kresoje, and et al. 2012. "Microbial 16S rRNA Ion Tag and community metagenome sequencing using the Ion Torrent (PGM) Platform." *Journal of Microbial Methods* 91 (1):80-88.

Whitman, R. , V. J. Harwood, T. A. Edge, and et al. 2015. "Microbes in Beach Sands: Integrating Environment , Ecology, and Public Health." *Reviews in Environmental Seience and Bio/Technology* 13 (3):329-368.

Wilhelm, B.T., and et al. 2008. "Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution." *Nature* 453:1239-1243.

Winfield, M.D., and E.A. Groisman. 2003. "Role of nonhost environmental in the lifestyle of Salmonella and E. coli." *Applied and Environmental Microbiology* 69:3687-3694.

Wright, T. . 2018. "Email to Discuss Pricing on QIAGEN NGS Products."

Wu, GD, C Compher, EZ Chen, and et al. 2014. "Comparitive metabolomics in vegans and omnivores revelas constraints on diet-dependent gut microbiota metabolite production." *Gut* 65 (1):63-72.

Xu, J, MA Mahowald, RE Ley, and et al. 2007. "Evolution of symbiotic bacteria in the distal human intestine." *PLoS Biology* 5 (7).