PROJECT NO.
4961

The Use of Next Generation Sequencing (NGS) Technologies and Metagenomics Approaches to Evaluate Water and Wastewater Quality Monitoring and Treatment Technologies

CALIFORNIA
Water Boards
STATE WATER RESOURCES CONTROL BOARD
REGIONAL WATER QUALITY CONTROL BOARDS

The Use of Next Generation Sequencing (NGS) Technologies and Metagenomics Approaches to Evaluate Water and Wastewater Quality Monitoring and Treatment Technologies

Prepared by:

Emily Garner
West Virginia University

Matthew Forrest Blair, Connor Brown, Abraham Cullom, Benjamin C Davis, Sudeshna Ghosh, Mariah Gnegy, Suraj Gupta, Ishi Keenum, Krista Liguori, Ayella Maile-Moskowitz, Erin Milligan, Jin Pan, Aaron J Prussin II, Katie Scott, Lenwood S. Heath, Linsey C. Marr, Peter J. Vikesland, Liqing Zhang, Amy Pruden
Virginia Tech

Co-sponsored by:
California State Water Resources Control Board

2023

THE Water Research FOUNDATION

The Water Research Foundation (WRF) is the leading research organization advancing the science of all water to meet the evolving needs of its subscribers and the water sector. WRF is a 501(c)(3) nonprofit, educational organization that funds, manages, and publishes research on the technology, operation, and management of drinking water, wastewater, reuse, and stormwater systems—all in pursuit of ensuring water quality and improving water services to the public.

For more information, contact:
The Water Research Foundation

1199 North Fairfax Street, Suite 900
Alexandria, VA 22314-1445
P 571-384-2100

6666 West Quincy Avenue
Denver, Colorado 80235-3098
P 303-347-6100

www.waterrf.org
info@waterrf.org

WRF ISBN: 978-1-60573-638-9
WRF Project Number: 4961

This report was prepared by the organization(s) named below as an account of work sponsored by The Water Research Foundation. Neither The Water Research Foundation, members of The Water Research Foundation, the organization(s) named below, nor any person acting on their behalf: (a) makes any warranty, express or implied, with respect to the use of any information, apparatus, method, or process disclosed in this report or that such use may not infringe on privately owned rights; or (b) assumes any liabilities with respect to the use of, or for damages resulting from the use of, any information, apparatus, method, or process disclosed in this report.

Prepared by Virginia Tech and West Virginia University. All figures and tables indicated by a citation are re-printed under the terms of the Creative Commons Attribution-Non Commercial License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

This document was reviewed by a panel of independent experts selected by The Water Research Foundation. Mention of trade names or commercial products or services does not constitute endorsement or recommendations for use. Similarly, omission of products or trade names indicates nothing concerning The Water Research Foundation's positions regarding product effectiveness or applicability.

Acknowledgments

The authors are grateful to the organizations and experts who contributed information, opinions, time, data, and samples to this project.

Research Team

Principal Investigators:
Amy Pruden, PhD
Peter J. Vikesland, PhD
Linsey C. Marr, PhD
Liqing Zhang, PhD
Lenwood S. Heath, PhD
Virginia Tech

Emily Garner, PhD
West Virginia University

Project Team:
Matthew Forrest Blair, PhD
Connor Brown
Abraham Cullom
Benjamin C. Davis, PhD
Joseph O. Falkinham, III, PhD
Sudeshna Ghosh, PhD
Mariah Gnegy
Suraj Gupta, PhD
Ishi Keenum, PhD
Krista Liguori, MSPH
Ayella Maile-Moskowitz, PhD
Erin Milligan
Jin Pan
Aaron J. Prussin II, PhD
Maria Virginia Riquelme, PhD
Katie Scott
Ni Zhu
Virginia Tech

Utility Partners:
Flagstaff Water Services
Hampton Roads Sanitation District
Metropolitan Water Reclamation District of Greater Chicago
Orange County Water District
Southern Nevada Water Authority

WRF Project Subcommittee or Other Contributors

Vicente Gomez-Alvarez
US EPA

Bina Nayak
Pinellas County Utilities

Yale J Passamaneck
Bureau of Reclamation

Cindy Figueroa
State Water Resources Control Board

WRF Staff

John Albert, MPA
Chief Research Officer

H. Grace Jang, PhD
Research Program Manager

Contents

Tables

Figures

Acronyms and Abbreviations

ACLAME	A CLAssification of Mobile genetic Elements
AMR	Antimicrobial resistance
ANAMMOX	Anaerobic ammonia oxidation
ARB	Antibiotic resistant bacteria
ARG	Antibiotic resistance gene
ASV	Amplicon sequence variant
BACMET	Antibacterial Biocide & Metal Resistance Database
BLAST	Basic Local Alignment Search Tool
bp	Base pair
CARD	Comprehensive Antibiotic Resistance Database
CARD	Comprehensive Antibiotic Resistance Database
CAZy	Carbohydrate-Active Enzymes Database
CDC	Centers for Disease Control and Prevention
cDNA	Complementary deoxyribonucleic acid
CLI	Command line interface
COG	Clusters of Orthologous Groups of proteins
COMAMMOX	Complete ammonia oxidation
CTAB	Cetyltrimethylammonium bromide
CV	Coefficient of variation
dBgs	De Bruijn Graphs
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphates
DWDS	Drinking water distribution system
exDNA	Extracellular deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
EPA	Environmental Protection Agency
ESBL	Extended-spectrum beta-lactamase
FARME-DB	Functional Antibiotic Resistance Metagenomic Element Database
FISH	Fluorescence in situ hybridization
GAC	Granular activated carbon
Gb	Giga-base pairs
gc	Gene copies
GTDB-tk	Genome Taxonomy Database Toolkit
HMW	High molecular weight
ITS	Internal transcribed spacer
KEGG	Kyoto Encyclopedia of Genes and Genomes
LMW	Low molecular weight

LOD	Limit of detection
LOQ	Limit of quantification
MAG	Metagenomic-assembled genomes
MGE	Mobile genetic elements
MixS	Minimum Information about any (x) Sequence
MLS	Macrolide-lincosamide-streptogramin
MNP	Magnetic nanoparticles
MPD	MyPathogen Database
MRG	Metal resistance genes
mRNA	Messenger ribonucleic acid
MRT	Microbial residence time
NCBI	National Center for Biotechnology Information
NGS	Next generation sequencing
NMDS	Non-metric multidimensional scaling
NTM	Non-tuberculosis mycobacteria
OP	Opportunistic pathogen
OTU	Operational taxonomic unit
PCR	Polymerase chain reaction
PPE	Personal protective equipment
QC	Quality control
qMeta	Quantitative metagenomics
QMRA	Quantitative microbial risk assessment
qPCR	Quantitative polymerase chain reaction
RDP	Ribosomal database project
RNA	Ribonucleic acid
rRNA	Ribosomal ribonucleic acid
RWDS	Reclaimed water distribution systems
SDS	Sodium dodecyl sulfate
SNP	Single nucleotide polymorphism
SOP	Standard operating procedure
VBNC	Viable, but non-culturable
WGS	Whole genome sequencing
WHO	World Health Organization
WWTP	Wastewater treatment plant

Executive Summary

ES.I Key Findings

- This report intends to help the water industry understand, apply, and benefit from next generation sequencing (NGS). This guidance was informed by a comprehensive literature review, as well as interviews with water utility staff.
- NGS technologies are now widely applied to identify and quantify microbes in wastewater, drinking water, recycled water, and surface water environments.
- In the scientific literature, a vast array of methods is applied in the preparation of samples for NGS and also in the analysis of NGS data, calling for uniformity and standardization when possible, to improve comparability of the data.
- NGS technologies are beginning to be used by larger water utilities to provide information about pathogens and organisms that play a role in processes of interest (e.g., nitrification). This document provides a framework to help align the choice of methods and analyses to learn more about these pathogens and organisms.
- Case studies are provided to demonstrate key applications (profiling of pathogens, antibiotic resistance, assessment of viral risks, and key microbial functions of interest). A validation study was conducted to assess DNA extraction methods, sequencing depths, and overall quantitative capacity of NGS.

ES.2 Background and Objectives

The emergence of next generation sequencing (NGS) technologies is revolutionizing the use of molecular techniques for understanding complex microbial communities. NGS is poised to address key issues of importance to the water industry by bringing new understanding to various dimensions of water quality including antibiotic resistance, pathogen occurrence, functional capacities of microbial communities, contaminant biodegradation, virus occurrence, and many others. However, application of NGS in the water sector has been somewhat limited due to cost, need for specialized expertise and equipment, challenges with data analysis and interpretation, lack of standardized methods, and the rapid pace of new technological developments. Given the immense potential of NGS, effort is needed to overcome these obstacles and make NGS technologies accessible to water professionals including utility staff, consultants, researchers, and regulators. As NGS technologies are beginning to be applied within the water industry, there is a need for guidance not only to facilitate and expand their application, but also to ensure that resulting interpretation is meaningful and reliable. The overarching objective of this project is to advance the application of NGS in the water industry, first by identifying key opportunities for their application, assessing barriers to their implementation, proposing and validating approaches to overcome these barriers, and developing a comprehensive guidance document to educate water professionals and aid in navigating options for their application.

ES.3 Project Approach

Six key tasks were completed to support the objective described above:

Task 1: Conduct a comprehensive literature review detailing existing and emerging NGS approaches relevant to the water industry. The findings of this comprehensive literature review are summarized in Chapter 2.

Task 2: Solicit water utility and stakeholder input to identify field-scale applications of interest and challenges to implementation of NGS technologies. Insights from these interviews with utility and other industry stakeholders are described in Chapter 3.

Task 3: Develop standard operating protocols for application of NGS by the water industry. A comprehensive guidance document was generated that details the available methodologies and technical factors that NGS users should consider and is available in Chapter 4.

Task 4: Demonstrate promising field applications of NGS using existing data sets to advance understanding of water, wastewater, and water reuse. Five case studies utilizing NGS for a variety of applications are described in Chapter 5.

Task 5: Conduct validation experiments to address key knowledge gaps needed to expand application and consistency of NGS technologies. A description of the approach and findings resulting from these experiments are detailed in Chapter 6.

Task 6: Compile a comprehensive guidance document detailing the overall results of Tasks 1-5 and incorporate utility feedback. This final project report serves as this guidance document.

ES.4 Results

There is enormous value that NGS technologies pose to bring to the water industry. Through a comprehensive literature review and input from water utility stakeholders, comprehensive guidance and standard operating procedures were developed. The literature review indicated wide application of NGS in the realm of scientific research, but more limited translation into day-to-day use by utilities. Results from utility stakeholder interviews indicated increasing awareness of NGS approaches, but also revealed barriers to implementation. In addition to cost of analysis, the vast array of field, laboratory, and data analysis methods applied makes it difficult for utilities to identify an entry point. Standardization would be helpful, but may not be realistic, given the wide array of applications and the rapidly developing nature of the field. Guidance provided here can help the water industry to understand and navigate these options, by matching methods with the information that is desired to be gained. Cost-effective, user-friendly protocols and tools that produce readily interpretable, actionable results would be ideal.

This report provides a comprehensive overview of how NGS is being applied to the study of aquatic environments, especially wastewater, drinking water, and recycled water. Detailed explanations of various protocols for sample collection, sample processing, nucleic acid extraction, NGS, and data analysis are provided, including pros and cons of each approach. Case studies are provided to illustrate how NGS can be effectively applied, using examples of tracking pathogens, profiling antibiotic resistance, and profiling functional gene composition.

Finally, comprehensive validation studies were conducted to expand the application and consistency of NGS investigations of wastewater. Key knowledge gaps that were addressed include evaluation the absolute quantitative capacity and the limit of quantification (LOQ) and detection (LOD) of shotgun Illumina sequencing, as well as the direct effect of DNA extraction methodologies for reconstruction microbial communities. The development of NGS technologies is on a sharp trajectory and will likely continue to rapidly develop both at the benchtop and the laptop. The guidance in this report will help to inform this trajectory in a manner that benefits application in the water industry, helping to improve the representativeness, accuracy, reproducibility, and interpretability of water quality monitoring data. With the implementation of robust experimental controls and informed experimental designs, the current state of NGS technologies can serve as a powerful tool to investigate dimensions of water and wastewater systems that were previously inaccessible.

ES.5 Benefits

NGS is now widely applied across scientific literature for the purpose of studying microbes in various aquatic systems including wastewater, drinking water, recycled water, and surface water. NGS is a very powerful tool that can access vital information about microbes inhabiting these systems in a non-targeted fashion, i.e., without the need to select targets *a priori*. Water utilities can substantially benefit from NGS technologies, especially for the purpose of verifying pathogen and antibiotic resistance removal/attenuation by various treatment processes, and also for profiling microbial metabolic functions of interest. There has especially been interest in NGS technologies recently in the realm of wastewater-based surveillance of pathogens, such as SARS-CoV-2, and there are tremendous potential public health benefits. However, there are many barriers to water utilities benefiting from NGS technologies. Cost and access to equipment to process and prepare samples for sequencing is viewed to be a key barrier, especially for smaller water utilities. Also, because of the wide variety of methods applied in the literature, it is difficult for utilities to identify an entry point. This report serves as such an entry point, providing a guidance framework for aligning the selection of methods and analysis pipelines with the information that is desired to be obtained. In addition to a detailed explanation of NGS "from bench-top to laptop," case studies are included to illustrate specific examples of how water utilities might apply NGS. Finally, a validation study was conducted to address key knowledge gaps, including assessing the effect of sequencing depth, comparing short and long-read sequencing, and evaluating the quantitative capacity of NGS through internal standards.

ES.6 Related WRF Research

- Characterization of the Microbiome of a State-of-the-Art Water Reuse System to Enhance Treatment Performance (4784)
- Advancing Understanding of Microbiomes in Drinking Water Distribution Systems and Premise Plumbing Using Meta-omics Techniques (4733)
- Literature Review: Advancing Understanding of Microbiomes in Drinking Water Distribution Systems and Premise Plumbing Using Meta-omics Techniques (4700)

CHAPTER 1

Background, Motivation, and Objectives

1.1 Motivation

The emergence of next generation sequencing (NGS) is revolutionizing the potential to address complex microbiological challenges in the water industry. NGS technologies can provide new and holistic insight into microbial communities and their functional capacities in water and wastewater systems, thus eliminating the need to develop a new assay for each target organism or gene. For example, NGS is beginning to be applied towards tracking a myriad of pathogens and viruses of concern in water systems, assessing functional capacities of microbial communities, enhancing biodegradation of recalcitrant compounds, and identifying treatment technologies that minimize potential for antibiotic resistance to spread, among many other applications. However, several barriers have hampered wide-scale adoption of NGS in the water industry, including cost, need for specialized expertise and equipment, challenges with data analysis and interpretation, lack of standardized methods, and the rapid pace of development of new technologies. This project is aimed at helping overcome such obstacles and making NGS technologies accessible to water professionals, including not only researchers, but also utility staff, consultants, and regulators.

1.2 Objectives

The overarching objective of this project is to advance the application of NGS in the water industry. This objective was achieved by documenting the current state-of-the-science of NGS technologies; identifying key knowledge gaps and barriers to their implementation; validating relevant approaches; and developing a guidance document to educate water professionals and aid in navigating options for NGS applications. Six specific tasks were completed:

Task 1: Conduct a comprehensive literature review detailing existing and emerging NGS approaches relevant to the water industry. The findings of this comprehensive literature review are summarized in Chapter 2.

Task 2: Solicit water utility input to identify field-scale applications of interest and challenges to implementation of NGS technologies. Insights from these interviews with utility and other industry stakeholders are described in Chapter 3.

Task 3: Develop standard operating protocols for application of NGS by the water industry. A comprehensive guidance document was generated that details the available methodologies and technical factors that NGS users should consider and is available in Chapter 4.

Task 4: Demonstrate promising field applications of NGS using existing data sets to advance understanding of water, wastewater, and water reuse. Five case studies utilizing NGS for a variety of applications are described in Chapter 5.

Task 5: Conduct validation experiments to address key knowledge gaps needed to expand application and consistency of NGS technologies. A description of the approach and findings resulting from these experiments are detailed in Chapter 6.

Task 6: Compile a comprehensive guidance document detailing the overall results of Tasks 1-5 and incorporate utility feedback. This final project report serves as this guidance document.

1.3 Background

Drinking water sources, such as lakes and rivers, treated drinking water, and wastewater are all complex environments that span a range of water qualities. In addition to being defined by variable physicochemical water quality characteristics, these waters each comprise a rich and diverse microbial ecosystem. Even treated drinking water contains multitudes of microorganisms, with 10^3-10^5 cells ml^{-1} typically detected in the water column (Hammes et al., 2008; Hoefel et al., 2003; Vital et al., 2012) and 10^6-10^{11} cells cm^{-2} lining distribution system pipe walls in biofilms (Morvay et al., 2011; Zacheus et al., 2001). Understanding the composition of the microbial communities in these waters can be beneficial for detecting pathogens and improving the understanding of their ecological niches, tracking changes in the abundance of organisms responsible for adverse effects, such as corrosion or biofouling, and characterizing the assemblages of microbiota responsible for degradation of contaminants and microbial substrates in treatment processes.

While monitoring of water sources and systems remains heavily reliant on culture-based approaches for enumeration of pathogenic bacteria and indicator organisms, these methods often vastly underestimate true microbial numbers, failing to detect viable, but non-culturable (VBNC) cells, slow-growing phenotypes, and organisms for which the nutritional requirements and environmental niches are not easily replicated in a laboratory setting (Alleron et al., 2008; Byrd et al., 1991; Staley, 1985). The use of targeted molecular techniques, such as quantitative polymerase chain reaction (qPCR), has become prevalent, but requires *a priori* knowledge of organisms of interest while offering little insight into overall microbial community dynamics. The emergence of NGS technologies is revolutionizing the use of molecular techniques for understanding complex microbial communities. However, the application of NGS in the water sector has been somewhat limited due to cost, need for specialized expertise and equipment, challenges with data analysis and interpretation, lack of standardized methods, and the rapid pace of new technological developments. Given the immense potential of NGS for improving the understanding of complex microbial communities in water and wastewater, effort is needed to overcome these obstacles and make NGS technologies accessible to water professionals.

1.3.1 NGS Technologies

NGS is used to describe a variety of high-throughput nucleic acid (i.e., deoxyribonucleic acid (DNA) and ribonucleic acid (RNA)) sequencing technologies, which now make it possible to directly and rapidly recover millions of DNA or RNA sequences from environmental samples (Shokralla et al., 2012). NGS is a major advancement relative to the traditional “first” generation sequencing technology (i.e., Sanger sequencing), which generates single sequences at a time at as much as two orders of magnitude higher cost per base pair (bp) (Niedringhaus et al., 2011).

1.3.1.1 Short Read Sequencing Technologies

The majority of NGS studies published to date in the water and wastewater fields rely on the use of short read sequencing technologies, primarily those marketed by Illumina, Inc. (San Diego, CA). Studies relying on short read sequencing have also utilized other platforms, including Ion Torrent semiconductor sequencing (Thermo Fisher Scientific, Inc.) and 454 Pyrosequencing (Roche, Inc.), though the latter platform was discontinued in 2016. Illumina sequencing produces millions or even billions of reads in each run, ranging in length from 50-600 bp. In this approach, genomic DNA (or complementary DNA [cDNA] in the case of RNA sequencing) is fragmented or gene regions amplified and adapters are used to tether short DNA fragments to the solid surface of a flow cell. Bridge amplification PCR serves to generate millions of copies of each DNA fragment tethered to a solid surface, resulting in the formation of millions of "clusters." The sequencing surface is flooded sequentially and repeatedly with each of the four deoxyribonucleotide triphosphates (dNTPs; i.e. adenine, "A"; thymine, "T"; guanine, "G"; cytosine, "C"), and as dNTPs complementary to the first available position in a single-stranded template are incorporated, fluorescent signals are emitted (Shendure and Ji, 2008). Detection of fluorescent signals (or lack thereof, as is the case when dGTPs are incorporated on newer Illumina models) unique to each dNTP is used to generate a sequence corresponding to each cluster. The Ion Torrent family of sequencing technologies also produces short read sequencing data, but, rather than detecting a fluorescent signal, the instrument records base calls by detecting a change in pH that occurs upon nucleotide incorporation (Wanger et al., 2017). Instead of tethering DNA fragments to a solid substrate to form clusters, each template DNA strand is tethered to a bead and amplified via emulsion PCR. Each bead is then loaded into a microwell on a chip, with each well producing one of millions of reads. Short read sequencing technologies are advantageous because they are massively high throughput, capable of producing millions of sequences for a single sample, and accurate, with researchers documenting 0.26-0.80% of base calls being erroneous on Illumina and 1.71% on Ion Torrent instruments (Quail et al., 2012). This ability to generate millions of sequences for each sample is critical if metagenomics is to be used to randomly subsample a mixed microbial community and obtain representative profiles of its composition. The primary disadvantages associated with short read sequencing technologies are the inability to generate longer reads and cost. Analysis of short reads is limited in the ability to understand the context of sequenced genes, for example, often making it impossible to identify the organism of origin for a gene. While the cost to generate sequencing data on a per base basis has plummeted from over $5,000 per megabase in 2001 to just $0.01 per Mb in 2015 ("DNA Sequencing Costs: Data," 2018), the cost of preparing DNA for sequencing and of generating sufficient data to be representative of a complex microbial community can still be quite expensive and is often a barrier to adopting the technology.

1.3.1.2 Long Read Sequencing

While short read sequencing technologies have dominated the scientific literature with respect to water, wastewater, and other environmental applications, single molecule, long read sequencing (a.k.a., "Third Generation Sequencing") is emerging as a powerful alternative. The key advantage of long read sequencing methods is that they generate extremely long reads, often producing DNA sequences exceeding 20 kilobases in length (Ardui et al., 2018). The ability to generate long reads is particularly valuable for sequencing genomes, mobile genetic elements (MGEs), and entire operons without the need for assembly. Long reads are especially critical for understanding gene context and ascertaining with higher certainty from which organism a gene originates or which other genes flank a gene of interest. Further, the ability of long read systems to sequence a single DNA molecule eliminates the need for amplification steps, such as PCR, and associated biases. Two platforms dominate the long read market: Pacific BioSciences, Inc. and Oxford Nanopore. While Pacific Biosciences offers a traditional lab-based instrument, Oxford Nanopore has developed a range of platforms, including highly portable sequencing platforms that can be easily transported and operated on-site. The key disadvantage of long read sequencing methods is that they often have error rates as high as 5-13% (Ardui et al., 2018; Batovska et al., 2017), limiting the ability to detect gene mutations or characterize hypervariable regions using these methods. However, adaptations to long read sequencing have been developed, such as circular consensus sequencing, that can vastly improve the accuracy of these methods, with an average of 99.8% accuracy reported (Wenger et al., 2019). Additionally, in December 2020, Oxford Nanopore introduced improvements to their PromethION flow cells reported to have an accuracy as high as 99.1% (Oxford Nanopore Technologies, 2020). While long read sequencing technologies have historically yielded shallower sequencing depths than short read platforms (Judge et al., 2015; Kranz et al., 2017), recent advances have substantially improved long read output, with the recent PromethION advancements reporting outputs as high as 10 terabases (Oxford Nanopore Technologies, 2020).

1.3.2 NGS Methodologies

The application of NGS technologies relevant to the water industry generally can be classified into four subcategories: whole genome sequencing (WGS), metagenomic sequencing, metatranscriptomic sequencing, and targeted sequencing of amplified gene regions (i.e., amplicon sequencing). While there are numerous approaches in practice, an example workflow for each NGS approach is presented in Figure 1-1.

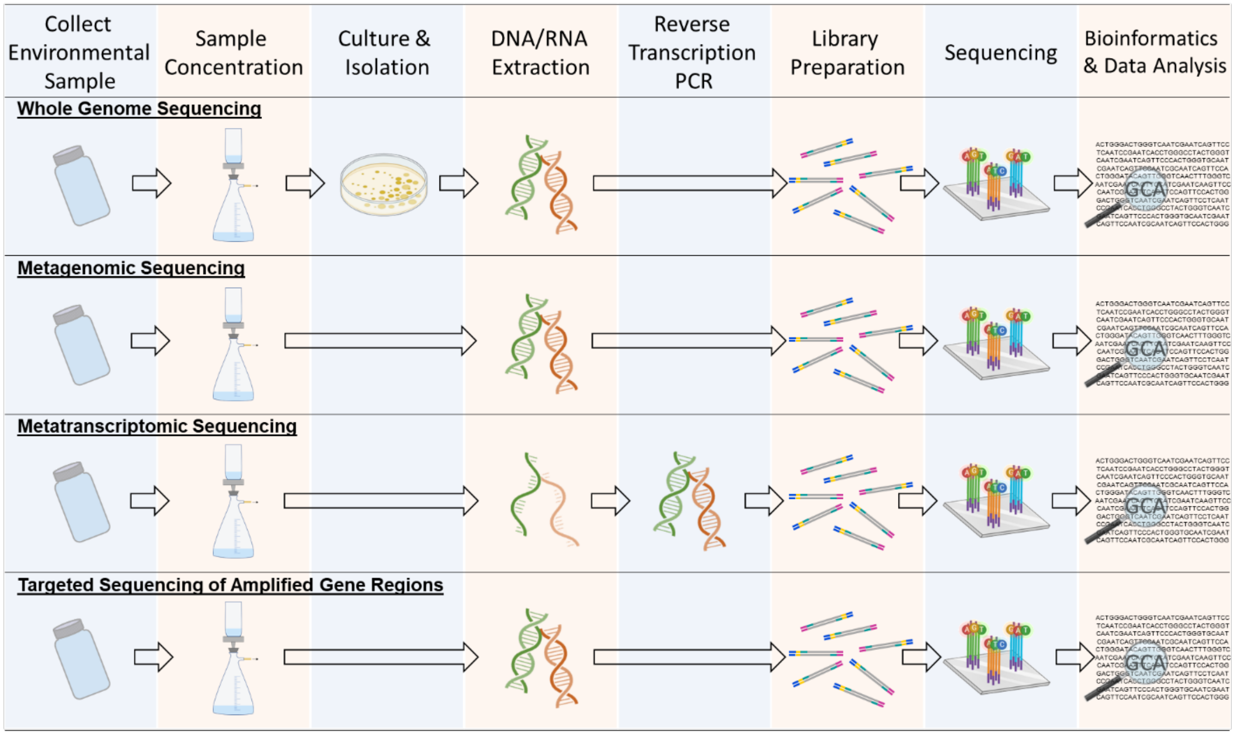

Collect Environmental Sample
Sample Concentration
Culture & Isolation
DNA/RNA Extraction
Reverse Transcription PCR
Library Preparation
Sequencing
Bioinformatics & Data Analysis
Whole Genome Sequencing
Metagenomic Sequencing
Metatranscriptomic Sequencing
Targeted Sequencing of Amplified Gene Regions
GCA

Figure 1-1. Example of a Typical Workflow for Each NGS Approach.
Steps may be added or adapted to meet the specific needs of the user. Specific library preparation also varies according to the needs of the user and may differ from the steps presented here. For example, the library preparation for metagenomic sequencing targets the entirety of genetic material present, while the library preparation for targeted sequencing of amplified gene regions includes amplification of a specific gene of interest.
Source: Reprinted from Garner et al. 2021 with permission from Elsevier.

1.3.2.1 Whole Genome Sequencing

WGS is a powerful means of characterizing microorganisms that have been isolated via culture-based methods. Once isolated as a pure culture, DNA is extracted and sequenced, typically via short read shotgun sequencing approaches. Short read data sets are then assembled to construct the whole genome sequence using either *de novo* or guided assembly. *De novo* relies on assembling short reads to create full sequences without using a template, while guided assembly maps reads to a specified reference genome (Ng and Kirkness, 2010). In cases where an organism has not been previously sequenced, or when the species of the isolate is unknown, as is often the case with isolates obtained from complex water and wastewater communities, *de novo* assembly must be used. *De novo* assembly is computationally challenging, but several open-source tools have been developed to facilitate this need, such as IDBA-UD (Peng et al., 2012), SPAdes (Bankevich et al., 2012), Velvet (Zerbino and Birney, 2008), and MEGAHIT (Li et al., 2015c), among others. The key principle disadvantages of short read sequencing technologies for WGS are that enrichment of the target organism in pure culture is typically required to generate sufficient uncontaminated DNA for sequencing and that short read technologies often result in fragment genomes. While WGS has typically relied on isolation, culture, and short read sequencing technologies, the emergence of long read sequencing platforms capable of generating the sequence of a single DNA molecule may advance and simplify WGS in the future (Shapiro et al., 2013). Use of these long read technologies is also advancing the field of WGS by providing greater genome coverage than short read technologies, often resulting in improved ability to generate closed or complete genome sequences. WGS has been applied extensively in the water and wastewater field to track sources of drinking water outbreaks (Garner et al., 2019a; Raphael et al., 2016), as well as to identify catabolic pathways associated with nutrient removal (Chao et al., 2016; Meng et al., 2019).

1.3.2.2 Metagenomic Sequencing

Metagenomic sequencing, also known as shotgun metagenomic sequencing, refers to the random subsampling and sequencing of genetic material from an environmental sample or other mixed community (Schloss and Handelsman, 2005). This approach has grown rapidly in its application to examining microbial communities and their functional capacities in water and wastewater. Typically relying on short-read sequencing technologies, reads are annotated to genes in existing databases to determine their taxonomic origin or putative functions. For example, metagenomics is now commonly applied to identify antibiotic resistance genes (ARGs) (Garner et al., 2018a; Stamps and Spear, 2020; Zhang et al., 2015), genes associated with nitrification and denitrification in wastewater (Cai et al., 2016; Ye et al., 2012), catabolic genes involved in biodegradation (Folch-Mallol et al., 2019; Sidhu et al., 2017), viruses (Bibby et al., 2011; Tamaki et al., 2012, shifts or differences in overall microbial community structure (Brumfield et al., 2020; Hull et al., 2017), and genes associated with pathogens from within a mixed community (Cui et al., 2019; Kumaraswamy et al., 2014; Li et al., 2015d; Saleem et al., 2018). Metagenomics is also increasingly used to support the assembly of complete or partial genomes from short-read sequencing data generated from uncultured microbial communities (i.e., metagenomic-assembled genomes (MAGs)) (Alneberg et al., 2018; Handley et al., 2014). While short read sequencing is common in metagenomic sequencing, use of long read sequencing is also emerging (e.g., Driscoll et al., 2017).

1.3.2.3 Metatranscriptomic Sequencing

Metatranscriptomic sequencing relies on principles similar to metagenomic sequencing, but targets RNA rather than DNA. RNA sequencing is especially key for identifying RNA viruses (e.g., SARS-CoV-2), although this is typically still referred to as “metagenomics” because it targets RNA genomes rather than mRNA transcripts. Metatranscriptomics also can be applied to directly sequence RNA transcripts and thus yield greater insight into microbial activity than metagenomics. For example, targeting messenger RNA (mRNA) provides direct information about which genes are actually being expressed (Carvalhais et al., 2012). Sequencing transcribed mRNA can similarly provide a picture of which microbes are functionally active. However, it is important to recognize that metatranscriptomics is often hampered by the fact that RNA degrades rapidly in environmental samples. Degradation rates may differ among species, making preservation and analysis of mRNA challenging (Pascault et al., 2014), and large technical and biological variation is often observed in metatranscriptomics datasets (Tsementzi et al., 2014). Approaches for preservation of RNA have been applied to help reduce decay in sensitive samples for subsequent analysis (Kohl et al., 2017; Tap et al., 2019). An excess of ribosomal RNA (rRNA) in transcriptomes also creates challenges in characterizing mRNA in environmental samples, thus application of methods that deplete rRNA prior to sequencing are of great value in generating high quality mRNA datasets (He et al. 2010). Due in part to these challenges, the application of metatranscriptomics has been limited relative to methods that target DNA in full-scale water and wastewater systems.

1.3.2.4 Targeted Sequencing of Amplified Gene Regions

Targeted sequencing of amplified gene regions, or “amplicon sequencing,” relies on polymerase chain reaction (PCR) amplification of a gene of interest, followed by short read sequencing of the product. Most commonly, this approach is applied to the 16S rRNA gene, which is universal to bacteria and archaea, to determine the phylogeny (i.e., evolutionary lineage) or taxonomy of the members of the microbial community (Caporaso et al., 2011). The method relies on the use of primers for PCR amplification that target highly conserved regions of the target gene, but capture sufficient hypervariable regions to distinguish various forms of genes. Typically, reads will be clustered based on similarity into operational taxonomic units or sequence variants and compared to existing databases to determine taxonomy. Although amplicon sequencing is powerful for resolving phylogenetic and taxonomic variation among members of microbial communities, such approaches as applied to the 16S rRNA gene are still most accurate at the phylum, class, and order levels, with the ability to differentiate between microbes at the family, genus, species, or strain level often limited. Amplicon sequencing of 16S rRNA genes is often used to understand overarching shifts in microbial communities, such as disturbances, succession, and temporal trends. Similar approaches have also been used to identify waterborne eukaryotes (e.g., amoebae and protozoa) by targeting the 18S rRNA gene (Bradley et al., 2016), fungal communities by targeting the internal transcribed spacer (ITS) region (Bokulich and Mills, 2013), and specific virus families, for example by targeting the hexon gene specific to adenovirus (Iaconelli et al., 2017; Kuo et al., 2015). Commercial options are available that rely on this approach to target multiple gene targets simultaneously to profile multiple ARGs or pathogens, such as the AmpliSeq panels produced by Thermo Fisher Scientific. Amplicon-based approaches have become widely popular given their broad applicability and modest cost compared to metagenomic sequencing. While short read sequencing is typically used to facilitate amplicon sequencing (e.g. Garner et al., 2019b; Ji et al., 2015), the use of long read sequencing to characterize amplicons is emerging (Haig et al., 2018).

CHAPTER 2

Literature Review: Applications of NGS to Study Water and Wastewater

Portions of this chapter are reprinted from *Water Research*; Vol 194; Emily Garner, Benjamin C. Davis, Erin Milligan, Matthew Forrest Blair, Ishi Keenum, Ayella Maile-Moskowitz, Jin Pan, Mariah Gnegy, Krista Liguori, Suraj Gupta, Aaron J. Prussin, Linsey C. Marr, Lenwood S. Heath, Peter J. Vikesland, Liqing Zhang, Amy Pruden; Next generation sequencing approaches to evaluate water and wastewater quality; Copyright (2021), with permission from Elsevier.

2.1 Introduction

Initially used to characterize soil bacterial diversity, the application of NGS to environmental systems has grown to be commonly employed for studying a vast array of environments, ranging from ancient permafrost samples (D'Costa et al., 2011) to surfaces in the International Space Station (Be et al., 2017; Venkateswaran et al., 2014). NGS is now beginning to be proven as a valuable tool for expanding understanding of water, wastewater, and water reuse systems. To document the breadth of ways that NGS technologies have been used for the study of water and wastewater, a systematic literature review was conducted. In addition, key knowledge gaps and research needs were identified based on the scope of the available literature.

2.2 Approach

To synthesize the breadth of existing applications of NGS for studying water and wastewater, a systematic review of the existing literature was conducted, and key themes were identified as ways NGS is being applied to address water quality challenges. Briefly, a search was conducted of the Clarivate Analytics Web of Science Database for studies published between January 1, 2000 and December 31, 2019 written in the English language. A three-tiered search strategy was employed to identify (1) studies that utilize NGS, (2) studies focusing on water or wastewater environments that are related to the water industry, and (3) studies that focus on key applications. The number of references identified in each tier are depicted in Figure 2-1.

In total, 15,367 publications were identified that utilize NGS methodology, of these, 651 papers were returned via a keyword search as being relevant to the water industry and were further examined in this systematic review. After manual exclusion of irrelevant publications, six key application areas were identified where NGS is most commonly applied across drinking water, source waters, wastewater, receiving water bodies, and water reuse. These areas include taxonomic classification and pathogen detection, functional and catabolic gene characterization, antimicrobial resistance (AMR) profiling, bacterial toxicity characterization, *Cyanobacteria* and harmful algal bloom identification, and virus characterization. The number of references for each application area identified within the water and wastewater area utilizing each NGS methodology are summarized in Figure 2-2 and key findings of the systematic literature review are summarized in Table 2-1. All articles compiled for this systematic review have been catalogued in a Zotero library, accessible at https://www.zotero.org/groups/2593738.

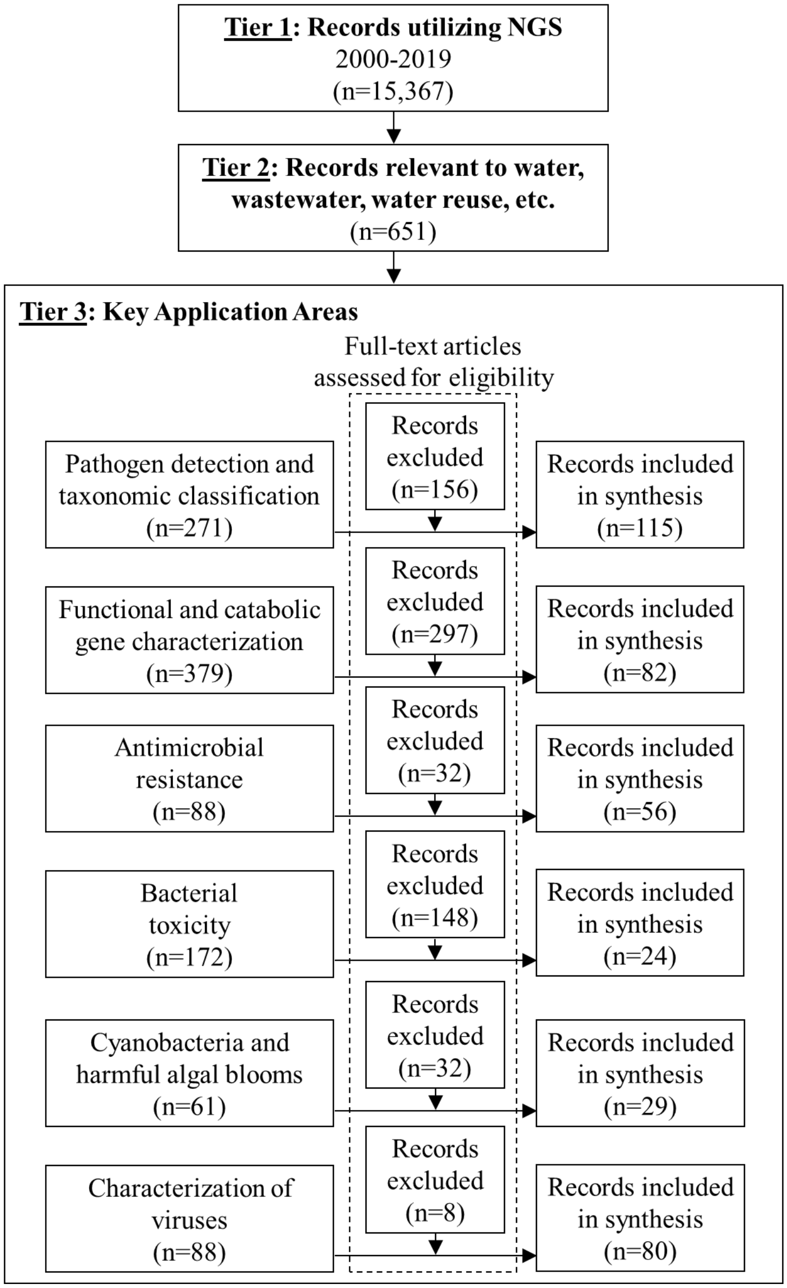

Tier 1: Records utilizing NGS
2000-2019
(n=15,367)
Tier 2: Records relevant to water, wastewater, water reuse, etc.
(n=651)
Tier 3: Key Application Areas
Full-text articles assessed for eligibility
Pathogen detection and taxonomic classification (n=271)
Records excluded (n=156)
Records included in synthesis (n=115)
Functional and catabolic gene characterization (n=379)
Records excluded (n=297)
Records included in synthesis (n=82)
Antimicrobial resistance (n=88)
Records excluded (n=32)
Records included in synthesis (n=56)
Bacterial toxicity (n=172)
Records excluded (n=148)
Records included in synthesis (n=24)
Cyanobacteria and harmful algal blooms (n=61)
Records excluded (n=32)
Records included in synthesis (n=29)
Characterization of viruses (n=88)
Records excluded (n=8)
Records included in synthesis (n=80)

Figure 2-1. Flowchart of the Three-Tiered Search Strategy and Screening for Eligibility.
Source: Reprinted from Garner et al. 2021 with permission from Elsevier.

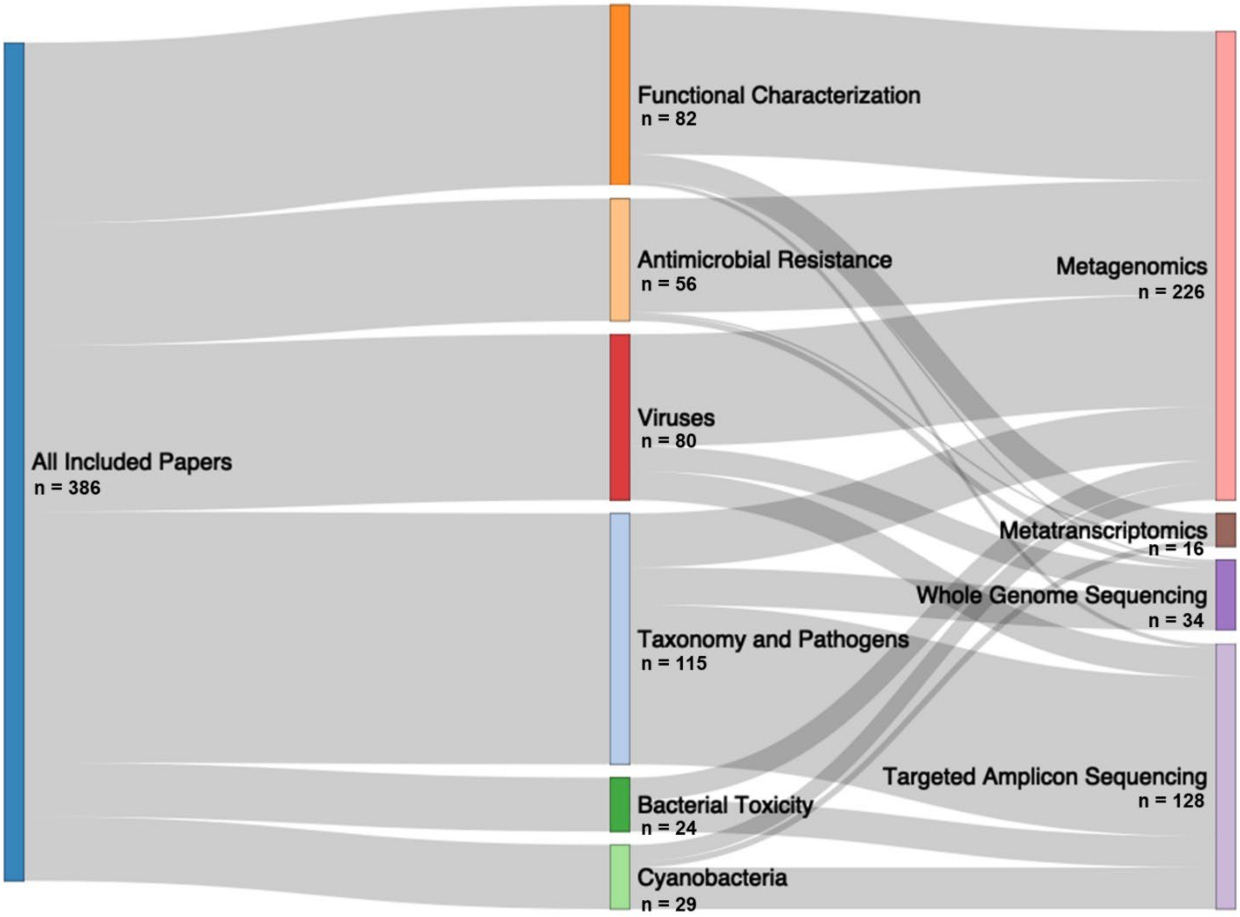


Figure 2-2. Distribution of NGS Methodologies Applied for each Highlighted NGS Water Industry Application. In cases where multiple NGS approaches were used in one study, the study was included in each relevant NGS approach category.
Source: Reprinted from Garner et al. 2021 with permission from Elsevier.

Table 2-1. Summary of Key Trends Observed for Six Key Areas of Application of NGS to Relevant Water Environments.

Source: Reprinted from Garner et al. 2021 with permission from Elsevier.

Application	Frequently Used Methodologies	Frequently Used NGS Platforms	Common Annotation Databases	Key Themes	Challenges & Limitations
Pathogen Detection and Taxonomic Classification	Targeted Sequencing of Amplified Gene Regions, Metagenomics, WGS	Illumina MiSeq, Illumina HiSeq, Roche 454 Pyrosequencing	RDP,[1] GreenGenes,[2] SILVA,[3] PATRIC,[4] NCBI RefSeq[5]	Taxonomic surveys; Pathogen detection; Microbial community shifts and trends	False negatives; Method variability and biases; Limited taxonomic resolution at species/strain level for amplicon sequencing
Functional and Catabolic Gene Characterization	Metagenomics, Metatranscriptomics	Illumina HiSeq, Roche 454 Pyrosequencing, Illumina MiSeq	SEED,[6] KEGG,[7] COG,[8] eggNOG,[9] MG-RAST M5nr[10]	Functional profiling; Nitrogen metabolism; Phosphorus uptake; Metals; Methanogenesis; Biofilms; Metabolism; Transport; Virulence, Defense, and Stress Response; Foaming; Disinfection	Limited experimental, biological, and technical replication
Antimicrobial Resistance	Metagenomics, WGS	Illumina HiSeq, Roche 454 Pyrosequencing	CARD,[11] Resfams[12]	Resistome profiling and trends; Resistance profiling of isolates; Co-occurrence with MGEs and MRGs	Linking findings to risk and human health outcomes; Normalization; Validation of short read assembly
Bacterial Toxicity	Metagenomics, Targeted Sequencing of Amplified Gene Regions	Illumina HiSeq, Roche 454 Pyrosequencing, Illumina MiSeq	GreenGenes,[2] SILVA[3]	Impact of toxic compounds on microbial communities: arsenic and other heavy metals; chlorine; nanomaterials; anthropogenic contaminants	Normalization; Validation of short read assembly
Cyanobacteria and Harmful Algal Blooms	Targeted Sequencing of Amplified Gene Regions, Metagenomics, Metatranscriptomics	Illumina HiSeq, Illumina MiSeq, Roche 454 Pyrosequencing	IMG,[13] KEGG,[7] MG-RAST M5nr[10]	Characterization of cyanobacterial communities; Shifts in gene expression; Removal of cyanobacteria during water treatment	Strain variability compared to reference strains

Characterization of Viruses	Metagenomics, WGS, Targeted Sequencing of Amplified Gene Regions	Illumina HiSeq, Illumina MiSeq, Roche 454 Pyrosequencing		Virome characterization; Viral genotyping	Lack of conserved gene targets, Obtaining sufficient biomass; Viral RNA degradation; Lack of virus reference databases; Lack of standard approaches for data analysis
[1](Cole et al., 2014); [2](DeSantis et al., 2006); [3](Quast et al., 2013); [4](Davis et al., 2020b); [5](O'Leary et al., 2016); [6](Overbeek, 2005); [7](Kanehisa and Goto, 2000); [8](Galperin et al., 2019); [9](Huerta-Cepas et al., 2019); [10](Meyer et al., 2008); [11](Alcock et al., 2020); [12](Gibson et al., 2015); [13](Markowitz et al., 2012)					

2.2.1 Pathogen Detection and Taxonomic Classification

In total, 115 articles were identified that apply NGS for taxonomic studies of aquatic environments. Of these, 77 utilized targeted sequencing of amplified hypervariable gene regions (74 targeting the 16S rRNA gene, three targeting ITS regions), 26 utilized metagenomic sequencing, and 18 utilized WGS. Sequencing across studies was primarily conducted on Illumina platforms (82), with the second most prevalent being 454 pyrosequencing (18). Ion Torrent (5), Nanopore (3), and PacBio (2) were far less common, with the remaining studies using various outdated DNA analysis technologies. After 454 pyrosequencing was discontinued, Illumina's MiSeq became the dominant platform for amplicon sequencing (60) with the HiSeq platforms primarily used for metagenomic studies. Bioinformatic approaches to amplicon sequencing of 16S rRNA hypervariable regions used a handful of databases and software, the most common being the RDP (Cole et al., 2014), GreenGenes (DeSantis et al., 2006), SILVA (Quast et al., 2013), PATRIC (Davis et al., 2020b), and the National Center for Biotechnology Information's (NCBI) RefSeq (O'Leary et al., 2016) databases, with data typically analyzed utilizing QIIME (Caporaso et al., 2010), mothur (Schloss et al., 2009), BLAST (Altschul et al., 1990), or MEGAN software (Huson et al., 2007), or the MG-RAST (Meyer et al., 2008) and RDP services (Cole et al., 2014). Freshwater environments were the focus of the majority of studies (81), including drinking water and drinking water distribution systems (DWDS) (28), premise plumbing and various drinking water infrastructure (22), and river water (21). Wastewater treatment was analyzed in 16 studies, including profiling the treatment train (10), biosolids (3), membrane bioreactors (2), activated sludge (2), and anaerobic digestion (1).

General taxonomic surveys were the focus of 42 studies. These studies examined the microbiomes of aquatic environments either through metagenomic or 16S rRNA targeted sequencing to profile microbial communities and corresponding shifts in composition in response to various changes in condition with time to infer the effects of differential water quality parameters. Recently, taxonomic surveys have revealed the surprising diversity of microbes inhabiting drinking water and DWDSs. It has been discovered that drinking water microbiomes are strongly shaped by disinfectants, water age, pipe materials, and seasonal and source water variations (Bae et al., 2019; Chiao et al., 2014; Douterelo et al., 2018; Gomez-Alvarez et al., 2016; Liu et al., 2018; Mukherjee et al., 2016; Pinto et al., 2014; Potgieter et al., 2018; Roeselers et al., 2015; Saleem et al., 2018; Shaw et al., 2015; Stamps et al., 2018; Wang et al., 2014). For example, Potgieter et al. (2018) performed a two-year study of a full-scale DWDS to investigate the spatial and temporal dynamics of the microbial community across water age and disinfection regimes. Long-term NGS studies are important as they avoid stochastic variations from single time point studies and can reveal the systemic processes, infrastructure, and environmental factors that affect the microbial communities in drinking water systems, thus informing downstream management decisions. In another exemplary study, Mansfeldt et al. (2019) examined the impact of microbial residence time (MRT) on taxonomic composition in activated sludge. The authors found that longer MRTs resulted in a greater range of growth parameters enabling microorganism persistence, thus increasing richness and diversity. Findings were verified using a model based on Monod-growth kinetics, demonstrating the importance of efficient resource capture and survival during low production for persistence in activated sludge. Collectively, such studies that utilize NGS to examine the underlying ecology of microbial systems make important progress towards addressing how microbiomes might be engineered to support effective degradation of wastewater substrates and optimize nutrient removal.

Pathogen detection was a principal focus of the majority (79) of taxonomy-related papers. Several important sub-topics emerged with respect to studying pathogen profiles specifically, including increasing the range and sensitivity of pathogens detected during screening of drinking water sources and recreational waters (Cui et al., 2019, 2017; Fang et al., 2018b; Haig et al., 2018; Hamner et al., 2019; Huang et al., 2014; Jin et al., 2018; Layton et al., 2014; Nadya et al., 2016; Shrestha et al., 2019, 2017; Vadde et al., 2019; VanMensel et al., 2020; Yang et al., 2020a), exploring pathogen content and ensuring adequate pathogen removal during wastewater treatment (Bibby et al., 2010; Cai and Zhang, 2013; Chen et al., 2019; Guo and Zhang, 2012; Harb and Hong, 2017; Hembach et al., 2019; Huang et al., 2018; Leddy et al., 2017; Li et al., 2015a; Lu et al., 2015; Osunmakinde et al., 2019; Wan et al., 2018; Yergeau et al., 2016; Zhang et al., 2019), and monitoring for pathogen regrowth during wastewater reuse applications and drinking water distribution and storage (Garner et al., 2018b, 2019a; Kumaraswamy et al., 2014; Obayomi et al., 2019; Proctor et al., 2015; Qin et al., 2017). Li et al. (2015a) used metagenomic sequencing of the influent, activated sludge, aeration basin biofilm, aeration basin foam layer, anaerobic digestion sludge, and the final effluent of various wastewater treatment plants (WWTP) in Hong Kong to detect pathogens using a genetic marker taxonomic classifier (MetaPhlAn) and a curated pathogen index. The authors reported 98% removal efficiency of all pathogens and 4-log removal of *Escherichia coli* and *Enterococcus faecalis* from influent to effluent, showing that secondary treatment is an effective means to reduce most bacterial pathogens in wastewater. This study is important to the water industry because its straightforward bioinformatic workflow and application of engineering removal efficiencies closely resemble culture-based and qPCR studies, exemplifying the application of NGS for *in situ* pathogen detection and relative quantification. In a related study, Kumaraswamy et al. (2014) used amplicon sequencing with a nested PCR approach targeting the V4 region of the 16S rRNA gene paired with annotation against Greengenes and a preconstructed human pathogenic bacteria 16S rRNA database to assess the presence of potential pathogens in treated wastewater intended for reuse. They found abundant pathogen markers in treated wastewater post-disinfection that were not detected by culture, particularly Gram-positive pathogens, revealing an elevated health risk and warranting further screening and risk assessment. These examples are of importance to the water industry because they highlight the utility of NGS techniques for pathogen screening to uncover trends in overall behavior of multiple putative pathogens that would not be possible by culturing, thus more comprehensively informing downstream applications.

Approximately one-third (27) of the pathogen-centered literature focused on monitoring opportunistic pathogens (OP), such as *Pseudomonas aeruginosa*, *Mycobacterium avium*, and *Legionella* spp. in relevant water environments. Eleven studies assessed the epidemiological relationship between potential reservoirs of OPs and patient isolates (Bartley et al., 2016; David et al., 2017; Fitzhenry et al., 2017; Fleres et al., 2018; Graham et al., 2014; Lande et al., 2019; Lévesque et al., 2016, 2014; Quick et al., 2014; Raphael et al., 2016; Runcharoen et al., 2017;

Wüthrich et al., 2019). For example, Bartley et al. (2016) used WGS to analyze isolates from a hospital water distribution system, infected patients and retrospective patients, and found that a single *L. pneumophila* population was responsible for all nosocomial infections in 2011 and 2013. WGS has proven to be an effective tool for real-time outbreak investigations that rely on the rapid identification of the associated organisms to identify and mitigate sources of infection. Additionally, nine studies focused on studying OP colonization of cooling towers (Farhat et al., 2018; Fitzhenry et al., 2017; Lévesque et al., 2014; Llewellyn et al., 2017; Nakanishi et al., 2019; Paranjape et al., 2020; Pereira et al., 2018, 2017; Wüthrich et al., 2019). In parallel to studying the distribution of *Legionella* in 196 cooling towers across eight US climate regions, Llewellyn et al. (2017) used targeted sequencing of the 16S rRNA gene to investigate the cooling tower microbial communities. Interestingly, they found the taxonomic distributions were homogeneous across the U.S.

Method development and validation was the focus of 14 papers and focused largely on improving precision and sensitivity for targeted sequencing techniques (Albertsen et al., 2015; Bautista-de Los Santos et al., 2016; Brandt and Albertsen, 2018; Greay et al., 2019; Guo et al., 2013; Lee et al., 2017; Saingam et al., 2018; Sato et al., 2019b; Spencer et al., 2016). For example, Lee et al. (2017) designed novel primer sets for 16S rRNA gene amplification to increase the specificity for the Cyanobacteria and Proteobacteria phyla. Organism-specific assays have also been developed. For example, Pereira et al. (2018, 2017) developed genus-specific NGS assays for both *Legionella* spp. and *Pseudomonas* spp., which employ organism-specific PCR primer sets that target the 16S rRNA hypervariable regions, with the intent of increasing the taxonomic resolution to the species level. This approach is particularly valuable in identifying potential risks associated with pathogen-containing genera because methods that amplify the 16S rRNA gene using conserved primers often lack the ability to resolve the presence of pathogenic species from genetically similar organisms. Sato et al. (2019b) developed a *Leptospira* spp. specific NGS assay to investigate epidemiological links in Leptospirosis endemic regions of rural Japan, increasing the range and specificity of risk assessments in environmental waters.

There are many challenges associated with taxonomic profiling and pathogen screening of bacterial communities using NGS, but the methods and platforms used are undoubtedly the most developed amongst other NGS applications. The first challenge is that, for any amplicon or shotgun-based sequencing analysis, an organism could still be present even if not detected (i.e., false negatives). Sensitivity of microbial community analysis is dependent upon a multitude of factors, including DNA extraction methods, PCR biases, databases and classifiers used, sequencing technique (e.g. 16S rRNA gene amplicon or metagenomic), and most of all, the sequencing depth (Albertsen et al., 2015; Brandt and Albertsen, 2018; Guo et al., 2013; Yergeau et al., 2016), none of which have been standardized. A second challenge, that the 16S rRNA hypervariable region often does not provide sufficient taxonomic resolution for pathogen identification. Greay et al. (2019) demonstrates that the V4 region did not exhibit the specificity necessary to differentiate the *Enterobacteriaceae* family of *Gammaproteobacteria* in wastewater, a critical group of organisms for water quality monitoring. This study also cross-checked 16S rRNA amplicons against the NCBI nucleotide collection and found that nine of the

12 pathogens detected were erroneously assigned using the GreenGenes database. This issue is a systematic error that has most likely led to false positives in numerous NGS studies to date.

It should be noted that, because strain-level differentiation is often necessary, only WGS has the resolution to reliably verify the presence of pathogens (e.g., by lineage and presence of key virulence genes). In contrast, read-based metagenomics and 16S rRNA gene amplicon sequencing at best only provide a screen for the potential presence of pathogens by identifying genus or higher ranks of taxonomic classification. However, when sequencing coverage is sufficient, MAGs can be generated from metagenomic data and support strain-level resolution (Quince et al., 2017a). While ongoing method development and validation is important for improving the application of taxonomic characterization of water and wastewater microbial communities, these methods are promising in their ability to characterize typical and atypical microbial communities in water systems, and to screen for the presence of a broad range of pathogens while bypassing culture bias.

2.2.2 Functional and Catabolic Gene Characterization

In total, 82 records were identified that applied NGS for studying metabolic and other functional activity in drinking water, wastewater, and recycled water. Of these, 70 studies were focused on genomic DNA, seven on RNA, and five on a combination of both DNA and RNA. Metagenomic sequencing was most common (67), followed by metatranscriptomics (7), or a combination of the two (5). Of the remaining papers, two utilized amplicon sequencing and one implemented WGS. Sequencing was primarily conducted on Illumina-based platforms, such as HiSeq (42), MiSeq (9), and NextSeq (2), with Roche 454 pyrosequencing making up the second largest sequencing technology (15) utilized. The rest of the studies employed various platforms. Wastewater was the focus of the majority of studies, targeting wastewater in general (5), A^2O process (1), anaerobic digestion (7), ANAMMOX (2), activated sludge (14), industrial treatment (8), membrane bioreactors (1), microbial fuel cells (3), biosolids (2), and upflow anaerobic sludge blanket digestion (2). Characterization of functional genes in various drinking water sources was also prevalent: aquifers/groundwater (8), freshwater (1), freshwater sediments (5), lakes (5), and rivers/streams (6). Other studies targeted drinking water treatment and distribution systems (8), fracking (2), constructed wetlands (1), and petroleum reservoirs (1). Functional annotation was conducted using a variety of predefined (74) and custom databases (9). A large majority of studies utilized some combination of the SEED (Overbeek, 2005), KEGG (Kanehisa and Goto, 2000), COG (Galperin et al., 2019), and/or eggNOG (Huerta-Cepas et al., 2019) databases (COG and KEGG – 3, COG eggNOG and KEGG – 1, KEGG – 11, KEGG and eggNOG – 3, KEGG UniRef100 and Uniprot – 1, SEED and COG – 1, SEED and KEGG – 20, and SEED - 11) or the MG-RAST M5nr (Meyer et al., 2008) databases (9), which combines non-redundant sequences from GenBank, SEED, IMG, Uniprot, KEGG, and eggNOG databases. Though functional analysis played a major role in all of these studies, only six applied it as the sole focus, with 77 conducting it in conjunction with taxonomy-based analysis.

Several common themes emerged with respect to how NGS has been used to study metabolic and other functional activity in relevant water environments. The most common application of functionally-applied NGS was functional gene profiling, either comprehensively or by examining specific functions of interest. These applications were often conducted in tandem to some degree and were applied in 73 studies, with 27 focused heavily on general characterizations and 46 focused on specific functional profiles of interest. Among the 27 studies, the majority applied the SEED, COG, eggNOG and/or KEGG databases to characterize a wide range of functionalities including: cellular processes, metabolism, and information storage and processing. Here, the application of metagenomics aimed at identifying functional genes served primarily to shed light on biological processes, in effect, characterizing the functional profiles within collected samples or comparing them across samples. For example, Douterelo et al. (2018) applied NGS, annotated using MG-RAST and COG functional levels 1 and 2 (returning results for cellular processes & signaling, information storage & processing, metabolism, and poorly characterized), to characterize the differences between biofilm and bulk water functional profiles, finding that they were habitat dependent. Similarly, Sidhu et al. (2017) utilized the SEED and KEGG databases to comprehensively profile the differences between raw sewage and treated sludge from a WWTP finding key genes responsible for aromatic hydrocarbon degradation. A number of additional studies also applied general functional analysis to characterize their specific environments using similar approaches (Cai et al., 2016; Delforno et al., 2017a; Hemme et al., 2015; Ju et al., 2014; Mohan et al., 2014; Saxena et al., 2018; Song et al., 2019; Van Rossum et al., 2015; Vikram et al., 2016).

In contrast to comprehensive profiling of functional genes, a number of specific functions of interests were the focus of 46 of the 73 studies, including: nitrogen cycling/metabolism (Bai et al., 2013; Cai et al., 2019; Chao et al., 2016; Costa et al., 2015, 2015; Crovadore et al., 2017; Emmanuel et al., 2019; Jewell et al., 2016; Keller et al., 2015; X. Liu et al., 2019b; Mason et al., 2014; Peura et al., 2015; Pinto et al., 2016; Rehman et al., 2019; Reid et al., 2018; Sun et al., 2018; Varrone et al., 2014; Z. Wang et al., 2014; Ye et al., 2012), phosphorus uptake (Hemme et al., 2015; LeBrun et al., 2018; Silva et al., 2013; Smith et al., 2012; Tian et al., 2015), metal transformation (Abbai and Pillay, 2013; Bai et al., 2013; Costa et al., 2015; Dai et al., 2018a; Hemme et al., 2015; Varrone et al., 2014), methanogenesis (Bedoya et al., 2019; Biderre-Petit et al., 2019; Cai et al., 2019; Delforno et al., 2017a; Delforno et al., 2017b; Guo et al., 2015; Peura et al., 2015; Reid et al., 2018; Reis et al., 2016; Sidhu et al., 2017; Tomazetto et al., 2015; Varrone et al., 2014; Vikram et al., 2016; Wong et al., 2013; Xia et al., 2018; Yang et al., 2014b; Yu et al., 2018), biofilm formation (Chao et al., 2015; Douterelo et al., 2018; Li et al., 2019b; Rehman et al., 2019; Sun et al., 2018; Vikram et al., 2016), and a wide range of metabolic processes (Cai et al., 2013, 2016; Chao et al., 2013; Cleary et al., 2018; Dai et al., 2018b; Das et al., 2017; Fang et al., 2014; Hamilton et al., 2017; Lu et al., 2017; Ludington et al., 2017; K.-L. Ma et al., 2019; Medeiros et al., 2016; Mohan et al., 2014; More et al., 2014; Rahman et al., 2017; Ruiz-Moreno et al., 2019; Saxena et al., 2018; Silva et al., 2012; Sul et al., 2016; Van Rossum et al., 2015; Yadav et al., 2015; Yu and Zhang, 2012). Studies also investigated transport mechanisms (Keller et al., 2015; Meng et al., 2019; Ye et al., 2012), virulence, defense and stress response (Dai et al., 2018b; Medeiros et al., 2016; Schlüter et al., 2008), foaming (Rosso et al., 2018), and the impacts of disinfection (Chao et al., 2013; Mohan et al., 2014). Ye et al. (2012) provide an exemplar of the ability for functional NGS to identify nitrogen-related functional genes through utilization of the KEGG database's annotation to quantify and characterize genes related to both nitrification and denitrification pathways in full-scale and lab-scale wastewater treatment bioreactors. Pinto et al. (2016) also report metagenomic evidence for the presence of bacteria similar to *Nitrospira* with the ability to completely oxidize ammonia to nitrate in a drinking water system (i.e., COMAMMOX), by discovering ammonia oxidation genes that were divergent from canonical ammonia oxidizers. Gou et al. (2015) and Yang et al. (2014b) provide examples of how functional NGS can be applied to characterize the methanogenesis pathways present in anaerobic digesters, both of which found the acetoclastic pathway to be dominant. Douterelo et al. (2018) applied metagenomics to characterize and compare differences between bulk water and biofilms within DWDSs along with genes related to biofilm formation, finding that that the functional profile differed between habitats and that biofilms preferentially possessed resistance mechanisms related to radical-induced disinfection damage.

The majority of studies applying metagenomics focusing on identifying functional genes were aimed at assessing various metabolic processes, with a wide breadth of environments and pathways examined. Aromatic compound degradation (Abbai and Pillay, 2013; Bai et al., 2013; Delforno et al., 2017a; Jadeja et al., 2019; Mason et al., 2014; More et al., 2014; Sidhu et al., 2017; Silva et al., 2013; Yadav et al., 2015), arsenic metabolism (Cai et al., 2013; Costa et al., 2015; Das et al., 2017; Edwardson and Hollibaugh, 2017), energy metabolism (Cleary et al., 2018; Hemme et al., 2015; Keller et al., 2015; Medeiros et al., 2016; Reid et al., 2018), amino acid and fatty acid degradation (Emmanuel et al., 2019; Fang et al., 2014; Fang et al., 2018a; Lu et al., 2017; Ma et al., 2019), hydrogenases (Bai et al., 2013; Tomazetto et al., 2015; Varrone et al., 2014; Wexler et al., 2005), and oxygenases (Bai et al., 2013; Fang et al., 2013; Jadeja et al., 2014; Mason et al., 2014; Silva et al., 2013; Singh et al., 2010; Yu and Zhang, 2012) were of wide interest, in addition to the more general characterizations identified above. For example, Cai et al. (2016) effectively utilized metagenomics to reconstruct metabolic pathways related to methanogenesis and denitrification during a comparison of two biogas producing digesters, in addition to finding that each digester had a high level of functional redundancy underlying their stability.

Identification of novel functional genes was the focus of seven studies. These studies utilized metagenomics to identify new genes and evaluate their functions. For example, Folch-Mallol et al. (2019) utilized metagenomics to describe a novel thioesterase gene involved in phenylacetic acid degradation seemingly phylogenetically-related to *Actinobacterium.* Additionally, Silva et al. (2013) focused on novel genes/pathways related to phenol and aromatic compound degradation by screening a metagenomic library for phenol hydroxylase genes and phenol degradation activity, resulting in the identification of genes present in wastewater treatment involved in degradation of aromatic compound. Rather than directly examining functional genes, inferring functional capacity from 16S rRNA gene-derived microbial taxonomy was the focus of 2 studies. For example, Cleary et al. (2018) applied PICRUSt predictive functionality to two morphologically different *Cinachyrella* samples from marine Indonesian Lakes, with varying levels of connectivity to the sea, with the goal of comparing bacterial and archaeal communities and their respective functional capabilities to each other, and those found in lakes. The authors noted pronounced differences between both morphospecies' predicted functionality and shared enrichment of pathways related to energy metabolism, stress response, secondary metabolites, and information processing when comparing both morphospecies to prokaryotic communities in water. It should be noted; however, that approaches such as PICRUSt are inherently limited because they indirectly infer functionality based on taxonomy, rather than directly examining functional genes. Thus, accuracy is subject to the extent to which the organisms detected are characterized, bearing in mind that 16S rRNA gene sequencing methods at best have genus-level resolution, while functions may vary at the strain level. Further, the PICRUSt developers caution that a nearest sequenced taxon index score of 0.03 is typically required to predict a target taxon at the species level, and results should be carefully interpreted (Langille et al., 2013). Communities with diverse functional niches and the ability to adapt to different carbon sources would especially confound this approach. Thus, while taxonomically derived functional inference methods may point to some potentially useful hypotheses for follow-up study, they should be applied with extreme caution to complex microbial communities, such as those characteristic of WWTPs and other water environments.

Overall, limited sample replication was apparent in a number of studies (Bedoya et al., 2019; Chao et al., 2016; Debroas et al., 2009; Li et al., 2019b; Medeiros et al., 2016; Reid et al., 2018; Sánchez-Reyez et al., 2017; Xia et al., 2018), which was likely a result of the high costs associated with implementation of NGS. Challenges in characterizing complex metabolisms characteristic of wastewater treatment and a high proportion of 'poorly-characterized hits' during functional annotation were widely noted, which was impacted by limited databases for annotation. Roughly half of all studies were devoted to activated sludge reactors and anaerobic digesters, which are characterized by a relatively strong foundational understanding of key microbiological processes involved. However, focusing on snapshots of singular processes fails to provide context with respect to how functional profiles change temporally and spatially throughout treatment trains and continued operation. The lack of a substantial functional understanding of multiple treatment processes is especially notable with advanced treatment processes, where no studies were found that functionally analyzed advanced treatment trains for recycled water or potable reuse waters. It is worth noting that metagenomics has been applied to characterize the function of microbial communities associated with some emerging technologies in the water and wastewater field, specifically in the form of bioelectrochemical systems and microbial fuel cells (Kiseleva et al., 2015; Varrone et al., 2014; Yu et al., 2018). In general, future research would benefit from better coupling of functional profiles to operational conditions and water quality data. Collectively, available literature provides a good foundation for functional-based metagenomic analysis, with methods devoted to common metabolic pathways, nitrogen cycling, and general functional profiling being the most well-developed at this stage. Metagenomics aimed at elucidating functional capacity of microbial communities is an important contribution to the water industry as these methods facilitate mechanistic understanding of microbial processes that can support improved engineering of water and wastewater treatment processes.

2.2.3 Antimicrobial Resistance

In total, 56 studies were identified that applied NGS for examining AMR in relevant water environments. Of these, 54 utilized metagenomic sequencing, three utilized WGS, one utilized metatranscriptomics, and two utilized targeted sequencing to amplify the 16S rRNA gene to provide insight into the relationship between microbial community and ARG profiles. Wastewater, raw sewage and biosolids were the focus of the majority (34) of studies, followed by environmental source waters (34), drinking water treatment trains and distribution (7), and reclaimed water (2). Several common themes emerged, with most studies primarily focused on profiling "resistomes," (i.e., ARGs carried collectively across the microbial community) in both pathogenic and non-pathogenic bacteria. Additionally, studies applied WGS towards examining resistant isolates derived from water systems, while other studies systematically evaluated how water treatment mechanisms shape resistomes.

Several studies utilized metagenomics to probe wastewater effluents with specific emphasis on hospital and antibiotic manufacturing wastewaters as potential hotspots for ARGs (Baral et al., 2018; Chu et al., 2017; Fróes et al., 2016; Garrido-Cardenas et al., 2017; Guo et al., 2017; Gupta et al., 2018b; Hendriksen et al., 2019; Kumar et al., 2018; Lekunberri et al., 2018; Li et al., 2017; Liu et al., 2019c; Ng et al., 2017; Rowe et al., 2016; Schlüter et al., 2008; Szczepanowski et al., 2008; Tang et al., 2016; Yang et al., 2014a). Additionally these studies often used assembly of short reads to assess the co-occurrence of ARGs with MGE and metal resistance genes (MRG). This is important information in terms of assessing the potential for ARGs to be mobilized among bacterial populations or to be co-/cross-selected by other agents, such as metals. While most published studies to date employ short read sequencing, a few more recent studies have begun to apply long read sequencing. For example, Che et al. (2019) used Nanopore sequencing to identify ARGs associated with plasmids. The majority of studies have examined wastewater and the effects of effluent on receiving environments, but water reuse and drinking water are also of interest for AMR monitoring as potential direct routes of exposure (Chao et al., 2013; Dai et al., 2018b; Douterelo et al., 2018; Garner et al., 2018a; Jia et al., 2019; Ma et al., 2019; Yang et al., 2013). In one study combining metagenomics and metatranscriptomics, Liu et al. (2019c) noted that only 65 percent of ARGs identified by metagenomics in activated sludge were actively transcribed. The presence of unexpressed ARGs remains a concern, as they could still be expressed in clinical situations when antibiotics are applied, but such a study helps identify treatment processes that may also induce expression, and likely selection, of bacteria carrying ARGs. Identifying how specific wastewater and drinking water treatment processes impact the resistome was the focus of 8 studies. Several studies concluded that chlorination can selectively increase the relative abundance of ARGs compared to other genes (Chao et al., 2013; Jia et al., 2019; Kumar et al., 2018; Shi et al., 2013). Other studies have utilized metagenomics to characterize how activated sludge and bench-scale digesters can alter selection pressure with respect to ARGs encoding clinically relevant resistance (Bengtsson-Palme et al., 2016; Christgen et al., 2015; Yadav and Kapley, 2019; Zhang et al., 2015). Understanding how specific treatment processes affect resistomes is critical in generating practical recommendations for the water industry.

Sequencing resistant isolates derived from water systems was the focus of four studies, where WGS was applied towards characterizing ARGs and MGEs in bacteria such as *E. coli* (Cameron et al., 2019; Jiang et al., 2019; Parsley et al., 2010; Roy et al., 2018). In particular, WGS enables a high degree of confidence in assembly, which in turn improves the ability to characterize mechanisms of resistance and assess co-occurrence of ARGs with MGEs and other genes of interest. Coupled with phylogenetic analysis, WGS can be applied towards assessing mechanisms by which AMR is evolving across strains. Sequencing resistant *E. coli* and

Enterococcus, fecal indicators already in mainstream use by the water industry, may enable monitoring of background resistance and sequencing can help identify if new resistance patterns are emerging (Fróes et al., 2016).

While the use of NGS technologies is critical for the water industry in characterizing the breadth of AMR in water and wastewater systems that cannot be fully captured with culture or traditional molecular methods, there are still critical knowledge gaps regarding how to best apply these technologies towards guiding management of these systems for the purpose of mitigating the spread of AMR elements. Guidance is still needed to assess the utility of various NGS-derived metrics and the extent to which they provide information about ecologically important processes or represent risk to human health. For example, the relative abundance of total ARGs is thought to be indicative of the degree to which environmental conditions select for carriage of ARGs across a microbial community, while absolute abundance of total ARGs is arguably a more direct indicator of magnitude of health risk. Additional guidance is needed in terms of how to most meaningfully normalize metagenomic data in a manner that enhances quantitative capacity and comparability across systems. Notably, conclusions with respect to relative comparisons across samples can be substantially influenced as a function of sequencing platform, sequencing depth, and sample type. Many published studies employ assembly of short reads to assess co-occurrences of ARGs, MGEs, MRGs and pathogens, but the lack of a means to verify if assemblies are correct is a major shortcoming. Additionally, the quantitative value of metagenomic data is lost after assembling, though hybrid approaches are also available that rely on mapping of short reads to assembled or long reads. Thus, long read sequencing which eliminates the need for assembly holds particular promise for AMR monitoring, vastly increasing confidence in identification of bacterial hosts of ARGs and their co-occurrence with MGEs, MRGs, or other genes of interest. WWTPs in particular have been identified as key nodes both for the control and dissemination of AMR, and thus are an ideal point for surveillance of AMR in a community and mitigation of ARGs prior to release to the environment. Systematic evaluation of the efficacy of various treatment technologies for addressing AMR would be beneficial.

2.2.4 Bacterial Toxicity

In total, 24 studies were identified that apply NGS for studying bacterial toxicity in relevant water environments. Of these records, nine utilized metagenomic sequencing, 13 utilized targeted sequencing of the 16S rRNA gene, and two utilized both metagenomic sequencing and 16S rRNA gene amplicon sequencing. Wastewater and drinking water were the focus of the majority (6 each) of studies, followed by groundwater (5), aquifer (2), surface water (2), sediment (2), and constructed wetlands (1). The toxic compounds most frequently studied were free chlorine (5), arsenic (5), other heavy metals (5), nanomaterials (2), and multiple other compounds. Twenty-two studies met the criteria for both toxicity and AMR searches and were addressed above in the AMR section.

The effects of toxic contaminants in water sources and how they shift the microbial community were the focus of 13 studies. These studies examined both natural and anthropogenic pollutants. Heavy metals (Costa et al., 2015; Hemme et al., 2010), arsenic (Cai et al., 2013; Das et al., 2017; Jiang et al., 2019; Layton et al., 2014), and hydrocarbon (Abbai and Pillay, 2013;

Fahrenfeld et al., 2017) impacts were often studied in ground and surface water as well as stream sediments. Anthropogenic contaminants, such as nanomaterials (Binh et al., 2014; Liu et al., 2019b), tetrachloroethene (Reiss et al., 2016), and combinations of multiple pollutants (Lu et al., 2017; Sonthiphand et al., 2019) were examined in aquifers, constructed wetlands, and surface water. In general, these studies examined shifts in microbial community composition and function that occurred due to exposure to pollutants. For example, Fahrenfeld et al. (2017) applied metagenomics to investigate how the microbial community shifted along a stream running through an unconventional oil and gas disposal facility. The authors found that this wastewater release was associated with shifts in the microbial community structure and functions, including “functions related to dormancy and sporulation, respiration, and antimicrobial resistance”. Similarly, Sonthiphand et al. (2019) applied 16S rRNA amplicon sequencing to investigate the impact of anthropogenic activities on the diversity, abundance, and dynamics of microbial communities in groundwater. The authors found that microbial communities tended to cluster by the nature of the impact, specifically adjacent landfills, agricultural land, and community areas. These two examples provide helpful background understanding with respect to how toxins or pollutants can impact microbial communities associated with drinking water sources.

Examining microbial communities in drinking water treatment plants, WWTPs, and distribution or collection systems was the focus of nine studies. For example, some of the studies involving drinking water treatment plants examined the microbes within sand filters (Bai et al., 2013; Huang et al., 2014) and after disinfection (Huang et al., 2014); results indicated that microbial activity is important in removing harmful materials such as heavy metals, arsenate, and aromatic compounds. Free chlorine disinfection was found to reduce certain genera while enriching others as well as MGEs carrying virulence factors (Huang et al., 2014). Several studies have profiled how toxic compounds impact microbial communities in DWDSs (Douterelo et al., 2018; Saleem et al., 2018; Shaw et al., 2015), in wastewater reclamation and distribution systems (Lin et al., 2016), and in sewer biofilms (Gomez-Alvarez et al., 2012). For example, Douterelo et al. (2018) applied metagenomics to investigate the taxonomy and gene functions characteristic of biofilm and bulk water within a chlorinated DWDS. The authors identified “mechanisms of resistance and damage-repair to external stressors such as chlorine and antibiotics” within the biofilms. This study is an example of how metagenomic studies of DWDSs may help indicate infrastructure or treatment failures, thereby protecting and promoting water quality and safety.

One study was identified by the toxicity search criteria, but did not fall within any of the aforementioned sub-categories. Handley et al. (2014) assembled the whole genome of *Candidatus Sulfuricurvum* sp. from an aquifer-derived metagenome. While this approach has been employed to assemble complete genomes of uncultured microorganisms in other environments, this endeavor is inherently limited by achievable sequencing depth (Alneberg et al., 2018). The authors identified genes indicative of heavy metal and arsenic tolerance, which they presumed to be related to heavy metal contamination within the aquifer.

In summary, NGS studies involving antimicrobial or toxic compounds in water primarily focused on how these compounds shifted the microbial community in contaminated water environments or within the drinking or WWTP or distribution/collection systems. Of particular relevance to the water industry is new understanding being gained by examining the impact of free chlorine, primarily in drinking water disinfection.

2.2.5 Cyanobacteria and Harmful Algal Blooms

In total, 29 studies were identified that applied NGS to study cyanobacteria in freshwater, drinking water, and wastewater. Of these, 6 utilized metagenomic sequencing, 4 utilized metatranscriptomic sequencing, 18 utilized targeted sequencing of the 16S rRNA gene, 2 utilized both targeted and metagenomic sequencing, and 1 utilized both metagenomic and metatranscriptomic sequencing. Freshwater was the focus of the majority (20) of studies, including lakes (18), rivers (1), and reservoirs (1), followed by drinking water (6), wastewater (3), and aquatic biofilms (1). Several common themes emerged as ways NGS has been used to study cyanobacteria in relevant water environments, including characterization of cyanobacterial communities, especially toxic strains, in source water and water treatment systems, identification of shifts in gene expression, and control and removal of cyanobacteria in wastewater or drinking water.

Targeted sequencing with an emphasis on 16S rRNA gene and metagenomic sequencing were the main NGS approaches used to study the cyanobacterial genome and its position in the microbial communities relevant to freshwater (Abia et al., 2018; Bakal et al., 2019; Driscoll et al., 2017; Eiler et al., 2013; Ghai et al., 2014; Hou et al., 2019; Kurilkina et al., 2016; Kurobe et al., 2018; Lee et al., 2017; Pope and Patel, 2008; Qin et al., 2019; Reza et al., 2018; Saleem et al., 2019; Steffen et al., 2015; Vadde et al., 2019), drinking water (Kori et al., 2019; Otten et al., 2016; Pei et al., 2017; Saleem et al., 2018; Xu et al., 2018), or wastewater (Lee et al., 2017; Lu and Lu, 2014). For example, Kurobe et al. (2018) applied metagenomic sequencing to investigate the shift in the cyanobacterial community from freshwater to brackish water during a *Microcystis* bloom in the San Francisco Estuary. They found *Microcystis* was the predominant genus among cyanobacteria in both freshwater and brackish water, with up to six *Microcystis* genotypes identified. These studies emphasize the importance of NGS approaches for tracking cyanobacterial communities, especially toxic and bloom-forming genera, in freshwater systems. In another exemplar, Otten et al. (2016) utilized metagenomic sequencing to identify taste-and-odor producers and toxin-producing cyanobacteria in a drinking water reservoir. Their results indicated that *Anabaena* spp., *Microcystis* spp., and "an unresolved member of the order *Oscillatoriales*" are potentially responsible for some chemicals that produce taste or odor, including geosmin, microcystin, and 2-methylisoborneol. This research highlighted the need to apply NGS approaches to monitor cyanobacteria that have a negative impact on the drinking water quality.

Identification of shifts in gene expression associated with metabolic pathways was the focus of seven studies (Chen et al., 2018; Davenport et al., 2019; Hampel et al., 2019; Harke et al., 2016; Mou et al., 2013; Pascault et al., 2014). Most of these papers emphasized cyanobacterial gene expression at different times and under different nutrient conditions during a bloom season. For example, Davenport et al. (2019) observed "diel shifts in metabolic pathways of *Microcystis* spp. during a 48-h survey" using metatranscriptomic sequencing. Daytime gene transcripts favored photosynthesis and nutrient uptake, whereas nighttime transcripts were primarily related to protein synthesis (Lee et al., 2017). Similarly, Hampel et al. (2019) analyzed shifts in gene expression using metatranscriptomic sequencing under a high-nitrogen condition in early summer and a low-nitrogen condition in late summer, and found toxic *Planktothrix* and *Microcystis* relied heavily on regenerated ammonium when the nitrogen level became low in late summer. These studies addressed the need for NGS approaches to track the impacts of different time scales and nutrient conditions on the evolution of the cyanobacterial community and their gene expression levels and suggested potentially useful nutrient control methods to reduce toxic cyanobacteria in freshwater and water treatment systems.

Characterization of cyanobacteria in source water was the focus of 16 studies (Abia et al., 2018; Bakal et al., 2019; Driscoll et al., 2017; Eiler et al., 2013; Ghai et al., 2014; Hou et al., 2019; Kurilkina et al., 2016; Kurobe et al., 2018; Lee et al., 2017; Lu and Lu, 2014; Pope and Patel, 2008; Qin et al., 2019; Reza et al., 2018; Saleem et al., 2019; Steffen et al., 2012; Vadde et al., 2019; Xu et al., 2018), while an additional five looked at cyanobacteria in water treatment systems (Kori et al., 2019; Otten et al., 2016; Pei et al., 2017; Saleem et al., 2018; Xu et al., 2018). Control and removal of cyanobacteria in wastewater or drinking water was the focus of two studies (Bai et al., 2014; Zamyadi et al., 2019). These studies utilized NGS to investigate the influence of different water treatment processes on cyanobacterial communities in the relevant water environment. Bai et al. (2014) applied both targeted amplicon sequencing and metagenomic sequencing to assess the impacts of treated and untreated wastewater discharges on bacterial communities in the river, and concluded untreated water resulted in an increase of the relative abundance of cyanobacteria. Zamyadi et al. (2019) proposed a critical control point for cyanobacteria removal in the drinking water treatment process. They applied NGS along with microscopic analysis and cell integrity methods, and found after pre-oxidation treatment, the "remaining cyanobacterial cells (~80%) were undamaged with the potential to accumulate and grow within the plants post-$KMnO_4$ treatment, particularly in clarifier sludge." Both studies highlighted the importance of water treatment to control and remove potentially toxic cyanobacteria in the water systems and the need to apply NGS to improve the monitoring and management of water quality.

In summary, the focus of the majority of cyanobacterial studies using NGS approaches is on the characterization of cyanobacterial communities in both source water and water treatment systems, followed by evaluating of gene expressions. Key questions investigated included identification of community structure and toxic strains during bloom seasons and shifts in gene expressions under different environmental and nutrient conditions. The remaining studies, not covered in detail in this section, only discussed cyanobacterial metagenomics as part of the larger bacterial or prokaryotic community, rather than being the main focus.

2.2.6 Characterization of Viruses

In total, 80 records were identified that applied NGS for studying viruses in wastewater, recycled water and freshwater. Of these records, 54 utilized metagenomic sequencing, 12 utilized WGS, and 14 utilized targeted sequencing (5 targeting the hypervariable human adenovirus hexon gene, 8 targeting specific enteroviruses, and 1 targeting phage-borne 16S rRNA sequences). Wastewater was the focus of the majority (49) of studies, followed by freshwater (30), including lakes (17), streams or rivers (4) and surface water (3), groundwater (1) and reclaimed water (1). Several common themes emerged from the systematic review of how NGS has been used to study viruses in relevant water environments, including characterization of environments, method development and genotyping. Virome characterization of water environments was the focus of 54 studies. The majority of these papers reported viromes of previously unstudied environments (e.g., untreated sewage (Adriaenssens et al., 2018; Cantalupo et al., 2011; Ng et al., 2012; O'Brien et al., 2017; Strubbia et al., 2019a; Tamaki et al., 2012; Wang et al., 2018), dairy lagoons (Alhamlan et al., 2013), and freshwater lakes (Djikeng et al., 2009; Gong et al., 2016; Green et al., 2015; Gu et al., 2018; Hewson et al., 2018, 2012; Hornstra et al., 2019; Kavagutti et al., 2019; Malki et al., 2015; Mohiuddin and Schellhorn, 2015; Okazaki et al., 2019; Palermo et al., 2019; Rosario et al., 2009; Sible et al., 2015; Skvortsov et al., 2016; Tamaki et al., 2012; Tseng et al., 2013; Watkins et al., 2016)). Specific topics of interest within these studies included viral pathogen identification, DNA or RNA virome characterization, virus-host interactions, phage diversity, and ARG characterization within viruses. Tamaki et al. (2012) used NGS to characterize DNA viruses throughout the wastewater treatment process. This study highlighted the uniqueness of the virome of wastewater compared to 42 other environmental settings. Bibby and Peccia (2013) utilized NGS to characterize the virome of sewage sludge, with a focus on describing the diversity of human viruses and understanding the risks associated with land application. This study expanded the knowledge of human pathogenic viruses in wastewater samples and demonstrates that NGS can be used to identify human viruses in environmental samples.

Nine studies focused on method development, particularly improving methods to study viruses in aquatic environments (Bekliz et al., 2019; Brinkman et al., 2018; Fernandez-Cassi et al., 2018; Hjelmsø et al., 2017; Labonte, 2016; Oshiki et al., 2018; Strubbia et al., 2019b; Uyaguari-Diaz et al., 2016). These studies mainly focused on methods to concentrate, purify or extract viral nucleic acid from various water samples. For example, Hjelmsø et al. (2017) compared four viral particle concentration methods and four different extraction kits to evaluate viral community composition, specificity, richness and pathogen detection of raw sewage using NGS. Similarly, Fernandez-Cassi et al. (2018) developed an optimized ultracentrifugation method (or skimmed milk flocculation when ultracentrifugation is not available) to concentrate viruses and target adenovirus and other viruses that could not always be detected using metagenomics. This study addresses the sensitivity necessary for NGS to investigate the sewage virome to track human pathogens and highlights the importance of using viral metagenomics for public health surveillance.

Viral genotyping using next-generation amplicon or WGS was the subject of 12 studies (Boonchan et al., 2017; Brinkman et al., 2017; Fumian et al., 2019; Hata et al., 2018b, 2018b, 2018a; Kaas et al., 2019; Majumdar et al., 2018; Mancini et al., 2019; Ogorzaly et al., 2015; Oshiki et al., 2018; Suffredini et al., 2018; Yoshitomi et al., 2017). These studies used NGS to characterize the diversity of pathogenic viruses, including adenovirus, astrovirus, and norovirus, in wastewater or sewage-impacted surface water. For example, Hata et al. (2018a) used this approach to study occurrences of norovirus, sapovirus, rotavirus, Aichi virus and enterovirus;

they concluded that norovirus GII.17 strains were prevalent in surface water impacted by wastewater before becoming prevalent in gastroenteritis patients in Japan. Oshiki et al. (2018) used a microfluidic, nested PCR approach combined with NGS amplicon sequencing to detect and genotype 11 human pathogenic RNA viruses in human stool and sewage. These studies prove the usefulness of using amplicon-based NGS to study molecular epidemiology and monitor the emergence of viral pathogens using wastewater surveillance.

Undoubtedly, there are significant challenges associated with application of metagenomics for the study of viruses, compared to other microorganisms (Behzad et al., 2015; Prussin et al., 2014; Rose et al., 2016; Simmonds, 2015). The first challenge is that viruses do not share a conserved gene; thus profiling viromes can only be achieved via shotgun metagenomics, not targeted gene amplification (Ghurye et al., 2016). Shotgun metagenomics requires a substantial mass of genomic material, which is particularly difficult to obtain from viruses in environmental samples, in which concentrations are low (Behzad et al., 2015; Prussin et al., 2014). One way to overcome this challenge is by collecting large sample volumes (e.g., hundreds of liters) (Breitbart et al., 2002; Rodriguez-Brito et al., 2010; Rosario et al., 2009). However, this is both time-intensive and time-sensitive. Even with proper precautions during sampling and nucleic acid extraction, contaminants that could degrade genetic material may be introduced. Second, many viruses have RNA as their genetic material, which degrades much more rapidly than DNA. If the RNA degrades before sequencing, there would be false negatives in the virome analysis, and many potentially important viruses could be overlooked. The third significant challenge associated with viral NGS is that reference databases are very limited (especially compared to those for bacteria and fungi). Bibby and Peccia (2013) suggested that less than one percent of viruses have been sequenced and uploaded to databases and they demonstrated that over 75% of viruses in sewage sludge are unidentifiable. To improve viral databases, viruses must first be isolated, cultured, and subjected to WGS (Prussin et al., 2014; Rohwer and Edwards, 2002). However, viruses are very difficult to culture because they require a host (e.g., bacteria for bacteriophages) and some viruses are simply 'unculturable' using current tools and approaches. Finally, there is no "standard" user-friendly data analysis pipeline for viromes, as exists for bacterial/archaeal 16S rRNA genes (e.g., QIIME). Most virome pipelines require extensive bioinformatics and/or computer science backgrounds. With improved tools, knowledge of viruses in water and wastewater using NGS will rapidly improve. However, presently the understanding of viruses is fragmented, especially in water and wastewater. As such, viruses have been described as "the forgotten siblings of the microbiome family" (Williams, 2013).

2.3 Knowledge Gaps and Research Needs

2.3.1 Need for Standardized Methods

The field of NGS is rapidly evolving, with a continually emerging pipeline of new platforms, library preparation techniques, and data analysis approaches. Such an expanse of evolving options creates challenges to ensuring that data collected across studies is comparable, which is critical for informing broader conclusions across studies. Differences in sample volume, DNA extraction method, library preparation method, NGS platform, read length, sequencing depth, and data quality filtering are all critical aspects of NGS that are not standardized and make it difficult to compare results. Thus, standardized methods are critically needed to overcome key differences in approaches, as well as to facilitate the adoption of NGS technologies by new stakeholders, such as water utilities (Birko et al., 2015). However, the development of standardized methods for NGS application for studying water and wastewater should allow flexibility for the design of research studies aimed at asking novel questions that may not strictly conform to standard approaches. As the utility of NGS to different sectors of the water and wastewater industry increases, several agencies could potentially play a role in standardization of these methods for application in environmental samples, for example the U.S. Environmental Protection Agency, U.S. Centers for Disease Control and Prevention, World Health Organization, National Institute of Standards and Technology, and Clinical and Laboratory Standards Institute.

2.3.2 Sample Processing

Substantial variability in methodology exists for sample collection, concentration, and extraction of DNA and RNA and can have critical impacts on downstream NGS analysis. While the type of membrane used for sample concentration has been shown to have no discernable impact on microbial community structure (Djurhuus et al., 2017), there are likely to be substantial differences based on whether hollow-fiber filtration or precipitation and centrifugation is used compared to membrane filtration. Past studies have demonstrated that different DNA extraction kits used with water and wastewater samples tend to produce microbial communities and resistomes that are largely similar, but inconsistencies can be introduced by certain extraction kits or methods due to contamination or methodologies that favor certain groups of organisms (i.e., Gram negative vs. Gram positive) (Djurhuus et al., 2017; Li et al., 2017; Walden et al., 2017). Sample matrix also likely has an impact on the optimum extraction method, given that separation of microorganisms from complex organic matrices and carrier materials varies between matrices (Sanz and Köchling, 2019). Biomass is expected to vary substantially by sample matrix and low-biomass environments, such as drinking water, are likely to require more extensive concentration than more microbially dense environments. This introduces challenges for comparison across water environments (i.e., water, wastewater, sediment, biofilm, etc.). Standardization of pre-processing methods to collect and prepare samples for NGS is critical for producing comparable results across water and wastewater systems.

2.3.3 Bioinformatic Analysis

The continuous optimization and declining costs of sequencing platforms have facilitated generation of a vast volume of sequencing data. However, rendering these deposited data into meaningful results that can inform decision-making in the water industry is an ongoing challenge. The analysis of NGS data largely depends on principles and standards established by the broader genomics community, such as sequencing protocols and controls, read mapping algorithms, assembly algorithms, methods for phylogenetic analysis, and single nucleotide polymorphism (SNP) genotyping. However, each step poses a variety of challenges that are briefly described below.

Quality control (QC) is one of the most critical steps in the processing of NGS data. It usually involves removing or trimming low-quality reads, and the unwanted sequences (such as sequencing adapters, host contamination) that could pose problems in the downstream analysis. As different datasets have their own characteristics and challenges, it is advised to formulate the QC protocol that is best suited for the specific dataset in consideration. For example, metatranscriptome datasets could be contaminated with rRNA transcripts that needs to be removed prior to any downstream analysis (Esteve-Codina, 2018). Similarly, the QC tools framed for reads generated on short read platform may not be appropriate for reads generated on long reads platform due to the inherent differences in the two sequencing technologies (Fukasawa et al., 2020).

The traditional approach used for collectively detecting genes across microbial communities is using read alignment tools to annotate genomic or metagenomic data and applying a similarity search against a reference sequence using a best-hit approach (Wilke et al., 2013). The basic principle lies in the assumption that genes sharing homology perform similar functions. However, even a single bp change in sequence can alter protein function, thus the choice of stringency used in homology-based annotation can substantially affect the interpretation and reproducibility of the data (Randle-Boggis et al., 2016). For example, a lower percentage identity cutoff will fail to differentiate two highly similar sequences. While the best hit approach remains the most popular annotation strategy, it is prone to producing a high rate of false positives (Randle-Boggis et al., 2016). Sequencing aimed at characterizing SNPs is still an under-explored topic as it requires cost-intensive deep sequencing with high coverage of sequenced genomes. *De novo assembly* (i.e., assembly conducted based on overlapping reads, rather than mapping reads to reference genomes) holds promise for overcoming the high false positive rates associated with the best hit approach applied to short reads. While various assemblers are available to tackle complex NGS data, it is advised to be cautious while selecting the appropriate assembler as different datasets (metagenomics, metatranscriptomics) have their own challenges (Shakya et al., 2019). Further, assembly is computationally expensive and often generation of contigs of sufficient length for target analyses is limited when faced with high microbial diversity characteristic of water samples. Often too few contigs are produced and with low confidence in accuracy, diminishing the ability to carry out meaningful downstream analyses. Genome-resolved assembly of metagenomic sequencing data, though limited by the ability to obtain sufficient coverage, holds promise for a wide range of applications, including the ability to construct genomes of uncultured organisms and linking metabolism and function to phylogeny (Quince et al., 2017a).

The vast quantities of sequencing data being produced also create challenges for computational demands, data storage, and data security. Although public repositories such as the NCBI Sequence Read Archive and the European Nucleotide Archive are readily used to publicly store and share metagenomic data, substantial computational resources are still required to carry out in-house analyses, which can be a significant challenge to the application of NGS by the water industry. One possible solution to tackle this issue is to establish a cloud-based infrastructure, which is readily used by many biotechnology companies. NGS holds the capability of revolutionizing the water industry by providing a plethora of tools and frameworks to analyze water and wastewater microbial communities. However, owing to the daunting complexity of metagenomics, an interdisciplinary work force composed of engineers and bioinformaticians will likely be needed to perform these analyses and inform decision-making. Building such a work force would be essential in implementing and maintaining NGS capabilities in the water industry.

2.3.4 Availability of Annotation Databases

There are numerous well-curated databases available for annotating NGS reads, as demonstrated in Table 2-2. While these databases are useful for annotating microbial phylogenetic origin and for key functional capacities, such as antibiotic resistance, many key attributes are missing from the available databases. For example, very little information is available regarding non-bacterial microorganisms, such as viruses, amoebae, and fungi, which are often highly important in water and wastewater. In addition, few databases are available that curate metabolic genes, making it difficult to screen samples for functional capacity. The need for well-curated databases is a critical limitation that often encumbers the study of emerging topics via NGS technologies.

Table 1-2. Available Databases and Workflows for Annotation of Microorganisms and their Functional Capacity.
Source: Reprinted from Garner et al. 2021 with permission from Elsevier.

Microbial Target Type	Databases/Workflows
Bacterial Taxonomy & Phylogeny	Greengenes (DeSantis et al., 2006), SILVA (Quast et al., 2013), MetaPhlan2 (Truong et al., 2015), Ribosomal Database Project (RDP) (Cole et al., 2014), Genome Taxonomy Database Toolkit (GTDB-Tk) (Chaumeil et al., 2020), NCBI GenBank, NCBI RefSeq
Pathogen Identification	PATRIC (Wattam et al., 2017), NCBI Pathogen Detection, EuPathDB (Warrenfeltz et al., 2018), MyPathogen Database (MPD) (Zhang et al., 2018)
Antibiotic Resistance	Comprehensive Antibiotic Resistance Database (CARD) (Alcock et al., 2020), Functional Antibiotic Resistance Metagenomic Element Database (FARME-DB) (Wallace et al., 2017), Resfams (Gibson et al., 2015), ResFinder (Zankari et al., 2012), deepARG (Arango-Argoty et al., 2018), ARGminer (Arango-Argoty et al., 2020)
Mobile Genetic Elements	A CLAssification of Mobile genetic Elements (ACLAME) (Leplae et al., 2010), INTEGRALL (Moura et al., 2009), Isfinder (Siguier et al., 2006), The Transposon Registry (Tansirichaiya et al., 2019), ICEberg (Liu et al., 2019a), The Gypsy Database (GyDB) (Llorens et al., 2011)
Metal and Biocide Resistance	Antibacterial Biocide & Metal Resistance Database (BacMet) (Pal et al., 2014)
Metabolism	Carbohydrate-Active Enzymes Database (CAZy) (Lombard et al., 2013), Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa et al., 2017), evolutionary genealogy of genes: Non-supervised Orthologous Groups eggNOG (Huerta-Cepas et al., 2019), BioCyc (Karp et al., 2019), MetaCyc (Caspi et al., 2020)
Protein Function	Clusters of Orthologous Groups of proteins (COG) (Tatusov et al., 2000), SEED (Overbeek, 2005), NCBI RefSeq (O'Leary et al., 2016), UniProt (The UnitProt Consortium, 2019), KEGG (Kanehisa et al., 2017), Pfam (El-Gebali et al., 2019)

2.3.5 Determination of Viability

As with all techniques targeting DNA, most NGS techniques are unable to directly distinguish DNA originating from live versus dead cells. Metatranscriptomics has been used to overcome this obstacle by targeting mRNA rather than DNA, given that mRNA is produced only by live cells and associated only with genes that are actively transcribed (Moran, 2009). However, this approach is particularly challenging for analyzing low biomass water samples such as those collected from drinking water, which typically contain very low RNA concentrations. RNA also degrades rapidly in the environment and is prone to contamination by hosts (e.g., human cells in wastewater) (Bashiardes et al., 2016). The application of membrane-impermeable intercalating dyes, such as propidium monoazide and ethidium monoazide, have also been used to address challenges in detecting live vs. dead cells. While these methods initially showed great promise and raised significant interest among researchers for differentiating live and dead cells (Bae and Wuertz, 2009; Chen et al., 2011; Hellein et al., 2012; Vesper et al., 2008; Yáñez et al., 2011), many have highlighted critical limitations and shortcomings with the approach, such as a lack of reproducibility (Scaturro et al., 2016; Tavernier and Coenye, 2015; Taylor et al., 2014), matrix interference (Taylor et al., 2014), need for optimization according to target organism and sample characteristics (Taylor et al., 2014), and ability to penetrate some intact cells (Flekna et al., 2007; Kobayashi et al., 2009; Nocker et al., 2006). Flow cytometry using cell sorting for differentiating live versus dead cells has been employed and paired with NGS in some studies (Berney et al., 2007; Hammes et al., 2008; Kahlisch et al., 2010). Use of NGS methods may be more relevant for studying certain water environments than others. For example, immediately following a disinfection process, it is likely that a large portion of the genetic material present will be associated with inactivated cells or persisting extracellularly. In contrast, NGS is more promising for studying microorganisms in environments with more stable or slowly changing conditions affecting cell viability, as would be expected in most other stages of water and wastewater treatment, or in source waters and distribution systems. Improved methods are needed to accurately determine the viability of NGS targets.

2.3.6 Quantitative Capacity of NGS

The quantitative capacity of NGS is another key dimension of importance for application in water and wastewater. The ability to directly quantify targets is generally limited by sample and library preparation and on the normalization of samples to equal mass. Further, during sequencing, there is inherent variability in the number of reads obtained among samples. Therefore, with metagenomics, metatranscriptomics, and targeted amplicon sequencing, quantitative evaluation is usually restricted to a comparative fashion, in terms of “relative abundance,” i.e., by normalizing NGS data to a secondary metric internal to each sample, such as number of reads, percentage of total reads, or number of reads annotated as a reference gene, such as the 16S rRNA gene (Nayfach and Pollard, 2016; Weiss et al., 2017). While relative abundances are highly informative and provide valuable information for understanding the composition and functional capacity of microbial communities, in many cases, it is often desirable to be able to measure microbial targets in terms of absolute abundance. This is particularly important for informing risk assessment or for assessing infectious doses. Normalized NGS data has been transformed into absolute abundance data by pairing NGS measurements with independent quantifications of total cells or total 16S rRNA genes (Garner et al., 2016; Vandeputte et al., 2017). While this approach is promising, further validation is needed.

Achieving an appropriate sequencing depth and addressing differences in sequencing depth are also important for accurate and comparable quantification of microbial targets. There are currently no standards for establishing the necessary depth or coverage of sequences for water or wastewater samples, representing a key knowledge gap for facilitating comparison of samples across projects and ensuring sufficient reads are generated for each sample to capture genes of interest. Coverage and associated sequencing depth are likely to vary substantially based on the target environment (i.e., drinking water, wastewater, etc.) as well as the target research question. There are also a variety of approaches to address differences in sequencing depth among samples, such as through application of rarefaction, though no standards or consensus on appropriate methodologies currently exists (McMurdie and Holmes, 2014; Weiss et al., 2017).

2.4 Conclusion

NGS technologies are revolutionizing microbial monitoring relevant to the water and wastewater industry, including improving the ability to investigate topics such as taxonomic classification, functional and catabolic gene characterization, AMR, bacterial toxicity, cyanobacteria and harmful algal blooms, and characterization of viruses. NGS methods have been widely adopted for water research and are being translated to various applications in the water industry, with new applications continuously emerging. NGS has also recently been used to study the genome of SARS-CoV-2 in wastewater (Nemudryi et al., 2020; Rimoldi et al., 2020). While the application of NGS in water and wastewater practice presents many challenges such as cost, need for specialized expertise and equipment, challenges with data analysis and interpretation, lack of standardized methods, and the rapid pace of new technological developments, the synthesis provided herein of the myriad of ways NGS tools can be applied helps pave a path forward for practical application of NGS for the water industry. In sum, NGS has already catalyzed transformative new understanding of the role and activities of microbes in water systems and, with appropriate attention to addressing current limitations, it is well-poised for deeper integration into water industry practice and can aid in addressing numerous current and future water-related challenges.

CHAPTER 3

Utility and Industry Stakeholder Perspectives on NGS

3.1 Overview and Approach

Nine industry stakeholders were interviewed to gain insight into their familiarity with NGS, potential applications of NGS, and obstacles to implementation. Interviewees included seven personnel from U.S. drinking water, wastewater, or recycled water utilities. Two were other industry stakeholders who work closely with utility personnel in consulting roles. Table 3-1 summarizes the topics and specific questions that were utilized in these interviews. Interview findings have been synthesized and are summarized below.

Table 3-1. Questions and Prompts Utilized in Stakeholder Interviews.

Topic	Questions/Prompts
Familiarity with NGS	Describe the level of familiarity with NGS technologies among your utility's staff.
	Has your utility conducted NGS or worked with a collaborator to do so?
Applications of NGS	What applications of NGS interest your utility?
	What microbial challenges does your utility face that you need more resources to address or understand?
Protocols	Does your utility conduct DNA extraction? If so, what approach do you use?
	What level of familiarity does your utility personnel have with molecular techniques?
	How much time of your utility's personnel would you be willing to commit to learning to use NGS technologies?
	How much time of your utility's personnel would you be willing to commit to using NGS technologies on a routine basis?
Obstacles to Implementation	What are the key obstacles to implementing NGS at your utility?

3.2 Familiarity with NGS

The interviewees represented a range of expertise and experience with NGS, ranging from those who were unaware of the technology prior to the interview to extensive experience, with some utilities interviewed having conducted NGS methodologies in-house (Figure 3-1). All seven utilities had previously collaborated with academic partners or research institutes to conduct NGS aimed at characterizing water and wastewater samples. While others had utilized contract laboratories to conduct sequencing and data analysis, one had conducted NGS in-house, and one had recently purchased an NGS instrument with plans to initiate sequencing in-house soon. While several utilities had a team of scientists who were engaged with in-house or external application of NGS, several interviewees noted that they were the only member of the utility staff who were knowledgeable about NGS. One interviewee noted that the recent emergence of NGS impacts who on the utility staff is familiar with the methods, saying "For the utility staff that have been doing their job 15-20 years, even the microbiologists don't have familiarity [with NGS]. Some of the newer engineers are starting to at least be aware of the technologies." Interviewees also noted increasing awareness, which much more of the utility staff aware of NGS applications in water and wastewater now than just a few years ago. The interest and perceived ability of utilities to conduct NGS on-site or through contract laboratories varied widely. While some interviewees were eager to explore the ways that NGS could be applied to better understand and resolve water quality and operational challenges, others felt that NGS was not a priority. One utility staff noted, “It is important for some utilities to go first, because others will want or need to hang back and wait for standardized or established methods to be developed.” In contrast, a staff member from another utility noted that their utility would be unlikely to devote substantial time and resources to NGS unless it becomes part of the regulatory compliance requirements.

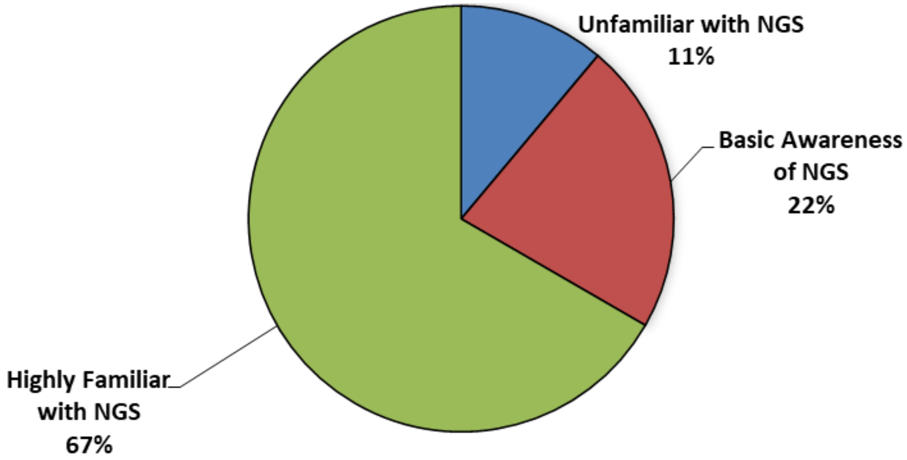


Figure 3-1. Summary of the Previous Familiarity of Stakeholder Interviewees with NGS (n=9).

3.3 Potential Applications

Interviewees noted numerous potential NGS applications of interest to their utility (Figure 3-2). Among those that were most frequently proposed were monitoring of SARS-CoV-2 and other viruses, pathogen tracking, and understanding overall microbial communities throughout treatment processes and distribution. A key area of interest was establishing long-term monitoring campaigns to better understand the baseline microbial community throughout treatment. This information could be valuable to elucidate when changes to microbial community composition may be indicative of process upset. Several utilities also noted a desire to use NGS to better understand microorganisms of concern beyond singularly focusing on bacteria, such as focusing on viruses and protozoa as targets of interest. Another common theme was the potential to use NGS to surpass the limitations of existing indicator methods, such as coliforms and *E. coli*, and identifying improved water quality surrogates. This was particularly attractive to utilities with an interest in direct potable reuse, where *E. coli* monitoring is of limited value. Some utilities also noted that NGS could be valuable as a screening tool, to inform when targeted methods for monitoring pathogens are needed. Other applications of interest included antibiotic resistance, changes during drinking and reclaimed water distribution (especially nitrification), source water quality, impact of outfalls on surface water quality, harmful algal blooms, foaming, biogas production, nutrient removal, bioprospecting, and *Legionella*.

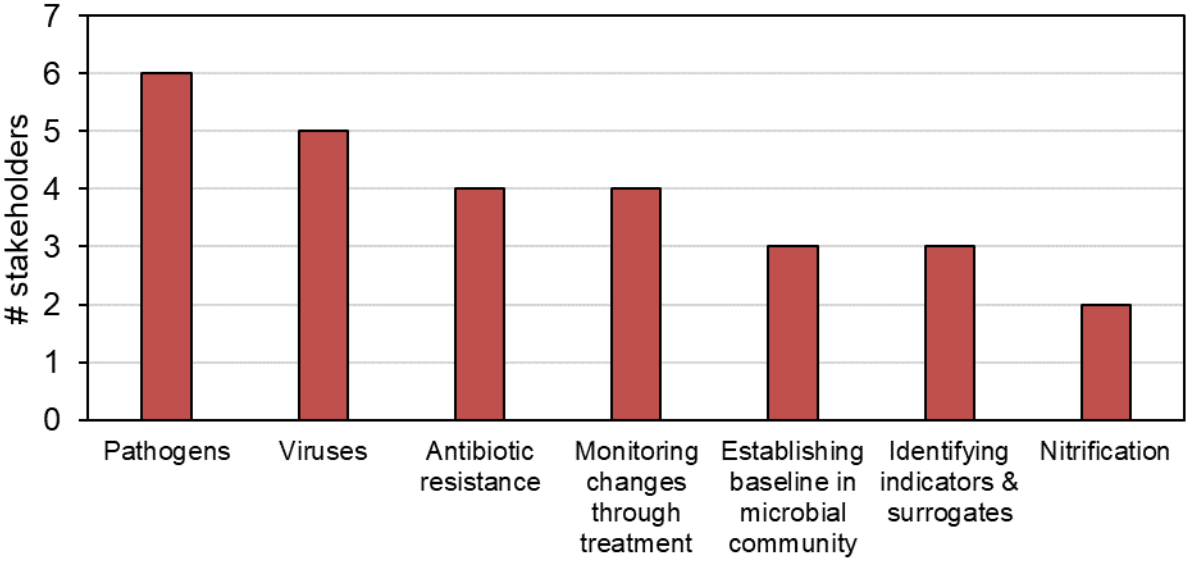


Figure 3-2. Key NGS Application Areas of Interest Communicated in Utility/Stakeholder Interviewees. Applications Proposed by Two or More Utilities were Included in this Synopsis.

3.4 Challenges and Obstacles to Implementation

A variety of obstacles were noted as potential barriers to implementation of NGS technologies in the water and wastewater industry (Figure 3-3). Interviewees noted a need for standardized methods, with multiple noting that it is important to recognize that the same methods will likely not apply across the spectrum of water, wastewater, and recycled water. They also noted that this applies to data analysis and interpretation as well, with guidance needed for how metagenomic data can be interpreted in a nuanced way. For example, what does detection of reads at very low quantities associated with pathogens mean in each of these cases? Challenges associated with data analysis and interpretation were a common concern among the interviewees. Many noted that utility staff typically do not have the bioinformatics training to do these analyses in house but that relying on outside contractors to assist with interpretation of findings can be problematic for understanding findings with respect to the water and wastewater regulatory structure. For example, some interviewees noted a lack of clarity for interpreting detection of pathogen-associated DNA sequences compared to culture-based methods. Other concerns that utilities noted as obstacles to their NGS implementation include collecting sufficient biomass for low concentration samples, lack of instrument access, presence of inhibitors and other matrix effects, cost, lack of quality assurance and quality control, and lack of continuing education for utility staff on the topic.

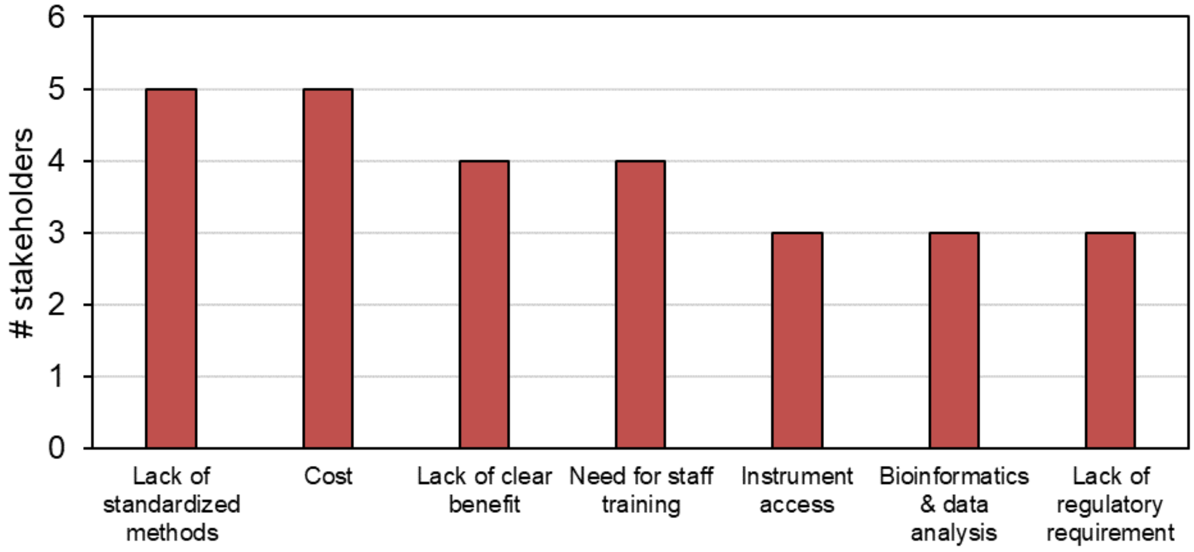


Figure 3-3. Key Obstacles to Implementation of NGS Communicated in Utility/stakeholder Interviews.
Obstacles Identified by Two or More Utilities were Included in this Synopsis.

3.5 Summary

While the familiarity of utility personnel and other stakeholders with NGS methodologies is generally limited, collaborations with academic partners and other research institutes are a key mode by which utilities have engaged with NGS-based applications to better understand their treatment and conveyance systems. Key applications of interest for utility personnel and water industry stakeholders include potential health threats; such as detection of pathogens, viruses, and antibiotic resistance. However, there are substantial challenges to implementation of NGS techniques for water utilities at this time. Chief among these challenges are a lack of standardized methods, cost, need for clearly defined benefits to motivate investment in these techniques, and a need for staff training.

CHAPTER 4

Available Methodologies and Technical Considerations for the Application of NGS to the Study of Water and Wastewater

4.1 Sample Collection

Specific scientific aims and budgetary constraints must be considered when deciding on sample collection methods. Established scientific aims and research questions will help to inform methods that will be needed for sampling. Research questions may be hypothesis driven, however, non-hypothesis driven research questions can be especially relevant for NGS. Some non-hypothesis questions may aim to improve tools (e.g., processing pipelines, reference databases) or standards (e.g., used to determine accuracy of results through the use of reference standards), may ask questions focused on spatial maps (how does a sample change over space or time) and abstract maps (principal coordinates analysis, non-metric multidimensional scaling) (Tripathi et al., 2018). Furthermore, how NGS helps answer one's research questions is important to consider as it could be that another established method (culturing, PCR, etc.) might be more appropriate or provide useful complementary information to help interpret NGS analysis. How one proceeds from the very start, sampling, to the very end of the analysis, is going to be dictated by those research questions, which is why they should be considered seriously before starting a project.

Prior to sampling, it is essential to develop a sampling and analysis plan. Guidance documents published by agencies such as the United States Environmental Protection Agency (US EPA) and procedures reported by others in the scientific literature can provide a useful reference. The sampling and analysis plan should be tailored to the questions and should consider which environments should be sampled, where and how many samples should be collected, what quality control measures should be included, and/or what metadata should be collected (EPA, 2014a)? Additional questions that should be considered before sampling can be found in Table 4-1. Note that special attention is necessary to preserve the integrity of microbial samples. Specifically, it is necessary to consider where the microorganisms or processes of interest are located, how to handle microbial samples and ensure their preservation, and how much mass or volume may need to be sampled to obtain adequate DNA or RNA for sequencing. It is also good practice to include negative (blank) and positive (mock community) controls during field sampling, sample processing, and throughout further processing steps, especially when low concentrations of DNA are anticipated. Inclusion of such controls help to distinguish true NGS signal representative of the sample from any background contamination. This section will review some general sampling approaches; however, current literature should be reviewed to ensure the most relevant and up-to-date operating procedures are employed for the environment sampled.

Sampling methods for metagenomic sequencing are not often established or described in sufficient detail in the literature (Mthethwa et al., 2021). Therefore, methods described below may be adapted from those used to sample the environment for other molecular techniques (e.g., qPCR) or culture-based methods. Sampling for NGS will typically follow similar methods, with the added nuance that a greater mass or volume may need to be collected to ensure there is sufficient genetic material at the end of the extraction process for successful sequencing.

Table 4-1. Example Research Questions to Consider Before Sampling.

Topic	Questions to Consider
Selecting methods that are appropriate to the research question	Is NGS the best way to answer these question(s) or are there other established methods (culturing, qPCR) that are more appropriate/easier/more affordable? What other tests or analyses might be needed in addition to NGS? Examples include analysis of nutrients, measuring viable microorganisms, quantification of certain genes (e.g., 16S rRNA genes or another gene quantified for normalization purposes) What additional metadata should be collected to help answer the research question(s), for example: Temperature and weather conditions Water quality measurements Flow rate Any impacts from the surrounding area (land use, factories, etc.) (*see MIxS for more detailed metadata recommendations for specific environments)
Designing a sampling plan that includes target environments/sample locations that are likely to contain the microbe(s) of interest	Do multiple sub-environments need to be sampled to adequately answer my research question (e.g., water in addition to sediment in an aquatic environment)? Do multiple locations need to be sampled to capture variability or some impact on the environment (capturing background or undisturbed environments, capturing changes over space or time)? Is the environment low in genetic material (DNA or RNA)? Am I sampling enough volume to obtain sufficient genetic material needed for sequencing and gain an accurate depiction of the sampled environment? Do I have appropriate negative controls to detect contamination that might skew results?
Field sampling considerations	Supplies needed to collect samples Appropriate biohazard and safety protocols Supplies must be decontaminated (materials to use at each site or what is needed to decontaminate those materials in the field)

	Sampling materials may differ depending on what is being tested (material should not interfere with analyses) Controls in place to prevent dynamic shifts in the microbial samples during transportation prior to preservation. (e.g., keeping samples on ice) Does the weather in the past 24+ hours impact conditions in the system of interest? Should sampling occur during dry or wet weather? Are there variables that impact flow (e.g., it might be most appropriate to target dry season, low-flow period for sampling)? What is the time requirement for sampling? Can samples be processed the same day as collection to minimize changes to samples during transport/storage? Should samples be processed in the field? How many technical and/or biological replicates need to be collected? How much volume or mass should be collected? What field, trip, and equipment blanks are needed? What are some common sources of contamination and what is being done to minimize these?

4.1.1 Sampling Considerations

One of the primary concerns associated with sample collection is whether the collected sample is truly representative of the environment of interest. Therefore, following established protocols helps to ensure that a sample is representative, that appropriate preprocessing steps are followed to maintain the integrity of the sample, and that downstream inhibitors are minimized. Despite the importance of sampling, current published literature does not often discuss sample collection procedures in detail (Mthethwa et al., 2021). It is also important to remember that methods validated for one environment may not be optimal for another. General information to consider when developing a sampling plan is summarized below.

4.1.2 Grab vs. Composite Samples

Environmental samples are typically collected either as grab or composite samples. Grab samples represent a snapshot in time as a sample is collected instantaneously at a given location. Grab samples are often easier to collect and are ideal when the environment is homogeneous over space and time. Grab samples can also be implemented in series to capture system dynamics over time and space, but this requires substantial labor and may be more appropriate for shorter term, more focused studies. Alternatively, composite samples provide a more representative average of the microbial community in a heterogeneous environment when there is variability over space and/or time. This can be advantageous if the target of interest is expected to vary diurnally e.g., pathogen targets in sewage. Composite samples may be collected by combining multiple grab samples or by using specifically designed automatic sampling devices. Composite samples may be collected over time through continuous sampling or by combining equal volumes collected at regular time intervals. Flow-proportional or flow-weighted composite samples, on the other hand, may be collected by mixing equal volumes collected at time intervals inversely proportional to the volume of flow or by mixing flow proportional volumes at regular time intervals (Baird and Laura Bridgewater, 2017). Though dependent upon project design, advantages of composite samples include not having to analyze multiple samples and having a sample that is more representative of the environment. Disadvantages of composite sampling include losing the connection between the microbial biomass and its relationship to an individual point in time, potential dilution of genetic material, and the potential for analytical interferences and probability of analyte interactions which can be especially important when monitoring for outbreaks or failures in the system. Because genetic material can degrade over time when not stored properly, it is important that the storage temperature be considered when using automatic composite samplers. When samples are collected over long intervals, there can be dynamic shifts in the collected microbial communities during storage, thus limiting the sampling and storage interval may be appropriate for the most accurate results.

4.1.3 Replication

Replicates help assess sources of variation that may skew results. Biological replicates refer to independent samples collected in separate containers or experimental replicates and capture variation in the system of interest. Technical replicates, on the other hand, are repeated measurements of the same sample and capture random noise associated with a protocol or analysis. Technical replicates may also be collected to assess data repeatability based on field conditions and to help distinguish technical from true variability. Despite the call for more replicates in studies (Prosser, 2010), they have not commonly been collected or analyzed in NGS studies (Quince et al., 2017b), likely due to the high cost of sample analysis. However, replication is key for robust experimental design.

4.1.4 Controls

It is especially important to incorporate negative controls into sampling when the sampled environment is characterized by low biomass to be able to measure contamination. Negative controls recommended for qPCR and dPCR should be considered for NGs, including controls for sampling and sample treatment (e.g., additional sample processing such as elution after concentration) as well as during extraction, reverse-transcription (if required), and sequencing (Borchardt et al., 2021). Negative controls related to sampling can include equipment blanks (e.g., laboratory grade reagent water is transported to the site and passed through the sample collection device or any other equipment used to sample), field blanks (e.g. laboratory grade reagent water is poured into a sampling container in the field that replicates the conditions and time frame of environmental samples), and trip blanks (laboratory grade reagent water in a sampling container is transported to and from the field unopened) (EPA, 2017a). Sampling blanks as well as negative controls included throughout sample processing may be sequenced along with biological samples (Hornung et al., 2019; Salter et al., 2014). To address this, positive control DNA can be spiked into the negative control DNA extractions, to ensure enough DNA to avoid failure of the sequencing run.

4.1.5 Sampling Environments or Matrices

This section will summarize sampling methods for the following environments:

- Wastewater
- Drinking water
- Biofilms
- Sludge/Biosolids
- Aquatic Sediment
- Surface water

4.1.5.1 Wastewater Treatment Plants

Samples collected at WWTPs may be grab or composite, though influent at WWTPs can be especially variable due to human behavior and/or weather and therefore time of sampling must be carefully considered. Hydraulic residence time should also be considered when sampling at multiple points throughout the WWTP. If automatic composite samplers are used, the storage temperature of the sample should remain cool to prevent degradation of nucleic acids. Operating procedures recommended by the EPA should be followed for sampling procedures, and in general, it is recommended that wastewater samples are collected at well-mixed locations (EPA, 2017b). The volume that one collects depends on where one is sampling in the treatment train. Wastewater collected from influent or activated sludge will require smaller minimum volumes as there is more genetic material in these samples. Samples collected later in the treatment train, especially after disinfection will require much larger sampling volumes. If chlorine is used during disinfection, sodium thiosulfate should be added (see drinking water section for more details).

4.1.5.2 Drinking Water

Recommendations for sampling drinking water will typically follow established methods for pathogen detection or molecular methods with the main caveat being the need to collect and process a larger volume to obtain enough nucleic acid for NGS. Studies have shown ranges of approximately 10 L to 2000 L may need to be concentrated for drinking water samples (Brumfield et al., 2020; Chao et al., 2013; Ma et al., 2019). For collecting tap water samples, the EPA recommends flushing the distribution line for a few minutes before sampling. It is also recommended to add sodium thiosulfate to any water that contains any remaining chlorine disinfectant residual to neutralize chlorine and minimize associated inactivation of microorganisms after sample collection (EPA, 2016).

4.1.5.3 Biofilm

It may be of interest to sample biofilm within environments of interest including distribution systems, drinking and WWTPs, in sediment, or marine environments. Biofilm has been collected by scraping the surface of valves or water meters, swabbing the interior pipe surfaces, by sonicating pieces of cut-out-pipe, or by inserting and later retrieving coupons placed into pipes. Separating the biofilm from the cut pieces of pipe or swabs can be done through sonication (Liu et al., 2020; Zhang and Liu, 2019) or vortexing (Ling et al., 2015). Swabbing the pipe interior is more common in the literature; it should be taken into consideration that studies have shown the microbial community varies depending on where sample is collected within the pipe (Liu et al., 2020). Biofilm from water treatment systems may be collected from sand or granular activated carbon (GAC); the filter material is often sampled where it is accessible, for example from the upper surface (Lautenschlager et al., 2014) or through core samplers (Palomo et al., 2016). Biofilm collected from streams is typically brushed or scraped off cobble sized rocks until a given surface area has been covered (Bekliz et al., 2019; Roberto et al., 2019). Surface area covered during sampling varied across methods and was not always reported. Processing of biofilms varied from mixing followed by filtering through 0.2-μm filters (Roberto et al., 2019) or initial sonication of rocks in water prior to scraping followed by filtration through 0.45-μm membranes (Pu et al., 2019).

4.1.5.4 Biosolids

Biosolids are the product of a treatment process (e.g., dewatering, palletization, anaerobic digestion) and may be sampled at different stages depending on one's research question(s). The input or output of these processes may be sampled to examine the effect of treatment. Mass required for NGS samples may vary depending on treatment process; a range has been reported in the literature from 100 g to 500 g (Bedoya et al., 2019; K. Bibby et al., 2011). However, as solids cannot be further concentrated, the mass used for extraction may depend on extraction method or kit. Total solids should also be analyzed so that results can be reported per gram of dry weight (EPA, 2014b).

4.1.5.5 Surface Water

Sampling surface water from streams, lakes, or oceans will each have different considerations. Some important considerations are to make sure that enough volume is collected to obtain adequate genetic material, that enough samples are collected to capture any spatial variability of interest across the surface water area and depth, and that sampling procedures do not introduce contamination when sampling. For example, the opening of the sampling device should face upstream and the water should be sampled first if sediment is also being collected at the same location, which will help to avoid resuspending sediment into the sampled water column (EPA, 2017a). The EPA's standard operating procedure (SOP) should be referenced for guidance in sampling different methods including the use of peristaltic pumps, discrete depth samplers, submersible pumps and more (Simmons, 2016). Sampling volumes have varied across studies ranging from 250 mL to 5 L (Chopyk et al., 2020; Mizusawa et al., 2021; Roy et al., 2018).

4.1.5.6 Sediment

The ways in which aquatic sediment is sampled will vary depending on where samples are collected. For example, lake sediment will have different sampling requirements than sediment from rivers or oceans. Lake sediment may be sampled using a gravity corer at the maximum depth when it is assumed that the genetic information at maximum depth is representative of the entire lake basin, this is especially common when comparing multiple lakes as this gives an idea of the microbiome for each lake (Garner et al., 2020). However, when a lake has a complex topography that leads to spatial variability, or it is expected that the sediment is not heterogeneous, it can be important to sample at multiple locations in the lake. Capo et al. (2021) provide a thorough summary of considerations for lake sediment sampling for DNA and identify the importance of replicates when target organisms are rare; increased replication increases the probability of detection while additional analytical replicates ensure reliability of the data. Sediment should be sampled using cores and may be subsampled using sterile implements immediately after core collection or splitting of the core. Sediment should be stored in cold, dark, anoxic conditions if not proceeding directly to DNA extraction.

There are a wide variety of sampling methods described for collecting sediment from rivers and streams; various parameters include the depth at which sediment was collected, volume of sediment collected, and how those samples were processed. Sampling depth varied from the top few centimeters of sediment to 15 cm depth or even sediment cores (Bier et al., 2020; Jacquiod et al., 2018; Suttner et al., 2020). Volume collected also varied from 10 mL to 250 mL (Mizusawa et al., 2021; Suttner et al., 2020). Some samples were sieved to remove coarse material and only the sandy fraction was retained (Fodelianakis et al., 2021). It is common for water to be collected at the same time and location as sediment.

4.1.6 Additional Considerations

4.1.6.1 RNA

Sampling methods for DNA or RNA are often quite similar with additional considerations taken when sampling for RNA because it is more likely to undergo physicochemical degradation (Veilleux et al., 2021) as well as degradation by RNases, which are ubiquitous in the environment (Zhang and Liu, 2019). An RNA stabilizer or preservative may be added to samples to prevent degradation (Bizic et al., 2022; Carvalhais and Schenk, 2013). Alternative methods must also be considered for sample storage, processing, extraction, and analysis for RNA samples. These are discussed in later sections (sample concentration – viral RNA targets, sample preservation – freezing samples and nucleic acid buffers, nucleic acid extraction – RNA, and sequencing methodology/platform – metatranscriptomic sequencing).

4.1.6.2 Metadata

Collecting relevant metadata is also important, especially if it can only be collected at the same time as sampling (for example, pH and temperature of a water sample). There are specific guidelines on minimum information that should be collected and recorded depending on the sample type and how it is analyzed. The Genomic Standards Consortium developed Minimum Information about any (x) Sequence (MIxS), which includes checklists for reporting technology-specific information related to the sequences themselves as well as sample data using environmental packages. MIxS has many packages depending on the environment being sampled including air, human-gut, sediment, soil, wastewater/sludge, water, and more (Yilmaz et al., 2011). The environmental packages can help one think through what information might need to be collected during sampling. For example, the checklist for wastewater/sludge includes the wastewater type (i.e., origin), suspended solids, secondary treatment, solids retention time, nutrients, etc., that can be important to collect the same day as sampling. The MIxS packages should be reviewed before sampling to make sure all additional information or samples are collected. Additional guidelines that should be considered include Minimum Information for Publication of Quantitative Real-Time PCR Experiments (the MIQE guidelines) published for qPCR (Bustin et al., 2009) and dPCR (Huggett et al., 2013; The dMIQE Group and Huggett, 2020) and the Environmental Microbiology Minimum Information guidelines

(Borchardt et al., 2021). These guidelines give information specifically related to qPCR and dPCR, however, similar considerations may be taken for NGS. For example, guidelines are given for collecting a sampling negative control, which evaluates equipment contamination, by sampling a sterile matrix (e.g., autoclaved water) that has passed through any sampling equipment used in the field. Which parameters need to be collected depends on one's specific hypotheses, more extensive metadata may need to be collected for additional water quality parameters, conditions associated with treatment operations, chemical contaminants, and influent sources or community data (e.g., socioeconomic, demographic).

4.1.6.3 Chain of Custody

Proof of chain of custody may be required for certain projects, for example, samplers for EPA's Contract Laboratory Program are required to maintain Traffic Report/Chain of Custody records (U.S. Environmental Protection Agency, 2020). The chain of custody documentation should include but is not limited to field logbooks, sampling trip reports, sample shipping and receipt logs, and analytical data sheets (Vasileski, 2000). Implementation of a chain of custody is not typically reported in literature, however, it can be important to minimize errors, especially when multiple people are involved in sampling and/or sample processing.

4.1.6.4 Safety Protocols

Water samples can contain a number of pathogens (especially untreated wastewater) potentially putting those handling the samples at risk, therefore, proper safety precautions should be taken to minimize exposure. Personal protective equipment (PPE) recommendations vary depending on where one is sampling; PPE typically recommended for wastewater utility personnel include durable gloves, safety glasses, Tyvek suits or coveralls, and respiratory protection (LeChevallier et al., 2020). Additional guidance may be provided by federal or regional agencies, for example, the CDC guidelines for handling human waste or sewage include proper PPE as well as training and vaccination recommendations (U.S. Centers for Disease Control and Prevention, 2022).

4.1.6.5 Sample Volume and Material

Considering volume and material or equipment needed to sample is also important before setting out to sample. It is crucial that enough microbial biomass is collected for sequencing (Quince et al., 2017b). If the environment being sampled has an unknown quantity of genetic material, it may be prudent to conduct preliminary analysis to quantify the genetic material (e.g., nucleic acid quantity) recovered using the planned sampling methods through devices such as Nanodrop or Qubit (Lim et al., 2016). It is important to use sterile sampling materials that will not introduce contamination or interfere with any downstream processes. Though not always reported, sampling materials are typically sterile plastic or glass (Brumfield et al., 2020; EPA, 2016, 2017b; Pu et al., 2019).

Some of the SOPs described here are not specifically for NGS, therefore prior to sampling the literature should be reviewed to determine if there are any updated standard methods or best practices. It is recommended that future publications include a more detailed description of sampling methods including information about sampling volume, sampling depth (where relevant), temperature of storage transportation until processing, and any relevant metadata.

Additional recommendations for study quality indicators related to AMR in wastewater and related aquatic environments was described by rEporting antiMicroBial ResistAnCE in WATERS (i.e., EMBRACE-WATERS) (Hassoun-Kheir et al., 2021). There is also a lack of SOPs focusing on a comparison of sampling methods, identifying potential for future research studies that might be able to help establish standard methods.

4.2 Sample Concentration

Aqueous samples typically require concentration prior to DNA/RNA extraction, although sludge, biofilm, sediment and other sample types often contain sufficient nucleic acids for direct extraction. Sample concentration should take place as soon as possible prior to DNA/RNA extraction. It is important to minimize further microbial growth and alteration of the microbial community profile prior to sample concentration, and therefore samples should be transported at < 10˚C (but not frozen) from the field to the lab, ideally processing them through to the concentration step in less than 6 hours (US Environmental Protection Agency, 2012). Freezing samples should be avoided because it can damage or break the cell envelope and interfere with subsequent sample concentration measures. Samples should be additionally preserved as needed following the recommendations detailed in the sample preservation section.

Most protocols for NGS sequencing require a minimum concentration of DNA/RNA to ensure quality sequencing. Sample concentration is critical to maximize the DNA/RNA available for sequencing with available approaches outlined in Table 4-2. However, there are multiple sample concentration methods available, and the optimal choice may depend on a number of factors, including the sequencing target (DNA vs. RNA). Some DNA/RNA extraction kits combine sample concentration with extraction. These will be addressed in the subsequent section devoted to extraction. For all approaches, process controls or sample treatment negative controls (i.e., processing sterile water in parallel with the sample processing) are recommended in order to capture contamination introduced by laboratory processes.

4.2.1 All Targets

If the aim is to capture nucleic acid from all possible microbial targets; including bacteria, eukaryotes, and virus DNA/RNA, then viruses typically become the limiting factor in selecting an appropriate concentration method. The method selected must take special care to avoid loss of smaller viral particles. Centrifugation can typically effectively concentrate all microbe types when there are already high solids concentrations (e.g., activated sludge, settled solids) (Bofill-Mas et al., 2006; Graham et al., 2020; Yu and Zhang, 2012). Hollow fiber filtration or ultrafiltration is relevant for lower turbidity water samples (e.g., reclaimed water, drinking water) or clarified samples (samples that have been subject to settling or passed through a 5 micron filter) and has been recommended for viral and bacterial surveillance(Ahmed et al., 2020; Chao et al., 2013). However, the efficiency of elution as well as choice of elution buffer and eluation pressure might affect the optimal recovery of organisms trapped in the filter. The Innovaprep concentrating pipette (using ultrafilter or 0.05 micron tips) is also a promising method that can effectively capture viruses and other microbes for water testing purposes (Gonzalez et al., 2020; Klempay et al., 2021; Pecson et al., 2021). This method, while versatile, is more expensive per sample than others. However, Innovaprep may be easier to implement in laboratory and field applications.

4.2.2 Prokaryotes

The most common sample concentration method for aquatic prokaryotic DNA is membrane filtration (U.S. Environmental Protection Agency, 2005). Water samples are vacuum filtered onto a small pore size membrane (0.22 - 0.45 microns; polycarbonate, mixed cellulose, or nitrocellulose) (APHA, 2017; Li et al., 2018). The volume filtered may depend on research questions, either a specific sample volume is filtered or the samples are "filtered until clogging." Filtering until clogging may be required for highly turbid samples, such as wastewater influent and activated sludge. This method is relatively inexpensive and requires minimal laboratory training beyond sterile technique. However, some high turbidity sample types may clog very quickly. It is also critical that this volume is recorded and reported accordingly to provide a denominator for subsequent volumetric normalization. Filters can then be immediately subject to DNA extraction, or stored in 50% ethanol at -20˚C until DNA extraction (Li et al., 2018). Low DNA samples such as drinking water may require large volumes of water (> 2L) to obtain sufficient DNA for sequencing (Keenum et al., 2021).

4.2.3 Viruses

Sample concentration for RNA viruses is rapidly evolving for wastewater detection with the SARS-COV-2 pandemic and subsequently evolving for metagenomics characterization. Pecson et al. (2021) recently showed that of 36 methods to process (and therefore concentrate) wastewater for SARS-COV-2 from wastewater, sample processing added minimal variation in enumerated copies (Pecson et al., 2021). Unique methods exist for viral extraction; such as magnetic particles (Julian and Schwab, 2012; Patnaik et al., 2020), skimmed milk flocculation (Gonzales-Gustavson et al., 2017), or positively charged cartridge filtration (Chaudhry et al., 2015; Williams et al., 2001). However little work could be found that applied these applications to viromics. The most commonly applied and effective method for viral sample concentration in aquatic matrices involved the precipitation of viruses using polyethylene glycol (Adriaenssens et al., 2018; Hjelmsø et al., 2017). This method lead to the highest richness of identified viromes via next generation sequencing, however it only targets viruses in a sample (Hjelmsø et al., 2017).

Table 4-2. Summary of Approaches Available for Sample Concentration.

Approach	Target	Strengths of Approach	Weaknesses of Approach
Concentrating Pipette	All targets	-Applicable to all types of organisms - Easy to implement	-Cost -New approach that is not well vetted
Centrifugation	All targets	-Applicable to all types of organisms -Inexpensive	-Only useful for samples with high solids concentrations
Hollow Fiber Filtration	All targets	-Applicable to all types of organisms -Best for low DNA/RNA concentration samples	-Variable elution efficiencies -Cannot be applied for high turbidity samples

Membrane Filtration	Organisms > 0.22 microns or that can be flocculated to > 0.22 microns	-Inexpensive -Easy to implement	-Time consuming to capture sufficient volume from high turbidity samples -May need very large quantities of water for low microbe samples -A significant fraction of viruses may be lost if not flocculated
Polyethylene glycol	Viruses	-Results in high concentration of viruses	-Method is specific only to viruses
Skim milk flocculation	Viruses	-High recovery rates	

4.3 Sample Preservation

Selection of sample preservation method is critical for maintaining the integrity of nucleic acids for downstream NGS of water and wastewater samples, particularly when the sample cannot be processed immediately after collection. Available methods for sample preservation are outlined in Table 4-3. Some of the most common methods of sample preservation include refrigeration at 4 °C or freezing at -20 °C or -80 °C, as temperature is a driver of microbial decay. Addition of chemical preservatives or buffers is another common preservation method that can deactivate or inhibit enzymes that degrade nucleic acid and prevent bacterial growth. Often, these methods are combined for water samples (e.g., freezer storage with the addition of a chemical preservative). Ultimately, the selection of a particular method is dependent on the availability of resources, anticipated timeframe of downstream analyses (filtration, nucleic acid extraction and sequencing), and targeted nucleic acid (DNA or RNA). The following sample preservation methods of interest are detailed in this section:

- Refrigeration at 4 °C
- Freezing at -20 °C or -80 °C
- Chemical preservation (ethanol fixation, cetyltrimethylammonium bromide (CTAB), silica beads and formaldehyde)
- Nucleic acid buffers (DNA/RNA Shield or RNA*later*)

4.3.1 Refrigeration at 4 °C

As noted above, sample concentration should ideally be completed within 4 hours of collecting the sample to maximize recovery of nucleic acids (Hinlo et al., 2017). However, sometimes this is difficult to achieve. Studies evaluating the impact of short-term storage (~7 days or less) at 4 °C of wastewater prior to filtering have focused primarily on RNA degradation, as it is less stable than DNA. SARS-CoV-2 RNA stability in wastewater remained constant for up to 9 days when stored at 4 °C (Markt et al., 2021) and up to 8 days at 4 °C for settled wastewater solids and decreased less than one order of magnitude beyond 35 days (Simpson et al., 2021). For treated sewage samples to be concentrated for viral quantification via PCR, immediate concentration (filtration, acid rinse and elution) followed by 13 days of storage at 4 °C showed higher poliovirus and norovirus recoveries when compared to 13 days of storage at 4 °C followed by concentration (Haramoto et al., 2008). For water samples, it has been recommended to keep samples cool prior to DNA extraction, and extraction should be done as soon as possible after collection (Mthethwa et al., 2021). Refrigeration is recommended if water samples are to be processed within 24 hours of collection for environmental DNA analyses (Hinlo et al., 2017).

4.3.2 Freezing at -20 °C or -80 °C

Samples are generally frozen at either -20 °C or -80 °C for long-term storage (>7 days). This method is used either before filtration (on bulk water samples) or after filtration (on filters). Caution is warranted in freezing samples before sample concentration via filtration because damage to cell envelopes incurred by freeze-thaw cycles will affect capture by the filter. SARS-CoV-2 RNA quantification in wastewater following freezing of bulk water has demonstrated mixed results. There was a significant signal loss when influent was frozen at -18 °C for 2-3 days, which could be a result of the release of RNase from bacterial cells in the sample (Markt et al., 2021). However, SARS-CoV-2 RNA influent concentrations were stable when stored at 4 °C for 28 days followed by freezing for 58 days at both -20 °C and -80 °C (Hokajärvi et al., 2021). SARS-CoV-2 RNA in solids was found to both increase and decrease in replicate samples when stored at -20 °C for 2-3 days. There was a decrease when solids samples were stored at -80 °C for 122 days, but still within an order of magnitude of the fresh samples (Simpson et al., 2021). It is unclear which temperature is preferable when choosing freezing for sample preservation, although freeze-thaw cycling affects DNA integrity (Hinlo et al., 2017).

4.3.3 Chemical Preservation with Ethanol

Ethanol fixation is one of the most common preservation methods for water samples including freshwater, sludge and wastewater filters. Ethanol as a preservative deactivates DNases and prevents bacterial growth (Kumar et al., 2020). When used immediately after sample collection in a 2:1 ratio of 100% ethanol to water, ethanol fixation resulted in similar DNA concentrations as freezing at -20 °C (Kumar et al., 2020), suggesting that ethanol fixation is a feasible substitute to freezing. In wastewater sludge samples, a 1:1 ratio of 100% ethanol to sample, followed by 12-hour storage at -20 °C, DNA purity was not significantly different than fresh samples. However, slight changes in community structure were observed, with a particular decrease in abundance of *Chloroflexi*, although this bias could be due to DNA extraction methods or other downstream processes. Fixation was recommended over freezing as a long-term storage method, but it is important to note that the choice of DNA extraction kit had a stronger effect on sample analysis outcomes than preservation method (Guo and Zhang, 2012).

The impact of ethanol fixation was compared to storage at -20 °C on antibiotic resistance gene (ARG) recovery from influent and effluent filters and activated sludge samples, with a particular focus on sample preservation for long-distance shipping. There was no significant difference in ARG diversity between fresh samples, frozen samples (-20 °C for three weeks) and 50% ethanol-fixed samples. For all samples analyzed, there was no difference in ARG and taxonomic profiles after long-distance shipping (Hong Kong to Virginia, USA) at ambient temperature fixed in 50% ethanol. This suggests that 50% ethanol fixation at ambient temperatures up to three weeks is suitable for metagenomic analysis of ARGs in wastewater influent, effluent and activated sludge (Li et al., 2018).

4.3.4 Chemical preservation with Longmire's Solution

Longmire's solution (100 mM Tris, 100 mM ethylenediaminetetraacetic acid (EDTA), 10 mM NaCl, 0.5% sodium dodecyl sulfate (SDS) and 0.2% sodium azide) has been used to preserve environmental DNA after filtration. EDTA and SDS inhibit enzyme activity and sodium azide prevents bacterial growth. This solution has been used to preserve environmental DNA in water samples (1:3 ratio of solution to water) and was just as effective as freezing at -80 °C for up to 28 days, but DNA concentration measured by qPCR declined by 56 days (Williams et al., 2016).

4.3.5 Preservation with Silica Beads

Preservation of filters dry on silica gel is a less common preservation method but has been shown to give the same number of operational taxonomic units in sampled river water, when compared to storage for one week at -20 °C and filter preservation in a mixture of 99% ethanol and Qiagen ATL buffer (Majaneva et al., 2018). Desiccation using silica gel for other water matrices for downstream analyses requires further research.

4.3.6 Chemical Preservation with Other Solutions

Other chemical solutions that have been used for sample preservation include formaldehyde and CTAB. Preservation in 3% paraformaldehyde and 50% ethanol fixation was successfully used on wastewater samples for 16S and 23S rRNA oligonucleotide probing (Manz et al., 1994). CTAB has also been suggested as a preservative for water filters (Renshaw et al., 2015), but there is limited research on CTAB as a preservation buffer compared to other methods.

4.3.7 Preservation with Nucleic Acid Buffers

Proprietary solutions designed to maintain integrity of nucleic acids, particularly RNA, can also be used for preservation of water, wastewater or sludge samples. RNA*later* (ThermoFisher Scientific) can be used to protect RNA at -20 °C for long-term storage or at ambient temperatures for short-term storage. RNA*later* added to activated sludge stabilized 16S rRNA at -20 °C for up to three months (with initial storage at 4 °C for 72 h); however, recovery of 16S rRNA was lower after six months as measured by reverse transcriptase PCR (Cydzik-Kwiatkowska and Wnuk, 2011). DNA/RNA Shield (Zymo Research) has been used to preserve samples for later quantification of RNA via real-time reverse transcriptase PCR in human stool samples (Coryell et al., 2021) and for metagenomic sequencing and 16S rRNA qPCR on potable water (Stamps et al., 2018). There is a need for further research to compare these nucleic acid buffers to other preservation methods for water or wastewater samples.

4.3.8 Limitations and areas of further research

Most studies use a combination of preservation approaches (e.g., freezer storage with a chemical preservative), therefore, elucidating the impact of a single preservation method is not straightforward. Currently, there is no apparent optimal preservation strategy if performing NGS on both DNA and RNA in the same sample. Most studies to date have focused on the impact of sample storage as it applies to quantitative PCR and not NGS. In general, more research is needed on RNA stability (other than SARS-CoV-2 RNA) in water samples, especially for the purpose of sequencing. It has also been suggested that the choice of sample preservation method and nucleic acid extraction kit maybe be closely intertwined, and these two approaches should be optimized simultaneously (Hinlo et al., 2017).

Table 4-3. Summary of Approaches Available for Sample Preservation.

Approach	Strengths of Approach	Weaknesses of Approach
Refrigeration at 4 °C	No freeze-thaw damage to nucleic acid Typically easiest method	Appropriate only for short-term storage (~1 week or less), particularly for RNA Changes in microbial community composition will be incurred and compound with time
Freezing at -20 °C or -80 °C	May allow for longer storage of samples for later analysis (e.g., SARS-CoV-2 RNA in solids)[1] Highly effective for long-term DNA/RNA storage	Freeze-thaw will damage cell envelopes and undermine their capture by membrane filtration More research needed to compare effects of -20 °C vs -80 °C for downstream analyses
Preservation (fixation) in ethanol	Less expensive than other chemical preservatives Allows for ambient temperature storage up to 3 weeks[3]	Potential for changes in bacterial community structure[4]
Preservation in other chemicals (e.g., Longmire's buffer or CTAB)	Allows for ambient temperature storage[5,6]	Can be more expensive than ethanol fixation Requires storage of hazardous chemicals (SDS, sodium azide) Limited research on CTAB
Preservation using silica gel beads	May result in more similar community compositions compared to other chemical buffers[7]	Limited research Recommended for filter preservation (not bulk water) May not be recommended for long-term storage
Preservation in nucleic acid buffers (e.g., DNA/RNA shield or RNA*later*)	Allows for ambient temperature storage[8]	Proprietary solutions More expensive than ethanol fixation

[1](Simpson et al., 2021) [2](Markt et al., 2021) [3](A.-D. Li et al., 2018) [4](Guo and Zhang, 2012) [5](Williams et al., 2016) [6](Renshaw et al., 2015) [7](Majaneva et al., 2018) [8](Mutter et al., 2004)

4.4 Nucleic Acid Extraction

Nucleic acid extraction is the foundation of all molecular-based microbial community analysis. The ideal nucleic acid extraction and purification method should capture all of the DNA and/or RNA in a sample such that:

- No bias is introduced (i.e., nucleic acids are captured from all groups of organisms with comparable efficiency);
- The yield is reflective of the true concentration of nucleic acids in that sample;
- The recovery captures the true diversity of species from which the nucleic acids in that sample originated;
- The final extract is not contaminated by other sources and purity is acceptable (the extract has a reasonable 260/280 ratio and exists as undegraded genomic DNA); and

- The extract is free of inhibitors that may affect downstream processes, such as PCR, library preparation, or sequencing.

While some traditional methods such as phenol-chloroform phase separation are still used today, the development of commercial extraction kits has offered many benefits; including reduced processing time, reduced exposure to hazardous reagents, and opportunities for standardization. Most of these kits use a combination of chemical, enzymatic, and mechanical treatment to lyse cells and viruses, release genomic DNA and/or RNA, and inhibit nucleases. Free nucleic acids are then bound to some matrix (a spin column or a suspension of beads/particles) and washed repeatedly to remove impurities. A comparison of lysis and purification strategies involved in commonly used microbial DNA/RNA extraction methods is provided in Table 4-4.

Efficiencies and accuracies of commercial extraction kits vary widely. A gut microbiome-based extraction experiment comparing nine DNA extraction methods found that yields for the same homogenized stool sample could vary from less than 0.5 µg to over 8 µg of total nucleic acid per extraction depending on the method used (Lim et al., 2018). However, yield is not usually correlated with diversity or richness in these comparison studies, underscoring the importance of finding a kit that not only exhibits high recovery, but also high extract quality and minimal bias (Knudsen et al., 2016; Yuan et al., 2012).

The two main factors influencing the efficiency of nucleic acid extraction are the extraction method used and the type of sample being processed. Extraction and purification interference can occur in non-bulk water samples (biofilms, activated carbon, activated sludge, etc.) due to the sample matrix. These samples can protect cells from lysis, contain high concentrations of inhibitors, or bind and retain extracellular DNA (Guo and Zhang, 2013; Kirtane et al., 2020; Lemarchand et al., 2005). The composition of wastewater and drinking water samples can vary widely in terms of inhibitors and disinfectants, both between and within treatment plants, presenting unique challenges (Lemarchand et al., 2005). Choice of extraction method is dependent on sample type and may differ between various sample types or within a set of complex samples (e.g. activated sludge) from different sources (Vanysacker et al., 2010; Walden et al., 2017).

The profile of the microbial community composition resulting from NGS analysis can be affected by choice of extraction method. Lysis-resistant phyla, such as *Actinobacteria* or *Mycobacteria,* can be severely underrepresented in metagenomes when compared against extraction-independent methods, such as fluorescence in situ hybridization (FISH) (Albertsen et al., 2012; Haig et al., 2018). Sample source, sample type and microbial community composition can introduce bias during initial extraction lysis, while the second purification step likely contributes more to overall yield and purity (Guo and Zhang, 2013).

Validation studies on different types of wastewater and drinking water samples have provided some guidance regarding the impact of different methods and kits (Table 4-5). While not comprehensive, these studies serve as a guide in selecting extraction methods based on the composition of the sample to be analyzed.

4.4.1 DNA Extraction

4.4.1.1 Wastewater and Biofilm Samples

Validation studies listed in Table have overwhelmingly recommended the Fast DNA Spin Kit for Soil (MP Biomedicals) for activated sludge, wastewater influent, and biofilm samples due to the kit's high yield, high extract purity, and higher representation of Gram-positive genera following community analysis (Albertsen et al., 2015; Guo and Zhang, 2013; Hwang et al., 2012; A.-D. Li et al., 2018; Niestepski et al., 2019). This kit has two unique features that likely contribute to its efficacy: glass beads used in lysis have a polydisperse distribution and the binding matrix used in purification is a suspended solution of silica particles with higher surface area that supports higher yields compared to a typical spin column (Guo and Zhang, 2013; A.-D. Li et al., 2018). However, this kit often produces lower measures of community richness than other kits or phenol-chloroform methods, though the most abundant OTUs are well-captured (Brandt and Albertsen, 2018; Guo and Zhang, 2013; Hwang et al., 2012).

Extraction kits using mechanical treatment (i.e., bead-beating) may result in DNA shearing into <10 kb fragments, which may preclude the use of those extracts in long-read sequencing or whole-gene 16S rRNA amplification (Albertsen et al., 2015; Guo and Zhang, 2013; Moss et al., 2020). Bead-beating may be appropriate for extracts intended for long-read sequencing if there is sufficient input material to yield enough DNA for library preparation, as in the case of sludge samples subjected to a modified Fast DNA Spin Kit method (Brandt et al., 2020). When input biomass is low, a protocol using an enzymatic cocktail for lysis and phenol-chloroform separation for purification may be a preferable alternative. Maghini et al. (2021) suggests a method using this approach for extraction of high molecular weight DNA from stool samples, though this protocol has not been tested on wastewater samples.

4.4.1.2 Drinking Water Samples

Low biomass levels and the presence of residual disinfectant can complicate extraction of drinking water samples. DNA yields may not be high enough to provide material for sequencing, even when filtering high volumes (Putri et al., 2021). Due to the nature of these samples, validation studies on tap water recommend the PowerWater DNA Isolation kit or a phenol-chloroform extraction method, with quenching of residual disinfectant prior to extraction if necessary (i.e. chlorine concentration is greater than 0.2 mg/L) (Brandt and Albertsen, 2018; Haig et al., 2018; Putri et al., 2021). A method developed by Vosloo et al. (2019) provides a modified protocol for the PowerWater kit with yields two to three times higher than the manufacturer protocol. A recent Water Research Foundation project, Project 4721, recently compared the DNeasy PowerWater DNA Isolation kit (Qiagen), the Fast DNA Spin Kit, the Fast DNA Spin Kit for Soil (MP Biomedicals), and a traditional phenol-chloroform method for extracting DNA from premise plumbing water samples (Raskin et al., 2022). They found that the best kit can vary depending on which OP was the primary target of interest (e.g., mycobacteria have a waxy outer layer and phenol-chloroform was optimal), but that overall, the DNeasy PowerWater kit was optimal when seeking to capture a variety of pathogens. The Fast DNA Spin Kit performed nearly equivalently when compared with the DNeasy kit.

4.4.1.3 Complex Media

Extraction methods for samples with extensive surface area for attachment, such as media from biologically activated filters or granular activated carbon, have not been extensively studied. Granular activated carbon is particularly challenging due to its apparent ability to bind free DNA (Kirtane et al., 2020). The high sorptive capacities and surface areas of these types of media may necessitate sample pretreatment prior to extraction through sonication and/or introduction of a competitively binding substance, or at least a modified mechanical lysis step (Zhang et al., 2010).

For previously uncharacterized sample types, the selection of an appropriate extraction method (and pretreatment, if necessary) may require preliminary testing. However, because it is difficult and usually unfeasible to measure the accuracy of several methods alongside large-scale community analyses studies, standard practices recommended here should serve as guidance for cases where preliminary testing cannot be carried out or can only be afforded if two or three methods are compared. Positive controls (e.g., external spikes) should also be considered to assess the reliability of the chosen extraction method and to guide interpretation of the resulting data according to potential biases.

4.4.1.4 Extracellular DNA (exDNA)

Depending on the research question, extracellular DNA (exDNA, i.e., free or adsorbed DNA originating from dead cells or live cell secretions) in water samples may either serve as a contaminant from inactive cells or a source of information. In the first case, wastewater samples intended for metabolic analysis can be pretreated with reagents, such as DNase or propidium monoazide, to remove extracellular nucleic acids, though differences in community structure between untreated and treated (exDNA-free) wastewater samples may not be large enough to warrant pretreatment of large sample sets (Albertsen et al., 2015). On the other hand, the capture of exDNA from sludge or influent filtrate can be achieved using CTAB precipitation or magnetic bead-based methods, which can provide information about the presence of extracellular antibiotic-resistance genes (Yuan et al., 2019; Zhang et al., 2013). Analysis of drinking water samples is much more likely to be affected by exDNA due to low concentrations of intact cells. In fact, exDNA may account for as much as half of the total DNA in drinking water samples depending on the extraction and treatment methods used, whereas it may make up less than 1.5% of DNA recovered from more concentrated samples such as activated sludge (Sakcham et al., 2019), Zhang et al 2013). Particularly for drinking water, the ubiquity of exDNA should be considered when selecting an extraction method to represent the intact microbial community.

4.4.2 RNA Extraction

Approaches to RNA extraction are similar in principle to DNA extraction. In fact, some commercial kits offer tandem isolation of both RNA and DNA. However, the relative instability of RNA, the ubiquity of RNases, and the sensitivity of reverse-transcription PCR present additional challenges compared to DNA extraction. RNA extraction is essential for transcriptomic and metatranscriptomic studies. However, RNA extraction in the context of wastewater is often focused on isolating viral genetic material, particularly targeting public health threats such as SARS-COV-2. Viral RNA extraction efficiencies for the same method can be significantly different when tested on enveloped and non-enveloped viruses (Torii et al., 2021). Unfortunately, validation studies on RNA extraction from different types of water and wastewater samples with different viral targets are lacking.

Table 4-6 provides an overview of recent viral RNA extraction validation studies. A more thorough overview of methods used in recent wastewater SARS-COV-2 detection studies is available in Mousazadeh et al. (2021).

Corpuz et al. (2020) provide a recent review of commonly used extraction methods targeting viral nucleic acids. The kits investigated utilize spin column purification methods (QIAamp Viral RNA Mini Kit, Qiagen; DNeasy Blood and Tissue kit, Qiagen; AllPrep DNA/RNA Mini Kit, Qiagen; etc.). Newer methods developed specifically for the isolation of SARS-CoV-2 RNA provide for faster extraction by adding magnetic nanoparticles (MNPs) directly to the lysis solution, allowing lysis and adsorption to occur in a single tube (Parra-Guardado et al., 2021; Ramos-Mandujano et al., 2021; Somvanshi et al., 2020; Zhao et al., 2020). Very few of the reviewed RNA extraction methods include a mechanical lysis step; this may be due to the hypothesized degradation of filter-concentrated RNA when subjected to bead-beating (Kaya et al., 2022). MO BIO Phenol-chloroform-guanidinium thiocyanate extraction and purification or extraction using a high-salt solution with spin column purification can also produce comparable results to RNA extraction kits (Torii et al., 2021; Whitney et al., 2021). Ultimately, no large-scale comparison studies exist to suggest that any kit or method produces the most reliable viral RNA extracts, potentially necessitating preliminary testing prior to extraction of a large sample set.

4.4.3 Automated Extraction Systems

Automated extraction systems such as epMotion® (Eppendorf), KingFisher (ThermoFisher), NucliSENS® easyMAG® (bioMérieux), and QIAcube (Qiagen) offer hands-free extraction of a limited number of samples. These systems are often able to accommodate a variety of extraction methods, particularly magnetic bead-based purification methods. Yields and community data from different automated systems are generally comparable (Verheyen et al., 2012) or even superior (Marotz et al., 2017) to manual extraction kits but require additional capital.

4.4.4 Controls

The selection of an appropriate extraction method based on literature review is critical but does not guarantee that the extraction efficiency will be optimal for a previously uncharacterized sample set. While it would usually be impractical and costly to test several methods on a particular sample type prior to extraction, the inclusion of a positive control analyzed through both extraction-dependent and -independent methods is recommended to determine efficiency. If it is likely that the community to be analyzed contains lysis-resistant taxa (e.g., *Mycobacteria*), it may be useful to include a positive control of an organism to account for biased extraction (Haig et al., 2018). The inclusion of a more susceptible organism as a positive control, such as *E. coli*, would provide a measure of overall DNA recovery to determine absolute abundance (Bonk et al., 2018). Some validation studies employ parallel extraction-independent techniques such as quantitative FISH to verify the results of amplicon data or metagenomics (Albertsen et al., 2015). Sequencing of negative controls (extraction blanks and no-template controls) is also essential in establishing that extracts are representative of the true community and are not affected by contamination from extraction reagents, especially in the case of low-biomass samples such as drinking water (Bautista-de Los Santos et al., 2016; Eisenhofer et al., 2019). At a minimum, publications should consider including raw data such as extract and library DNA mass concentrations alongside data for appropriate controls to provide transparency about potential bias associated with low-concentration samples (Brandt and Albertsen, 2018).

4.4.5 Conclusion

Ensuring a consistent nucleic acid approach with minimal bias and inhibitors is fundamental to producing reproducible and comparable NGS analysis. Extraction methods should be chosen according to two key variables: sample type (e.g., drinking water or wastewater) and the intended targets and sequencing platform (e.g., capturing specific categories of viruses or bacteria, short-read vs long-read sequencing, amplicon vs metagenomic sequencing). Most importantly, comparisons between samples of different types, sources, treatments, or timepoints should consider potential extraction biases, and comparison studies should keep extraction methods consistent in order to ensure that results are comparable. Pilot tests to select the optimal DNA extraction approach for a particular sample type are highly recommended. Ultimately, extraction methods should support analyses where the variables of interest are not overshadowed by differences in extraction efficiency or methodology.

Table 4-4. Commonly used Cell Lysis and Nucleic Acid Purification Methods.
Commercial kits generally achieve nucleic acid extraction through a cell lysis/nucleic acid binding stage followed by a purification step. Methods for these two processes vary between kits, with varying advantages and disadvantages. Detailed descriptions of the principles behind these methods can be found in Ruggieri et al., 2016; Shin, 2018; Tan and Yiap, 2009.

Process	Method	Description	Pros	Cons	Relevant Kits or Methods
Cell lysis	Chemical/ enzymatic treatment	Chemicals (Detergents, chaotropic salts, alcohols) and enzymes (Proteinase K, Lysozyme, cocktails) break open cell membranes/walls and inactivate nucleases.	Necessary in all protocols to provide the appropriate chemical conditions for cell lysis and DNase/RNase inactivation	When used alone, often insufficient in breaking open cells (especially Gram-positive)	QIAamp DNA Stool Mini Kit (Qiagen), Haig et al. 2018, Maghini et al 2021, Blood Mini Kit, DNA Mini Kit (Qiagen), Whitney et al. 2021, QIAamp Viral RNA Mini Kit (Qiagen), TRIzol Plus RNA Purification Kit (Invitrogen)
	Bead beating	Glass, zirconia, or garnet beads of different sizes disturb sample matrix and lyse cell membranes using high-speed vortexing.	Essential for lysing cells trapped in a complex matrix or difficult-to-lyse cells	Fragmentation of genomic DNA increases with increased intensity of bead beating Specialized equipment required	Fast DNA Spin Kit for Soil (MP Biomedicals), DNeasy PowerWater Kit (Qiagen), DNeasy PowerSoil Kit (Qiagen), UltraClean Fecal Kit (Zymo Research)
Purification	Phase separation (Phenol-chloroform)	Aqueous and organic phases in a biphasic emulsion are separated by centrifugation. Nucleic acids in the aqueous phase are precipitated with alcohol.Addition of guanidinium thiocyanate allows RNA to remain in aqueous phase while DNA is removed with organic phase.	Ideal for recovering high molecular weight DNA (minimal shearing) Can more efficiently lyse some cell types (e.g., mycobacteria)	Use of hazardous reagents Phenol can inhibit downstream processes Preceding lysis methods are variable and non-standardized	Haig et al. 2018, Hwang et al. 2012, Hill et al., 2015, Maghini et al. 2021, TRIzol Plus RNA Purification Kit (Invitrogen)
	Spin column	Polymer-coated columns bind nucleic acids in the presence of chaotropic salts; impurities are washed through the column.	Most commonly used method in commercial kits	Binding capabilities and recovery may be lower than those for a suspended solution of beads/particles	DNeasy PowerWater Kit (Qiagen), DNeasy PowerSoil Kit (Qiagen), UltraClean Fecal Kit (Zymo Research), QIAamp Viral RNA Kit (Qiagen), QIAamp DNA Blood Mini Kit (Qiagen), AllPrep DNA/RNA

				May encounter clogging issues	Mini Kit (Qiagen), etc. (i.e., most commercial kits)
	Magnetic beads	Micro- to nano- scale magnetic particles coated with polymer bind nucleic acids. A magnetic rack is used to draw beads out of solution for decanting during washing steps.	Less centrifugation steps, easily automated Provide a visible confirmation of retention of nucleic acids during washing	Often more expensive, not common in extraction kits	PowerMag Soil Kit (Mo Bio), MagAttract PowerClean Kit (Qiagen), KingFisher Flex (ThermoFisher), Parra-Guardado et al., 2021; Somvanshi et al., 2020; Zhao et al., 2020; Ramos-Mandujano et al. 2021, NUCLISENS® EASYMAG® (bioMérieux)
	Silica matrix	Suspended silica particles bind nucleic acids in the presence of chaotropic salts. Particles are loaded into a spin column and washed/centrifuged.	High surface area may provide more binding sites than a coated spin column	May require repeated loading of spin column	Fast DNA Spin Kit for Soil (MP Biomedicals)

Table 4-5. DNA Extraction Validation Studies on Wastewater and Drinking Water Samples.

Study	Samples Extracted	Methods Compared	Key Findings	Overall Recommendation
(Vanysacker et al., 2010)	3 types of activated sludge	Fast DNA Spin Kit MO BIO Ultraclean Soil kit Standard bead-lysozyme protocol based on Boon et al 2000 Combination of QIAamp Mini Kits	Observed a Sludge type/extraction method interaction Fast DNA Spin Kit produced highest yields, most obvious community differences between sludge types Ultraclean kit produced highest richness	Fast DNA Spin Kit (MP Biomedicals)
(Hwang et al., 2012)	Drinking water meter biofilms	PowerSoil DNA Isolation Kit Fast DNA Spin Kit for Soil 3 Phenol-chloroform extraction methods	Phenol-chloroform methods produced the highest yields, but Fast DNA Spin Kit for Soil produced comparable communities and higher purity extracts Fast DNA Spin Kit for Soil may have underestimated richness	Fast DNA Spin Kit for Soil (MP Biomedicals)
WRF Project 4721	Drinking water and premise plumbing water	Fast DNA Spin Kit Fast DNA Spin Kit for Soil DNeasy Power Water Phenol-Chloroform extraction	Kits vary in their recovery when targeting DNA from specific pathogens (*Legionella pneumophila*, *Pseudomonas aeruginosa*, *Acanthamoeba* spp., nontuberculous mycobacteria). DNeasy had optimal yield for most, but Fast DNA Spin Kit was comparable. Phenol-chloroform was best for mycobacteria, which have waxy coatings	DNeasy Power Water (Qiagen)
(Guo and Zhang, 2013)	2 types of activated sludge	MO BIO Power Soil DNA Isolation Kit MO BIO Ultraclean Soil DNA Isolation Kit Fast DNA Spin Kit for Soil ZR Soil Microbe DNA Kit	Fast DNA Spin Kit for Soil produced highest yields, highest purity, and highest abundance of Actinobacteria, but lower richness ZR kit produced low quality extracts, but resulting community still clustered with	Fast DNA Spin Kit for Soil (MP Biomedicals)

		EPICENTRE Soil Master DNA Extraction Kit MO BIO Ultraclean Fecal DNA Isolation Kit QIAamp Stool Mini Kit	Fast DNA Spin Kit-extracted community (unbiased lysis but inefficient purification) Extracts usually around 10kb in length	
(Hill et al., 2015)	Drinking water, WWTP Influent/Effluent	MO BIO UltraClean Microbial DNA Isolation Kit Fast DNA Spin Kit for Soil Optimized UNEX Buffer-bead beating method (developed by Hill et al. 2015)	Developed an extraction method for isolating DNA and RNA of viruses and bacteria using guanidinium isothiocyanate-based lysis buffer (UNEX Buffer), bead-beating, and silica column purification Based on results of PCR (CT values), UNEX method typically outperformed either of the other kits (however, no community data comparisons)	UNEX Buffer – bead-beating – silica column method
(Albertsen et al., 2015)	Activated sludge from a WWTP tank	Fast DNA Spin Kit for Soil MO BIO PowerLyzer PowerSoil DNA Isolation Kit	Variations in bead beating step produced significant variations in community composition and particularly relative abundance of Actinobacteria Variations in extraction efficiencies were also observed on Order level	Fast DNA Spin Kit for Soil (MP Biomedicals), with bead beating step modified to 6 m/s for 160s
(Walden et al., 2017)	Lake water, bulk wastewater, biofilms	QIAamp DNA Mini Kit QIAamp DNA Stool Mini Kit MO BIO PowerWater Kit MO BIO PowerSoil DNA Isolation Kit	All samples group more by sample type than extraction method No one method performed the best across all sample types	QIAamp DNA Mini kit (Qiagen) or MO BIO Power Soil kit
(Li et al., 2018)	Influent, effluent, and activated sludge of 2 WWTPs	FastDNA Spin Kit for Soil MO BIO PowerSoil DNA Isolation Kit	Fast DNA Spin Kit for Soil produced highest yields, highest purities, and	Fast DNA Spin Kit for soil (MP Biomedicals)

		ZR Fecal DNA MiniPrep	most consistent community results for all sample types	
(Niestępski et al., 2018)	WWTP Influent/Effluent	Fast DNA Spin Kit for Soil Genomic Micro AX Bacteria Gravity Kit	For 5 samples, Fast DNA Spin Kit for Soil detected ARGs in wastewater, but Genomic kit did not	Fast DNA Spin Kit for Soil (MP Biomedicals)
(Brandt and Albertsen, 2018)	Tap water	MO BIO PowerWater Isolation Kit Fast DNA Spin Kit for Soil	FastDNA Spin Kit for Soil produced extracts with concentrations below LOD	MO BIO PowerWater Isolation kit
(Haig et al., 2018)	Tap water from 15 households	Fast DNA Spin Kit for Soil Maxwell LEV Blood DNA kit Phenol chloroform method 1 (low concentration SDS) Phenol chloroform method 2 (high concentration SDS)	Phenol-chloroform method 2 recovered 8x more NTM and 3x more total DNA compared to kits Suggests that positive NTM controls quantified by cytometry or fluorescence should be included to confirm extraction efficiency	Phenol chloroform method 2 (high concentration SDS)
(Putri et al., 2021)	Quenched or unquenched tap water, sterile and non-sterile tap water spiked with *E. coli* to test effects of dechlorination on extraction	MO BIO PowerWater Isolation Kit	Only spiked samples that were dechlorinated produced DNA above LOD Higher filtration volumes did not produce higher DNA concentrations Residual chlorine levels at just 0.2 mg Cl2/L can dramatically disrupt DNA extraction	MO BIO PowerWater kit with dechlorination prior to extraction
Note: Some commercial kits described in the referenced studies are no longer available.				

Table 4-6. Viral RNA Extraction Validation Studies on Wastewater Samples.

Study	Samples Extracted	Spiked Target	Methods Compared	Key Findings	Overall Recommendation
(Iker et al., 2013)	Biosolids, feces, surface water	Adenovirus 2 Murine norovirus Poliovirus type 1	MO BIO PowerViral Environmental DNA/RNA Isolation kit Qiagen QIAamp Viral RNA Mini kit Zymo ZR Virus DNA/RNA Extraction kit	PowerViral kit performed best in qPCR for biosolids samples but was comparable to other kits for surface water samples, likely due to superior inhibitor removal	MO BIO PowerViral kit (specifically for high-inhibitor samples)
(O'Brien et al., 2021)	Concentrated wastewater	SARS-CoV-2	Qiagen All Prep PowerViral DNA/RNA kit NEB Monarch RNA MiniPrep Kit Zymo Quick RNA Viral kit with Inhibitor Removal Zymo Quick RNA Fecal/Soil Microbe MicroPrep Kit	Monarch RNA Miniprep Kit was especially sensitive to inhibitors Zymo Quick RNA-Viral with Inhibitor Removal produced highest yields	Quick RNA Viral kit (Zymo Research)
(Zheng et al., 2022)	Concentrated wastewater	SARS-CoV-2	QIAamp Viral RNA Mini Kit TRIzol Plus RNA Purification Kit	QIAamp Viral kit produced higher detection rates with a higher processing capacity over guanidinium thiocyanate method	QIAamp Viral RNA Mini kit (Qiagen)
(Kaya et al., 2022)	Concentrated wastewater	SARS-CoV-2	Qiagen Allprep PowerViral RNA/DNA Zymo Quick RNA Mini prep Zymo Direct-zol RNA Miniprep Plus TRIzol-chloroform (Wu et al 2020)	Zymo Quick RNA Mini prep produced highest recoveries with ultrafiltration as concentration method, followed by Direct-zol	Quick RNA Mini prep (Zymo Research)

(Torii et al., 2021)	Concentrated wastewater	Bacteriophage MS2 (nonenveloped), *Pseudomonas* phage ϕ6 (enveloped), murine norovirus (nonenveloped)	QIAamp Viral RNA Mini Kit TRIzol-based extraction with RNeasy PowerMicrobiome Kit	TRIzol method was more effective for enveloped virus when combined with PEG precipitation as the concentration method, but not when ultrafiltration was used Extraction on non-enveloped viruses produced significantly different results, also varying with the concentration method	Ultimately selected PEG + TRIzol concentration/extraction, but extraction kit choice may vary depending on concentration method.

4.5 Controls for NGS Workflows

There are various types of controls that should be incorporated into NGS workflows to ensure the integrity of the data that are produced. There is potential both for false positives and false negatives in NGS analysis, which can result from contamination in the workflow or from sub-optimal parameters employed in the bioinformatic analysis. Biases and errors introduced during sample preparation, DNA extraction, library preparation, sequencing, and bioinformatic analyses will undermine the accuracy and reproducibility of microbiome profiling (Bokulich et al., 2020). Experimental and process controls used to characterize, model, and correct for uncertainty and error inherent to NGS workflows are emerging (Hardwick et al., 2017), with special consideration when applied to environmental samples (Bharti and Grimm, 2021). Controlling for, or at least characterizing, experimental and technical error propagation is essential for defensible and comparable water quality monitoring data. Generally, there are four key categories of process controls that have been used in NGS workflows to date:

- Replication
- Sequencing mock microbial communities
- Inclusion of spike-in controls
- Analysis of *in-silico* datasets

Used in combination, these controls can greatly aid in reducing the uncertainty in NGS data, help to benchmark new workflows, and provide a path toward reproducible data generation.

4.5.1 Replication

Replicates are essential for capturing and quantifying both experimental and technical variation. Experimental replicates, i.e., biological replicates, refer to the collection of statistically independent samples representing a field or lab condition of interest. Experimental triplicates are common in the analysis of environmental samples for other targets, but it is common to encounter NGS studies in the literature that do not employ replication. For example, biological replication is almost entirely absent from studies employing deep shotgun metagenomic sequencing (Davis et al. 2023). Cost of sequencing is likely a major factor in this deficiency. However, it is important to recognize that no amount of investment is worthwhile if in the end the data generated cannot support the objectives of the project. In such situations, it would be better to scale back the research, than to sacrifice replicates. While triplicates are typically the default, a power analysis is advisable to ensure the appropriate number of replicates to test the hypotheses. Experimental replicates will account for random variation in experimental conditions and sample processing, so that the experimental signal can be distinguished. Technical replicates, i.e., multiple sequencing runs on the same sample, are also advisable, as they will inform if there is any random variation in the sequencing. Most studies that employ technical sequencing replicates report minimal variation (Hendriksen et al., 2019; Roy et al., 2018). Thus, biological replicates are more critical than technical replicates.

4.5.2 Mock Microbial Communities

Mock microbial communities are mixtures of known organisms, or sometimes just their nucleic acids, at known proportions that serve to validate the NGS workflow. Originally, mock microbial communities were developed by the Human Microbiome Project to optimize and standardize DNA extraction techniques for different matrices (Highlander, 2013). Mock microbial communities consisting of intact microbial cells are carried forward from DNA extraction onwards. In this scenario, microbes are chosen to capture typical biases incurred during DNA as a result of different microbial morphologies and recalcitrance to cell wall lysis (e.g., Gram-positive versus Gram-negative bacteria). For example, one commercially available mock community introduced by Zymo Research (ZyMO BIOMICS, cat# D6300) consists of 10 microorganisms at different relative abundances, including three easy-to-lyse Gram-negative bacteria (*Escherichia coli* - 12%, *Pseudomonas aeruginosa* - 12%, *Salmonella enterica* – 12%), five tough-to-lyse Gram-positive bacteria (*Listeria monocytogenes* - 12%, *Bacillus subtilis* - 12%, *Lactobacillus fermentum* – 12% *Enterococcus faecalis* - 12%, *Staphylococcus aureus* - 12%), and two difficult-to-lyse yeasts (*Saccharomyces cerevisiae* - 2%, and *Cryptococcus neoformans* - 2%).

Typically, intact cell mock communities are used as standalone samples that are processed in parallel with experimental samples. DNA extracts from the mock microbial communities are subjected to the same library preparation and sequencing as the experimental samples. Because the organisms are present at known abundance ratios, the efficacy of the DNA extraction protocol to evenly lyse different microbe types can be evaluated based on the ability to accurately reconstruct the mock community downstream (Highlander, 2013). Standards with logarithmically staggered cell counts (e.g., 89.1, 8.9, 0.89, 0.089% relative abundance) have also been developed to assess the LOD of workflows across a 6-log dynamic range (e.g., ZyMO BIOMICS, cat# D6310). These control types are useful for benchmarking new DNA extraction protocols or, more importantly, as regularly included controls for ensuring the reproducibility of existing NGS workflows (Harrison et al., 2021; Weinroth et al., 2022).

When reconstructing mock communities derived from intact cells, it cannot be confirmed if any mis-assemblies are a result of problems with DNA extraction or subsequent library preparation and sequencing. To address this issue, mock communities that consist of exposed genomic DNA only have also been developed to specifically test the effects of library preparation and sequencing platforms on NGS data produced (Kapustina et al., 2021; Manzari et al., 2020). In particular, a limitation to some PCR-based Illumina library preparation kits is amplification bias towards GC-rich or GC-poor regions of microbial genomes (Sato et al., 2019a). Mock communities have therefore been constructed to represent and model the effects of GC variation. Similar in structure to the intact cell mock communities discussed above, these mock communities contain collections of bacterial genomes at known abundances and GC contents ranging from 15% to 85%. Used as standalone samples that are processed alongside experiments, library preparation and sequencing biases can be illuminated and considered in downstream analyses.

4.5.3 Spike-In Controls

Spike-in controls are added directly into samples at a given point of interest along the workflow. These controls can help to account for various biases in nucleic acid extraction, characterize library prep and sequencing bias, and can enable absolute quantification of genomic targets. For example, DNA extraction efficiencies (i.e., the fraction of the total genomic DNA available in a sample that was recovered by the extraction procedure) can be estimated using spike-ins of known abundances of microbial cells that are added directly into samples before extraction. The known abundance of the organism, or set of organisms, can be compared to the resulting read counts after analysis to calculate an efficiency ratio. qPCR can also be used in such situations to determine absolute abundances of recovered cell counts as a fraction of the spike-in as secondary check (Crossette et al., 2021). In principle, this method can only work if the added microbes are not already present in the sample, therefore rare microbial strains are typically used. For example, the ZyMO BIOMICS Spike-In Control II (cat# D6321) uses intact cells from a set of marine organisms that are not expected to be present in freshwaters or wastewater: *Allobacillus halotolerans* (Gram-positive, 10^3 cells per spike), *Imtechella halotolerans* (Gram-negative, 10^4 cells), and *Truepera radiovictrix* (resistant to lysozyme, 10^5 cells). These control types are designed to not only calculate extraction efficiency, but also assess the efficacy of lysing different cell types simultaneously. Additionally, because they are present with a log-abundance distribution, they enable absolute cell number quantification, although these methods still need extensive benchmarking in the field.

More recently, spike-in reference standards have been applied to shotgun metagenomic sequencing for the absolute quantification of gene targets (i.e., quantitative metagenomics). These reference standards are spiked into samples after nucleic acid isolation at known abundances. A recent spike-dependent study by Crossette et al. (2021) spiked in an exogenous genome (*Marinobacter hydrocarbonoclasticus*) at known copy numbers after DNA extraction to determine absolute quantities of ARGs on the Illumina platform. *M. hydrocarbonoclasticus* is a marine organism that was not expected to be in the sample matrix (digested and undigested cow manure) and can therefore act as an independent reference standard. Sequenced reads were mapped to all 4,272 genes comprising the genome and the average ratio of known spiked-in gene quantities to reads mapping to the genome were used to calculate absolute abundance. Quantitative metagenomics is a very promising avenue to pursue, particularly for water samples, where volumetric concentration estimates (i.e., copies per Liter) are of value for modeling and risk assessment purposes.

4.5.4 In-Silico Datasets

The bioinformatic steps used to process NGS datasets are often complex and can be a substantial source of error and bias in workflows (Hardwick et al., 2017). Various software tools have been developed to simulate mock microbial communities with known compositions, library sizes, and sequencing error rates that can be used as in silico datasets. These simulated datasets have proven useful for developing and troubleshooting newly developed software tools or benchmarking existing pipelines. The datasets, typically in FASTQ or BAM format, are constructed to represent a “ground truth” to assess the sensitivity and specificity of bioinformatic analyses (Sczyrba et al., 2017). The obvious limitations of in silico datasets is that they are restricted to assessing bioinformatic process and do not reflect the inherent variation of real data types. The use of simulated data should therefore be used only to supplement the testing of bioinformatic steps and should not replace the use of the experimental standards and controls discussed previously.

4.6 Library Preparation

Library preparation, in the context of NGS, is the procedure for amplifying targeted genetic material and processing the DNA/RNA to be available for sequencing. Library preparation is the first step in the sample processing pipeline where NGS workflows will differ from the processing techniques shared among other molecular analyses. There are a variety of methods available for library preparation, but they all utilize the principle of ligating or labeling DNA or RNA fragments from a sample to adaptors, which are necessary for the specific sequencing method that will be used (van Dijk et al., 2014).

The NGS sequencing method and research questions will impact how library preparation should be accomplished. For metagenomic sequencing, library preparation will target the entirety of genetic material present in the sample. For amplicon sequencing, library preparation will target a specific gene or region of the DNA for amplification, either in simplex (one sample at a time) or multiplex (many samples simultaneously). Most sequencing platforms manufacture library prep kits that are specific to their system, which can be purchased commercially and users can simply follow their manual or adapt the protocol to their specific research questions. Kits are beneficial for library preparation as they are more user-friendly, improve comparability of data, streamline labor requirements, and minimize technician-error, though they are not always standard in amplicon sequencing.

4.6.1 Overview of key steps

While library preparation methods differ based on the specific research or monitoring questions of interest and the sequencing platform used, most library preparation workflows will include the following steps: DNA quantification, fragmentation, size selection (optional), adaptor ligation, library amplification (optional), clean-up, and normalization. These steps are graphically presented in Figure 4-1.

Initial quantification of DNA provides a quality assurance for the DNA extraction process, ensures mass requirements for the selected library prep kit, and is sometimes necessary to determine the amount of reagent needed for subsequent steps such as enzymatic fragmentation. Fragmentation breaks up genomic DNA into a distribution of smaller fragments. There are two methods of fragmentation: mechanical and enzymatic. Mechanical fragmentation physically shears DNA molecules apart, and is commonly performed via sonication. The duration of sonication determines the fragment size distribution. Enzymatic fragmentation utilizes enzymes such as restriction endonucleases to cleave the DNA. The ratio of enzyme to DNA and incubation time determines the fragment size. Mechanical fragmentation is associated with a loss of DNA mass (Tanaka et al., 2020), while enzymatic fragmentation may result in small errors in the final called sequences such as small insertions or deletions or SNPs (Tanaka et al., 2020). Thus, mechanical protocols may be preferable if mutation identification is an objective, and enzymatic fragmentation like that provided by the KAPA Frag Kit for Enzymatic Fragmentation (Roche) may be preferred when DNA concentrations are low and any mass loss is undesirable. Enzymatic fragmentation also typically requires less specialized equipment. Size-selection may be performed afterwards to narrow the range of fragment sizes carried to the next step, though this may be performed at different stages depending on the library preparation approach.

Several steps may follow to repair DNA and achieve attachment of adaptors. After fragmentation, one strand of the DNA may exhibit an overhang. This generally necessitates the end repair step, in which these gaps are filled. Then A-tailing (or adenylation) is performed, preventing ligation of target fragments to each other while creating a suitable ligation site for the thymine on the end of the adapters. With the target fragments prepared, the adapters necessary to index and sequence them for kits such as Illumina TruSeq and New England Bioloab's are attached in a process called ligation. Alternatively, PCR amplification may be used to attach adaptors to library fragments. This approach is particularly common in amplicon sequencing, and simultaneously achieves isolation and enrichment of a target gene (Caporaso et al., 2011). PCR amplification may also be inserted at other points along the workflow for some approaches, particularly when DNA mass is low.

Tagmentation is an alternative approach that simultaneously acts to fragment and ligate the target DNA via random enzymatic insertion of adapters (Burke and Darling, 2016). Accordingly, tagmentation may not integrate into the workflows of other kits. It is employed by the Ilumina Nextera XT and Oxford Nanopore Rapid Sequencing kits.

A final cleanup step removes unattached adapters. This may be achieved by PCR or gel electrophoresis. After this cleanup, libraries can be normalized, pooled, and sequenced.

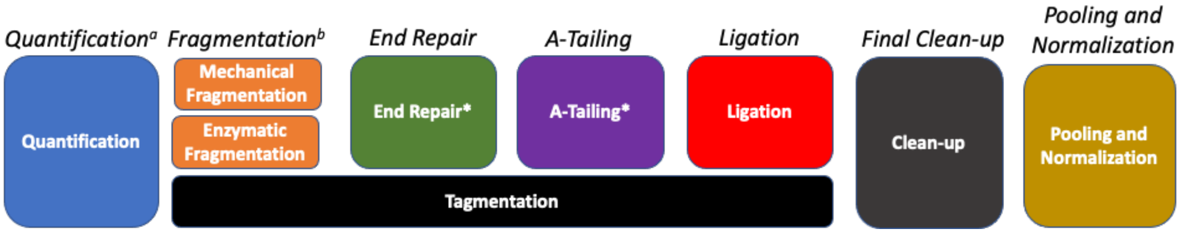


Figure 4-1. Library Preparation Processing Schematic.

[a]May be followed by amplification step for low-mass samples. [b]May be followed by size-selection. Alternatively, may not be desired for long-read sequencing. *Typically followed by an additional clean-up step.

Source: Modified from Hess et al., 2020.

4.6.2 Appropriate QA/QC measures

Many kits offer QA/QC guidance that should be consulted prior to sample collection. Some kits, such TruSeq, include QA/QC oligo controls that should be incorporated into the workflow at specific steps. The reads from these sequences can help identify where failure occurred in library prep.

Other considerations for inputs include:

- DNA should be concentrated appropriately to meet mass requirements for the kit and minimize the need for PCR.
- Check the mass requirements for the library preparation kit. The mass of DNA required may depend on the type of sequencing performed. Also note that the higher mass requirements for HMW sequencing may preclude successful library prep in certain settings.
- Fragment size should be checked using gel electrophoresis or Bioanalyzer and should be appropriate for the kit.

- EDTA, chelating agents, and salts can inhibit library preparation (KAPA Biosystems, 2019)
- Analysis purpose (e.g., antibiotic resistance gene annotation or taxonomic analysis)
- Due to PCR amplification during library preparation and normalization prior to sequencing, results can often only be expressed in relative abundance (i.e., normalized to the number of reads or an internal reference gene).

4.6.3 Commercial Kits

Commercial library preparation kits allow quicker processing and assure more comparable results when using the same kit within or across studies. The most-used commercial kits for NGS are outlined in Table 4-7. When selecting a kit, coordination of the DNA extraction approach and the library preparation kit may be beneficial for improved processing and quality of resulting reads.

Table 4-7. Comparison of Common Commercial Kits for NGS.

Commercial Kit	Manu-facturer	Compatible Sequencing Platform(s)	Fragment-ation[a]	Recommend-ed DNA Starting Concentration	Amplification Step	Reference
TruSeq DNA	Illumina	Illumina	Mechanical	1000 ng	Yes	(Illumina Inc., 2010)
Ultra II DNA Library Prep Kit	New England Biolabs	Illumina	Not indicated	0.5-1000 ng	Yes	(New England Biolabs,Inc., 2020)
Nextera XT	Illumina	Illumina	Tagmentation	1 ng	Yes	(Illumina Inc., 2019)
KapaHyperPrep	Roche	Illumina	Not indicated[b]	1–1000 ng	Yes	(KAPA Biosystems, 2019)
Ligation Sequencing Kit	Oxford Nanopore	Oxford Nanopore	Not indicated, but not necessary	1-3 ng	No	(Oxford Nanopore Technologies, 2019a)
PCR Sequencing Kit	Oxford Nanopore	Oxford Nanopore	Mechanical	100 ng	Yes	(Oxford Nanopore Technologies, 2022)
Rapid Sequencing Kit	Oxford Nanopore	Oxford Nanopore	Tagmentation	400 ng	No	(Oxford Nanopore Technologies, 2019b)
SMRTbell Express Template Prep Kit	PACBIO	PACBIO	Mechanical	1000-2000 ug	No	(PacBio, 2022)

[a] Hess et al., 2020; [b] Enzymatic fragmentation suggested

4.6.4 Guidance for Prep Selection

When designing an NGS study, researchers should consider: equipment access; technical expertise; anticipated DNA mass/concentration; sequencing platform; amplification bias; and pooling. Some library preparation kits require more specialized or expensive equipment than others, or equipment that may be difficult to use in the field, such as magnetic beads, magnetized plates, or sonicators. These considerations will be important ahead of kit selection.

Technician expertise will be another key criterion to consider. Typically library preparation is performed by dedicated technicians employed by sequencing centers. However, some protocols are suitable for lab staff with more basic training. Tagmentation protocols, for example, tend to require fewer steps and time, and so may be preferable for those with less training.

The amount of DNA in the samples or the amount of DNA needed for downstream NGS will be a crucial consideration. When working with low-mass matrices (<1 ng), such as drinking water, a kit with an amplification step may be necessary in order to produce sufficient DNA mass for long-read sequencing. Enzymatic fragmentation may also be preferable in these circumstances to avoid shearing-associated DNA loss in mechanical fragmentation. However, for samples with ample DNA such as WWTP influent, a kit without amplification may better represent the community while still producing sufficient mass. The amount of DNA needed will also depend on the kit, sequencing platform and protocol selected (Table 4-7).

As recent review articles have emphasized the need to systematically characterize and quantify the types of bias incurred in the library preparation step (Coenen-Stass et al., 2018; Hess et al., 2020; van Dijk et al., 2014). Some biases may have minimal relevance on the data interpretation, depending on the research objective. Developing best practices for avoiding or minimizing bias incurred by library preparation would also be helpful.

4.6.5 Limitations & Research Needs

Producing comparable NGS data can be challenging due to variable methods used for sample collection, concentration, library preparation, and sequencing. The biases of library preparation are particularly understudied. More research is needed to determine which kits and methods provide the highest fidelity results for each type of NGS, as well as which preparation method generates the largest quantity of DNA for downstream long-reads sequencing applications, without introducing substantial bias. In addition, further research on matrix-specific (wastewater, drinking water, etc.) method adjustments/optimization would improve accuracy and specificity of environmental studies utilizing NGS.

4.7 Sequencing Methodology and Platform

Sequencing methodology is important to consider when characterizing the microbial community from water and wastewater samples using NGS. Selecting the right sequencing methodology depends on the goals of the analysis, the acceptable cost of analysis, and the intended data analysis approach. It is important to bear in mind the selected sequencing approach and platform when developing the overall workflow from sample collection to analysis. The following are the five main categories of sequencing applications that are commonly applied for the characterization of microbial communities in water or wastewater samples:

- Whole genome sequencing (WGS)
- Shotgun metagenomics

- Amplicon sequencing
- Targeted metagenomics
- Metatranscriptomics

Each of these applications are described briefly in Chapter 1 and presented alongside a synthesis of the manner in which these approaches have been applied to study water and wastewater in Chapter 2. Applications and key considerations for each are detailed below and summarized in Table 4-8. There are a variety of sequencing platforms that are commonly applied for each of the above approaches. These platforms include:

- Illumina
- Ion Torrent
- Pacific Biosciences
- Oxford Nanopore
- ABI SOLiD

The technological basis for sequencing utilized by each of these platforms as well as an overview of the types of data generated by each are summarized in Chapter 1 and below in Table 4-9.

4.7.1 Sequencing Applications

4.7.1.1 Whole Genome Sequencing

WGS is applied to capture the sequence of the entirety of the genome from an microorganism. This organism must first be isolated and purified via culture-based methods before sequencing. Once the organism is isolated and DNA extracted, a variety of sequencing platforms can be utilized to obtain the genomic sequence. Short read sequencing technologies, such as Illumina are most common due to their relatively low cost and high accuracy. However, the emergence of long read sequencing platforms capable of generating the sequence of a single DNA molecule and their continual improvement in accuracy may make them increasingly utilized for this purpose in the future (Shapiro et al., 2013). To reduce the possibility of errors in sequencing carrying over to errors in the sequence of the genome, typically coverages (i.e., the extent to which reference strain genomes are represented in aligned sequencing reads) of 50X or even 100x per target organism are recommended.

To obtain the whole genome, sequencing reads must be strung together using either *de novo* or guided genome assembly. *De novo* assembly involves reconstruction of the genome from reads without using a template, while guided assembly maps reads to a specified reference genome (Ng and Kirkness, 2010). WGS is most appropriate for applications where detailed information about the genetic composition of a specific target organism is desired. For example, WGS has been applied for investigating sources of drinking water outbreaks through comparing genomic similarity among *Legionella pneumophila* isolates (Garner et al., 2019a; Raphael et al., 2016), assessing antibiotic resistance patterns among isolates of organisms relevant to human health, such as *Escherichia coli* and *Klebsiella pneumoniae* (Ekwanzala et al., 2019; Jiang et al., 2019), and identifying catabolic pathways associated with nutrient removal (Chao et al., 2016; Meng et al., 2019). A challenge of WGS is that it may be difficult or impossible to obtain an isolate of the organism performing the function of interest from complex environmental samples.

4.7.1.2 Metagenomic Sequencing

Shotgun metagenomic sequencing, often simply referred to as metagenomic sequencing, is the random subsampling and sequencing of genetic material from a mixed microbial community (Schloss and Handelsman, 2005). Metagenomic sequencing has been applied for a variety of sequencing goals, including profiling overall microbial community taxonomic composition and functional capacities. For example, in studies of water and wastewater, metagenomic sequencing has been used to profile ARGs (Garner et al., 2018a; Stamps and Spear, 2020; Zhang et al., 2015), genes associated with nitrification and denitrification (Cai et al., 2016; Ye et al., 2012), catabolic genes involved in biodegradation (Folch-Mallol et al., 2019; Sidhu et al., 2017), viruses (Bibby et al., 2011; Tamaki et al., 2012), overall microbial community structure (Brumfield et al., 2020; Hull et al., 2017), and pathogen-associated genes within a mixed community (Cui et al., 2019; Kumaraswamy et al., 2014; Saleem et al., 2018). Metagenomic sequencing is highly adaptable to a wide variety of applications, however, the availability of appropriate data analysis workflows and suitable databases for read annotation should be considered before pursuing a new application.

Metagenomic sequencing most often relies on the use of short read sequencing technologies. Short read sequencing generally provides the greatest sequencing depth, which is critical to maximizing overall coverage and correspondingly capturing a significant portion of diverse microbial communities. However, long read sequencing is beginning to also be applied to metagenomic sequencing (Driscoll et al., 2017). The disadvantage of long read approaches is that they yield a lower coverage of the overall metagenome per unit sequencing cost. However, they provide the added benefit of being able to examine long regions of microbial genomes and gain additional insight into genetic context within a genome.

4.7.1.3 Amplicon Sequencing

Amplicon sequencing relies on amplification of a gene of interest via PCR, followed by sequencing of the product. Amplicon sequencing is beneficial for obtaining a high-resolution profile of a known gene of interest, such as a taxonomic or functional gene. Ideal primers applied for amplicon sequencing should anneal to highly conserved regions of the target gene, which maximizes the number of genes of that category captured across the microbial community. In between the conserved primers, the sequence should ideally be highly variable, which will help to distinguish various variations of the gene and increase the overall resolution of the method. Most commonly, amplicon sequencing is applied to a phylogenetic marker for the purpose of taxonomically profiling the composition of a subset of the microbial community. For example, amplicon sequencing of the 16S rRNA gene is widely used to characterize the phylogeny or taxonomy of the members of the bacteria and archaea in a microbial community (Caporaso et al., 2011). Similarly, waterborne eukaryotes (e.g., amoebae and protozoa) have been profiled by targeting the 18S rRNA gene (Bradley et al., 2016), fungal communities by targeting the ITS region (Bokulich and Mills, 2013), and the adenovirus family by targeting the hexon gene (Iaconelli et al., 2017; Kuo et al., 2015).

Amplicon sequencing has become widely popular for characterizing microbial communities from water and wastewater given the broad applicability of the method and its modest cost compared to metagenomic sequencing. Short read sequencing platforms such as Illumina are typically used to facilitate amplicon sequencing, and the Illumina MiSeq is particularly popular for this approach given its ability to generate longer reads than similar platforms. However, the use of long read sequencing to characterize amplicons is also emerging. For example, the ARTIC pipeline can be used to analyze viral nanopore sequencing data generated from tiling amplicon schemes, to characterize targets such as SARS-CoV-2 (Loman et al., 2020). In addition, Haig et al. demonstrated the use of PacBio SMRT to characterize non-tuberculous mycobacteria at the species, subspecies and in some cases even the strain level (Haig et al., 2018).

4.7.1.4 Targeted Metagenomic Sequencing

A variety of methods have emerged that combine principles of metagenomic sequencing and amplicon sequencing for the purpose of enriching targeted fragments of genetic material prior to sequencing. In these approaches, which can be collectively called targeted metagenomic sequencing, primers or probes are designed that are specific to predetermined DNA targets. As part of the library preparation methodology, the primers or probes are utilized to either selectively capture or amplify these target gene regions for subsequent sequencing.

The primary advantage of this approach is that substantial sequencing depth (i.e., the number of times a given nucleotide in a metagenome has been read) can be devoted to characterizing only genes of particular interest. This will substantially reduce the sequencing costs compared to non-targeted approaches. However, it is important to recognize that there are trade-offs. The need to synthesize these targeted primers or probes can also introduce new costs. When novel assays are desired to target gene regions for which primers or probes are not available, the design of these can require extensive expertise and effort.

Targeted metagenomic sequencing was not found to be currently widely applied to study water and wastewater. However, they may become more popular options within the field as approaches are further developed and validated. Key examples of relevance to the water and wastewater field include assays targeting human pathogens (e.g., the Ion AmpliSeq™ Pan-Bacterial Research Panel, ThermoFisher) and antibiotic resistance genes (e.g., the Ion AmpliSeq™ Antimicrobial Resistance Panel, ThermoFisher, and other published research methods) (Lanza et al., 2018; Tamminen et al., 2020). Commercial assays have the benefit of ease of application and consistency of approach. However, a drawback can be that the proprietary nature leads to somewhat of a black box approach that is not suitable to research.

4.7.1.5 Metatranscriptomic Sequencing

Metatranscriptomic sequencing is beneficial when the goal is to capture the activity of the microbes. Metatranscriptomics achieves this by targeting mRNA, i.e., genes that are actually being expressed (Carvalhais et al., 2012). Because sequencing technologies target DNA, the extracted mRNA must first be converted to double-stranded complementary DNA (cDNA). Application of metatranscriptomic sequencing to water and wastewater has been limited to date. This is primarily due to the inherently low concentrations of mRNA and its rapid degradation in complex environmental matrices. This makes mRNA particularly difficult to extract in a manner that ensures sufficient quantity and quality for sequencing. These factors can also contribute to biases in the resulting metatranscriptomic sequencing data, which are difficult to capture and correct for.

To date, metatranscriptomics has been successful applied to characterize expression of antibiotic resistance genes in activated sludge (Liu et al., 2019c), expression of genes related to nitrogen cycling during wastewater treatment (Yang et al., 2020b), and identification of degradation pathways in contaminated wastewater (Delforno et al., 2019; Pei et al., 2020).

Table 4-8. Summary of NGS Sequencing Approaches.

Application	Data Generated	Strength(s)	Weakness(es)
Whole genome sequencing	Genomic sequence of target organism (i.e., chromosomes, plasmids)	Can be used to examine genetic context in terms of host organism and other genes	Must obtain a pure isolate of the target organism
Shotgun metagenomics	Random subsampling of sequences generated from mixtures of microbes	Non-specific approach eliminates need for targeting specific genes of interest	Sequencing depth may not be sufficient to capture target of interest
Amplicon sequencing	Sequences of a single target gene	Multiplexing improves cost-efficiency	Narrow context (i.e., only one gene target); Well-designed target primers required; PCR biases are introduced
Targeted metagenomics	Sequences of a collection of target genes	Multiplexing and enrichment of target genes improves cost-efficiency	Narrow context (i.e., a range of pre-determined gene targets); Well-designed target primers/probes required
Metatranscript-omics	Random subsampling of sequences generated mRNA across the microbial community	Profiles functions actively being carried out by microbes in sample of interest	mRNA is difficult to extract due to low concentration and tendency to degrade, introducing bias to the analysis

Table 4-9. NGS Sequencing Platforms.

Platform / Vendor	Technology	Applications
Illumina	Sequencing by synthesis	Short read sequencing (50-600 bp)
Ion Torrent	pH-based sequencing	Short read sequencing (200-400 bp)
Pacific Biosciences	Single-molecule real time sequencing	Long read sequencing
Oxford Nanopore	Single-molecule real time sequencing	Long read sequencing (up to 20+ kilo-bp)
ABI SOLiD	Sequencing by oligonucleotide ligation and detection	

4.8 Data Analysis

Here the research team refers to an “analysis pipeline” as a series of software, online analytical tools, or scripts used to complete a series of tasks required for the analysis and interpretation of NGS data. One option is to build the pipeline organically by the user, which presents the advantage of user control and tailoring the pipeline to a specific application or research question. However, building pipelines generally requires a high level of expertise. An alternative and more user-friendly approach is to work within pre-constructed pipelines that are designed for a specific application and have default settings and parameters built in that do much of the decision-making for the user. This latter approach is more user friendly and will tend to produce more consistent and comparable analysis; however, it is important to be aware that any assumptions being made are appropriate to the application.

4.8.1 16S rRNA Amplicon Sequencing Data

For amplicon sequencing pipelines (16S rRNA, 18S rRNA, or ITS genes), data analysis is typically done using prebuilt packages of software and programs that incorporate several, if not all, key steps required for data preprocessing, analysis, and visualization. The most widely used, user-friendly, and curated software for 16S rRNA amplicon sequencing is the Quantitative Insights into Microbial Ecology (i.e., QIIME2) (Caporaso et al., 2010; Hall and Beiko, 2018) software and its associated programs. Much of the guidance provided herein will pertain to 16S rRNA gene amplicon sequencing within the QIIME2 software, but many of the considerations and decision-making processes are parallel to other software (e.g., MOTHUR (Schloss et al., 2009)) and amplicons (e.g., 18S rRNA gene for eukaryotes or ITSs for fungi). QIIME2 is operated within the command line interface (CLI) for all computationally-intensive executions, but can typically be run on a standard modern laptop or desktop, although computation times will inevitably increase with decreasing computing power. Intermediate analysis files (i.e., .qza artefacts) can be downloaded locally and analyzed with QIIME2 R software packages and can be very convenient for users not as familiar with CLI operations. Note that Qiime2 is not compatible with non-Illumina sequencing platforms, such Nanopore or PacBio.

QIIME2 pipelines begin by importing raw amplicon sequencing data (either Illumina or IonTorrent platforms) in fastq format as QIIME “artefacts.” The first artefact consists of a feature table of raw sequence read information. The QIIME artefact is the primary currency for the execution of QIIME commands and the software will not recognize “raw sequencing data” before it is imported as an artefact. Once imported, raw sequences are demultiplexed into individual sample files based on user-provided metadata of sample barcodes, paired-end reads are merged via their overlapping regions, and primers and barcodes are removed to generate full-length and clean amplicon sequences Figure 4-2. To reduce the computational requirements of downstream analyses, amplicons are dereplicated to identify representative sequences for each putative, biologically-relevant feature present in the dataset whilst maintaining the abundance information for each amplicon. In some instances, sequencing cores can provide the clean amplicon data directly to customers that can then be used as input to the following analyses.

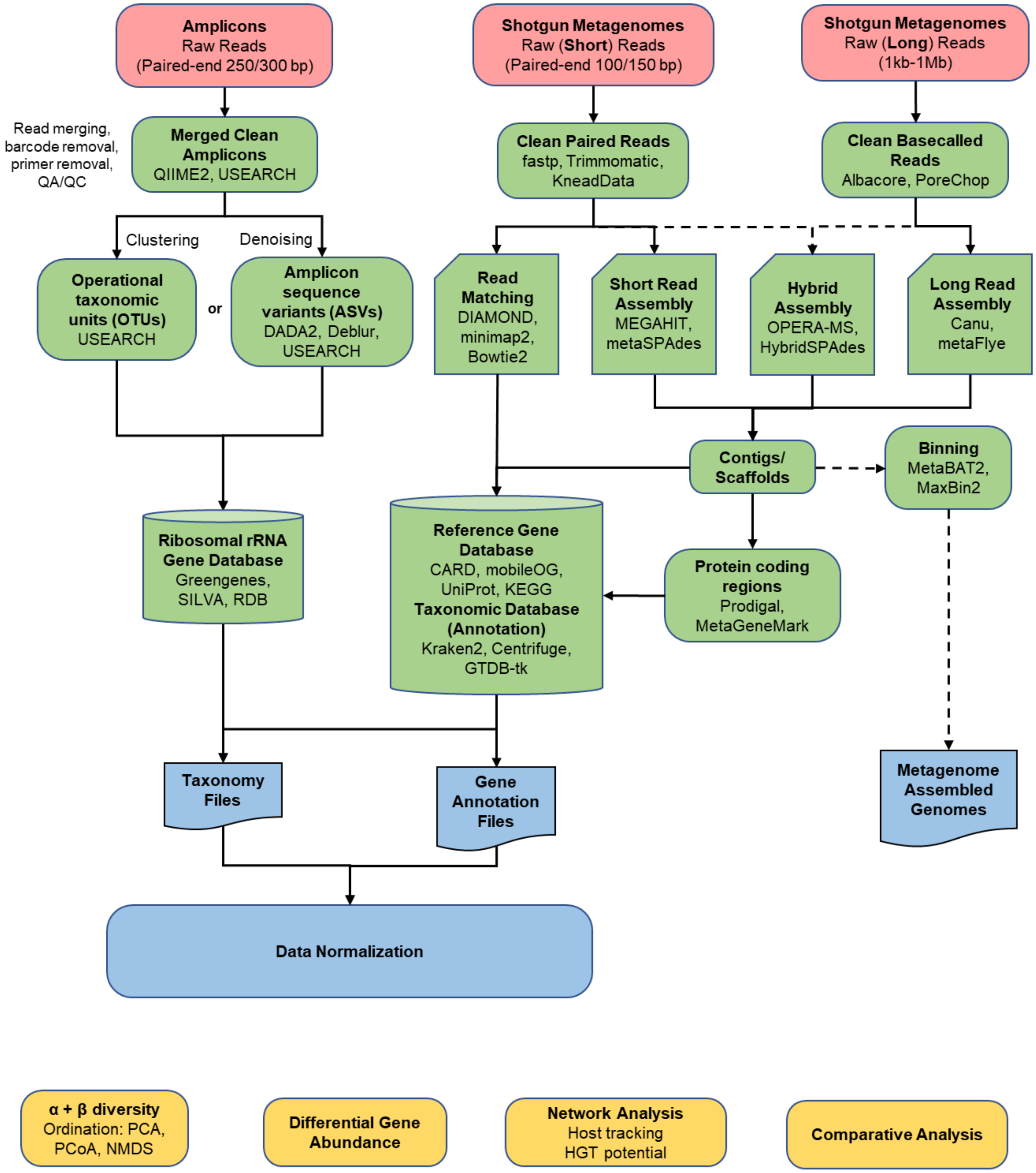


Figure 4-2. Overview of Possible Analysis Pipelines for Amplicon and Shotgun Metagenomic Sequencing Data.

The most critical decision made by the user/researcher for the analysis of amplicon sequencing data is whether to produce and analyze operational taxonomic units (OTUs) or amplicon sequence variants (ASVs) (Figure 4-2). The decision is made based on the relative specificity of taxonomic classification that is sought by the user and the quality of the amplicon data that has been generated.

OTUs are chosen by clustering amplicons by a defined percent nucleotide identity (typically 97-99%) to generate representative sequences for each detected biological unit. Lumping together sequences reduces the rate at which spurious sequences are interpreted as true biological variation, although this method may underutilize the accuracy of modern sequencing technologies and their ability to resolve fine-scale variation in taxonomic composition (i.e., at the species or subspecies scale). Amplicon clustering can be achieved by the USEARCH ((Edgar, 2010)) and UPARSE (Edgar, 2013) algorithms that are housed within the QIIME2 software package itself.

ASVs are becoming a more common choice because they represent more precise taxonomic resolution. ASVs consist of singular unique amplicons representing biologically-relevant features and are produced by denoising the sequencing data, i.e., statistically disentangling biological variation from sequencing errors. The denoising method can be executed with various software including DADA2 (Callahan et al., 2016), Deblur (Amir et al., 2017) executed in QIIME2, or USEARCH. Because of the ever increasing accuracy of modern sequencing platforms, researchers have begun to favor the analysis of ASVs for finer-scale taxonomic classifications to better illuminate ecological niches, or pathogens from commensal organisms (Callahan et al., 2017). Whether analyzing OTU or ASV data types, subsequent taxonomic classification steps described in the following sections, remain the same.

The most computationally-intensive process in 16S rRNA gene amplicon sequencing data analysis is training a taxonomic classifier. Taxonomic classifiers are probabilistic models (e.g., Naïve-Bayes classifiers) that have been trained on large databases of reference genes that enable the accurate classification of new, even novel query amplicons. The three main collections of databases used to train classifiers are the Greengenes (DeSantis et al., 2006), SILVA (Quast et al., 2013), and Ribosomal Database Project (Cole et al., 2014) and many researchers choose their preferred database for classifier training (and most perform similarly well (Park and Won, 2018). For accurate and comprehensive classifications, regardless of the database of reference genes, the QIIME2 classifier must be trained based on (1) the primer set used to generate the amplicons (i.e., the hypervariable region amplified) and (2) the percent identity used for clustering the raw data (100% identity in the case of ASVs). Classifiers do not need to be retrained and can be run on new sample sets given the same hypervariable region and clustering percentage. Finally, using the classifier, OTUs or ASVs can be assigned a taxonomy (kingdom, phylum, class, order, family, genus, species) providing neat count tables that can be downloaded and analyzed in third-party software, such as R or python. The QIIME2 software also provides a large array of in-house analysis and visualization tools including building phylogenetic trees, alpha- and beta-diversity analysis, ordination, and several others. All these techniques and analysis approaches can be found in the QIIME2 documentation (https://docs.qiime2.org/2022.2/tutorials/).

4.8.2 Shotgun Metagenomic Data

A typical metagenomics workflow will consist of (1) quality control of raw DNA sequence reads; (2) alignment of these reads to databases of monitoring targets or processing of taxonomic information, and, optionally, the assembly of quality filtered raw reads into longer stretches of contiguous genome sequences (“contigs”). Depending on where the data are generated and the choice of sequencing technology, there may be an additional preliminary step required to process files directly produced from the sequencer into raw DNA sequences. However, this is typically performed by the sequencing center. In the subsequent sections, it is assumed that the starting point for analyses are raw sequence files (fastq files in either plain text or compressed format (extensions may include .gz, .bz, .tar, among others).

4.8.3 Detecting Target Genes for Monitoring Purposes

Unlike qPCR, metagenomics can theoretically produce a single-run profile of multiple marker genes or monitoring targets simultaneously without *a priori* knowledge of sample composition or selection of targets. For many informative monitoring targets, e.g., ARGs, MGEs (e.g., *intI1*), public databases of nucleotide or amino acid (protein) sequences exist. These resources can be used as references to precisely detect their occurrence within samples through bioinformatics, the computational analysis of biological data. Importantly, many sequences recovered through shotgun metagenomics of environmental matrices will be similar in nucleotide or amino acid sequence to those in public databases. However, only a select subset of the sequences in the metagenome will plausibly be derived from the monitoring targets. To account for this, annotation criteria are established to ensure accuracy and sensitivity of the predictions in standard bioinformatic workflows.

Detecting monitoring targets using public databases relies on sequence aligners, a type of bioinformatic algorithm, that perform partial string alignment of sample-derived nucleotide or amino acid sequences with those of reference databases. The most widely used sequence aligners are the Basic Local Alignment Search Tool (BLAST), which scores partial string alignments based on the number of exact matches, partial matches, gaps, and an evolution-informed substitution matrix. While the original BLAST algorithm is still in use, it is generally infeasible for the size of NGS data. By contrast, aligners such as diamond blastp or blastx preserve the essential features of BLAST but in an efficient implementation, enabling the processing of NGS data.

The important output parameters for BLAST-related annotation software like diamond include the length of the alignment, the identity value (i.e., the proportion of perfect matches to mismatches), the bit-score (a normalized alignment score), and the e-value (a statistic that reports the likelihood of achieving a given bit-score given the search space (defined as length of sequence * length of reference database)). The appropriate choice of cut-off for these statistics depends on the reference database used and the nature of the targets themselves. For ARGs, 80% amino acid identity across a minimum of 25 amino acids is typical for the annotation of short Illumina reads (of 150 bp). The minimum alignment length of 25 amino acids (as generated using blastx) ensures that at least half of the 150 bp read will be aligned to a given potential ARG sequence. The 80% identity criterion is moderately strict. Identity between a reference sequence and a target sequence is based on the evolutionary history of the reference and target sequence making it difficult to assign a single value. However, higher identity values reduce the likelihood of false positives in general. Last, it should be noted that alignment-based annotation of long read data (such as that generated by nanopore sequencing) requires frameshift-aware methods of alignment. This is because some aligners, like diamond blastx, perform dynamic translation of the nucleotide sequence of the reads into amino acids in all six possible reading frames (i.e., three potential start positions on both the positive and negative sense strand). At present, long read technologies frequently contain insertion or deletion errors which would disrupt the dynamic translation process and thus provide the incorrect protein sequences. By contrast, using DNA-DNA alignment methods (e.g., minimap2) would circumvent this problem.

4.8.4 Assessing Taxonomic Composition of a Sample Using Illumina Short Reads

A standard objective of shotgun metagenomics is to assess the microbial composition of a sample and detect the presence of potentially harmful bacteria or other monitoring targets. Many tools for such analysis exist and in general rely on publicly available genomes of bacteria, viruses, fungi, and protozoans. Importantly, the size of the typical Illumina short read (e.g., 151 bp) means that the sequences will have relatively sparse signals for taxonomic classification and thus may be unreliable, particularly for poorly-characterized environments. However, metagenomics can theoretically provide a more sensitive detection of specific genera (e.g., pathogens). Thus, if the aim of the experiment is to characterize general changes to the profiles of microbes in a set of samples, 16s rRNA sequencing (as detailed above) may be preferable, why metagenomics may be more suited to specific pathogen targeting, with the added benefit of being able to identify genes of interest in parallel.

4.8.5 Assembling Data into Longer Stretches of DNA (Contigs)

As mentioned above, the short length of Illumina reads precludes some more refined analyses of metagenomes. Thus, it is often desirable to assemble the short reads into longer stretches of DNA, or “contigs.” Variably, some assembly algorithms will produce scaffolds as a final result. These are multiple contigs that, based on bioinformatic evidence, have been merged by the assembly algorithm into a single long genome fragment. Assembly is a computationally intensive and error-prone process. Yet, it can yield valuable insight into the functional potential and fine-grained taxonomic composition of a sample’s microbiome. One example of such a research question might be to identify the putative taxonomic host for an ARG. This would not otherwise be possible using exclusively short reads. A general rule of thumb among bioinformaticians has been that targets within a metagenome having about 5x coverage can be accurately assembled, although there are many confounding factors that can influence assembly accuracy. For complex environmental matrices, like that of wastewater, assembly is especially challenging because of the simultaneous occurrence of many closely related species or strains of microorganisms in a single sample. Some examples of errors that can occur include insertions and deletions, or more problematic, chimeras (a single contig that is the product of two distinct genomes).

4.8.6 Annotation Databases

A wide range of databases exist for annotation of target genes from NGS datasets. Some of the available databases for targets such as 16S rRNA genes, pathogens, and ARGs are presented in Table 4-10.

Table 4-10. Examples of Available Databases for Annotation of Specific Targets from NGS Data.

Target Genes / Organisms	Database Name	Last Updated*	Reference
16S rRNA genes	Greengenes	May 2013 (v. gg_13_5)	(DeSantis et al., 2006)
16S rRNA, 18S rRNA, 23SrRNA, 28S rRNA genes	SILVA	August 2020 (release 138.1)	(Quast et al., 2013)
16S rRNA, 28S rRNA genes	Ribosomal Database Project (RDP)	August 2020 (v. 18)	(Cole et al., 2014)
Pathogens	PATRIC	June 2019	(Wattam et al., 2014)
Pathogens	NCBI Pathogen Detection	August 2022	(NCBI, n.d.)
Pathogens	MyPathogen Database (MPD)		(Zhang et al., 2018)
Eukaryotic Pathogens	VEuPathDB	November 2022 (release 60)	(Amos et al., 2022)
Antibiotic Resistance Genes	Comprehensive Antibiotic Resistance Database (CARD)	September 2022 (v. 3.2.5)	(Alcock et al., 2020)
	Functional Antibiotic Resistance Metagenomic Element Database (FARME-DB)		(Wallace et al., 2017)
	Resfams	January 2015 (v. 1.2)	(Gibson et al., 2015)
	ResFinder	September 2021	(Zankari et al., 2012)
	deepARG	September 2017	(Arango-Argoty et al., 2018)
	ARGminer	April 2019 (v. 1.1.1)	(Arango-Argoty et al., 2020)
Mobile Genetic Elements	A CLAssification of Mobile genetic Elements (ACLAME)		(Leplae et al., 2010)
	mobileOG-db	August 2022 (v. 1.6)	(Brown et al., 2022)
	The Gypsy Database (GyDB)	2011 (v. 2.0)	(Llorens et al., 2011)
Plasmids	COMPASS	March 2020	(Douarre et al., 2020)
Integrons	INTEGRALL	2008 (v. 1.2)	(Moura et al., 2009)
Insertion Sequences	Isfinder		(Siguier et al., 2006)
Transposons	The Transposon Registry		(Tansirichaiya et al., 2019)
Integrative and conjugative elements	ICEberg	September 2018 (v. 2.0)	(Liu et al., 2019a)
Metal and Biocide Resistance Genes	Antibacterial Biocide & Metal Resistance Database (BacMet)	March 2018 (v. 2.0)	(Pal et al., 2014)
Metabolic Genes	Carbohydrate-Active Enzymes Database (CAZy)	January 2022	(Lombard et al., 2013)
Protein Function	Kyoto Encyclopedia of Genes and Genomes (KEGG)	November 2022 (release 104.1)	(Kanehisa and Goto, 2000)

	Clusters of Orthologous Groups of proteins (COG)	March 2022	(Tatusov et al., 2000)
	UniProt	April 2022	(The UnitProt Consortium, 2019)
	SEED		(Overbeek, 2005)
	UniRef		(Suzek et al., 2007)
Varies	NCBI RefSeq	November 2022	(O'Leary et al., 2016)
*As of December 7, 2022- blank cells indicate that the information could not be found, which could be a result of either no update since original publication or a continuous update mode			

4.8.7 Read Assembly

Metagenomic assembly is a computational technique used to reconstruct microbial genomes from shorter fragments of DNA sequences obtained from NGS. In general, assembly of reads is beneficial for gaining a better understanding of the genetic composition of difficult to culture microorganisms from within complex microbial communities (Alneberg et al., 2018; Handley et al., 2014). Assembly of metagenomes derived from environmental samples is commonly applied for two main purposes:

- Obtaining genetic context of the gene of interest (e.g., is an ARG present on a plasmid and therefore potentially mobile)
- Recovering whole genomes from metagenomes, i.e., MAGs

For such purposes, assembly is commonly applied to Illumina sequencing data, because the reads generated (100 - 150 bp) are too short to convey information about neighboring genes (Forsberg et al., 2014; Mao et al., 2015; Pal et al., 2016). The short length of these reads precludes many important analyses for understanding the taxonomy, function, and abundance of microbes present in the environment.

Long read sequencing provides enough sequence length to capture neighboring genes, but typically not whole genomes. Recent advances in long read sequencing technology have enabled the generation of reads of greater than 25,000 bp in length, allowing for more in-depth genome sequence analyses (Ardui et al., 2018). Still, assembly of long reads is often desirable, especially for obtaining MAGs. For instance, metagenomic assembly has been used to recover complete genomes from activated sludge, aiding in the characterization of key microbes involved in phosphate removal, nitrification, and other functions of interest.

Modern day assemblers are typically classified according to the kinds of reads that they are appropriate to assemble, as either short read (i.e., Illumina), long read (e.g., nanopore or PacBio sequencing), or hybrid assemblers which leverage both read types. A summary of approaches available for read assembly is provided in Table 4-11.

4.8.7.1 Assembly Approaches

Short Read Assembly

Depending on choice of sequencing methodology, Illumina reads range in size from 25 – 250 bp, with 150 bp representing a common choice. By contrast, the average bacterial genome size is 5 × 10^6 base pairs (Land et al., 2015). Thus, while 150 bp may capture individual genes, the amount of biological information contained in each individual read is insufficient to capture important data, such as the microbial species of origin, the potential horizontal mobility of an antibiotic resistance gene, among others. Most modern assemblers rely on de Bruijn Graphs (dBgs), which are *n*-dimensional directed graphs that represent overlaps of DNA sequences of length *k* (i.e., *k*-mers) extracted from the short reads, and connections between adjacent *k*-mers as edges (Ayling et al., 2019). However, short read assembly using dBgs is a computationally expensive method and can be prohibitive for metagenomic libraries with millions of reads, or of environments that contain many closely related strains (Ayling et al., 2019; Bradley et al., 2015). In this latter case, the dBgs are confounded by the presence of closely related strains due to the partial overlap of *k*-mers originating from different organisms. Furthermore, the selection of the size of *k* (i.e., the length of the read fragment extracted to construct the graph) has important implications for assembly quality. To address this issue, the iterative de Brujin based assembler (IDBA) was proposed (Peng et al., 2010). Briefly, IDBA performs multiple rounds of dBg-based assembly, with the contigs produced from preceding run used as reads in the subsequent iteration. In general, depth of coverage of different of genomic fragments (i.e., $\frac{N_{reads\ mapped} \times Length_{reads}}{fragment\ length\ (bp)}$) are used to filter spurious *k*-mer overlaps based on the assumption that the bacterial genome present in the sample should have approximately uniform coverage over the length of its genome. However, this assumption is not valid for metagenomes where multiple related species or strains might co-occur (Ayling et al., 2019; Peng et al., 2012). For this reason, IDBA-UD (uneven depth) was proposed. IDBA-UD builds upon IDBA by including iterative depth of coverage-based filtering that becomes more stringent with subsequent iterations. As previously mentioned, the construction of dBgs and subsequent analyses poses a computationally expensive challenge, and so MEGAHIT was introduced leveraging a similar iterative *k*-mer strategy with a compressed variation of dBgs, termed succinct de Brujin graphs (Li et al., 2016). The use of this compressed strategy enables the efficient processing of large and complex metagenomic libraries.

Finally, metaSPAdes (Nurk et al., 2017) is another popular short read assembly pipeline. In contrast to IDBA-UD and MEGAHIT, metaSPAdes analyzes an initial dBg for different structures that are caused by biological variation, that is, the presence of multiple related species or strains, including a final iterative repeat resolution step that aims to provide solutions to the multiple strain problem. metaSPAdes, MEGAHIT, and IDBA-UD remain popular assemblers, however one direct comparison to wastewater-derived metagenomes found superior performance of MEGAHIT and metaSPAdes relative to IDBA-UD (Brown et al., 2021).

Long Read Assembly

Long reads, such as those generated by nanopore or PacBio sequencing platforms, can span difficult-to-assemble regions of genomes in microbes that may not be resolved by short read assemblers (Pollard et al., 2018). Thus, long read sequencing has emerged as a way to directly capture long genomic regions without the need for the computationally expensive and error-prone process of assembly. However, at present, the error rate of long read sequencing is significantly higher than for short read sequencing (generally accuracies range from 85-95%) (Pollard et al., 2018) and it often remains desirable to assemble long reads to recover larger, genome-sized assemblies and thus several long-read assemblers exist. Though not designed with metagenomics in mind, Canu (Koren et al., 2017) is a slow but accurate long read assembler that first corrects errors in the reads, and then leverages a hash-based overlap detection algorithm followed by a sparse graph-based assembly. While providing accurate and high-quality assemblies, one wastewater study found that the computational resources required to assemble a medium size metagenomic dataset could be prohibitive, even for a high-performance computing cluster. By contrast, metaFlye (Kolmogorov et al., 2020) is a fast and computationally efficient method for long read assembly. Unlike Canu, metaFlye forgoes read-error correction. Instead, it employs solid or high-frequency *k*-mers at both a global (i.e., all reads) and local (individual read) level to identify overlaps at non-uniform coverages.

Hybrid Assembly

Hybrid assembly is another proposed solution for recovering genomes from metagenomic data. In this instance, both long and short read technologies are leveraged. Examples of hybrid assemblers that have been effectively harnessed for metagenomic data include HybridSPAdes (Antipov et al., 2016) and OPERA-MS (Bertrand et al., 2019).

As with all sections mentioned so far, assembly is an area of much research and thus the recommendations provided here should serve as a suitable starting point for interested users.

Table 4-11. Summary of Approaches Available for Read Assembly.

Approach	Strengths of Approach	Weaknesses of Approach
Short read assembly	The majority of existing assemblers and metagenomic tools have been designed with short reads in mind. **Key tools:** • **MEGAHIT:** fast, deals with uneven depth, appropriate for metagenomics. • **metaSPAdes:** accurate, generates scaffolds and thus can generate larger genome fragments.	Short reads are often unable to assemble repeat rich regions due to ambiguity in the de Bruijn graph. **Key tools:** • **MEGAHIT:** does not generate scaffolds. • **metaSPAdes:** slow and computationally expensive.
Long read assembly	Long reads can span difficult to assemble regions and thus can provide large assemblies. **Key tools:** • **metaFlye:** fast and expansive, options include advanced plasmid recovery. • **Canu:** produces highly accurate assemblies with error corrections. Assembly is highly customizable.	Long reads are a relatively new technology; best practices are still emerging. Error-rate can impact downstream analyses. **Key tools:** • **metaFlye:** has been shown to be less accurate in some cases. • **Canu:** extremely computationally intensive. Using advanced options

		require expertise or research.
Hybrid assembly	Combines the accuracy of short reads with the ability of long reads to span regions that are difficult to assemble with short reads. Can provide large, accurate assemblies. **Key tools:** • **HybridSPAdes:** produces long and accurate assemblies • **OPERA-MS:** produces long and accurate assemblies, designed for metagenomes, and includes a pseudo-binning step for genome recovery. Allows choice of short read assembler.	Requires sequencing on both short and long read platforms, which can be prohibitively expensive. **Key tools:** • **HybridSPAdes:** because it starts with metaSPAdes, it is slower. It was also not designed with metagenomics in mind.

CHAPTER 5

Case Studies Demonstrating the Application of NGS Technologies to Study Water and Wastewater

5.1 Overview

NGS has been applied in many innovative ways to address challenges faced by water utilities and to answer questions of great importance to improving water, wastewater, and reuse design and management. The breadth of applications of NGS in the water and wastewater industry were described in detail in Chapter 2. The research team has further developed a suite of case studies to exhibit key field-scale applications of NGS and demonstrate their relevance to answering questions relevant to particularly pressing challenges to the water industry. The following five case studies are presented in this chapter:

(1) Pathogens in simulated reclaimed water distribution systems identifying pathogens,

(2) Profiling antimicrobial resistance in surface water

(3) Pathogen screening using 16S rRNA gene amplicon sequencing in environments impacted by extreme flooding events

(4) Aerosolization of Viruses and Associated Risks at Wastewater Treatment Plants

(5) Functional application of metagenomics in wastewater

5.2 Case Study #1: Pathogens in Simulated Reclaimed Water Distribution Systems

This work is described in additional detail in Ghosh et al., (2021).

Detecting and quantifying pathogens in environmental samples can be particularly challenging due to their low abundances and the large numbers of previously uncharacterized bacteria that reside in environmental niches. This case study demonstrates that NGS targeting the metagenome coupled with an in-house pathogen quantification pipeline leveraging UniRef90 gene-family annotation by HUMAnN3 can be used to track changes in relative and absolute abundances of a broad range of pathogens in environmental water and biofilm samples. A major strength of this method is its ability to compare across multiple samples and accurately estimate changes in pathogen abundances. The NGS-based method described here can potentially be extended to quantify water-borne pathogens from diverse environmental sample, including drinking water or wastewater or the natural environment.

5.2.1 Introduction

Pathogenic microbes can evade removal during water and wastewater treatment and proliferate in distribution systems, as in the case of drinking or reclaimed water (Garner et al., 2018b; Savin et al., 2020; Waak et al., 2019). Culturing followed by enumeration of indicator bacteria, the fecal coliforms, has been the method of choice for tracking pathogens in the water industry. However, this approach has a significant drawback: enumeration of indicator bacteria may miss some pathogens, particularly pathogens not of fecal origin (Harwood et al., 2005). Non-fecal OPs, including *Pseudomonas* spp., *Legionella* spp. and non-tuberculosis mycobacteria (NTM), are emerging as major concerns in distribution system pipes and in house-hold plumbing (Falkinham et al., 2015). Culture-based assays and molecular methods, for example, qPCR, can be used to detect and quantify diverse target pathogens (Li et al., 2019a, 2015; Wang et al., 2012; Whiley and Taylor, 2016). However, culture-based assays are time intensive and developing and testing qPCR assays for different pathogens can be challenging.

NGS approaches targeting the metagenome provide a means of sequencing genetic material from all microorganisms, including potential pathogens. Many relevant taxonomic and functional markers can be captured, rather than just relying on the sequencing of a single housekeeping gene, such as the 16S rRNA gene. Representative strains of most pathogens have been subject to WGS, and thus it is possible to compare metagenomic sequences to these to identify pathogen markers. However, most publicly-available platforms act to annotate the whole microbial community, and not just pathogens. In addressing this challenge, an in-house metagenome-derived pathogen quantification pipeline was developed leveraging UniRef90 gene-family annotation by HUMAnN2/HUMAnN3, abbreviated henceforth as the "*pathogen quantification pipeline*" (Beghini et al., 2020; Franzosa et al., 2018) github.com/sudeshna-ghosh/Pathogen-fluctuations).

The *pathogen quantification pipeline* was validated against a mock community spiked into samples and independent measurement of targets via qPCR (Ghosh et al., *in review*). The in-house *pathogen quantification pipeline* was able to better estimate changes in abundance of pathogens between samples compared to two other available taxonomic annotation pipelines MetaPhlAn2 (Truong et al., 2015) and Kraken2 (Wood et al., 2019). The goal of this case study is to demonstrate the feasibility of using metagenomic NGS coupled with an in-house *pathogen quantification pipeline* to quantify and estimate changes in relative and absolute abundances of about forty water-borne pathogenic genera in reclaimed water distribution systems (RWDSs).

5.2.2 Methods

The persistence and re-growth of multiple pathogens were evaluated in six simulated RWDSs described in detail by Zhu et al. (Zhu et al., 2020a, 2020b). The RWDSs were fed WWTP effluent, collected from a local treatment plant in Virginia, USA, treated with or without BAC-filtration, followed by one of three secondary disinfectant conditions: chlorination, chloramination, or no disinfectant, and run in parallel. Replicate water quality and microbiological samplings were carried throughout the distribution system from water and pipe biofilm at different water ages and at different points in time. Results described here are specifically from samples collected during the 30°C run of the RWDSs (Zhu et al., 2020a). Sample processing, DNA extraction, metagenomic sequencing and data analysis methods are described in Table 5-1. Details about available pathogen databases and list of pathogens tested are available at github.com/sudeshna-ghosh/Pathogen-fluctuations.

Table 5-1. NGS Methodologies Applied in Case Study #1.

Procedure Component	Approach/Description	Reference
Source	Simulated Reclaimed Water Distribution System	
Sample Concentration	Bulk water and pipe biofilm samples were collected from a simulated RWDS	(Zhu et al., 2020a)
Nucleic Acid Extraction	FastDNA® SPIN Kit (MP Biomedical, Inc., Solon, OH) and FastPrep® Instrument (MP Biomedical, Inc., Solon, OH)	
Sequencing Approach	Metagenome sequencing	
Library Preparation	Nextera XT DNA Library Prep Kit (Illumina, USA). Note: the method has also been tested on libraries prepared using the Nextera Mate Pair Library Prep Kit (Illumina, USA)	
Sequencing Platform/Configuration	Illumina HiSeq 2500 rapid run mode with 2×100 bp pair-ended reads and the Illumina NextSeq with 100 bp pair-ended reads. Note: the method has also been tested on reads from Illumina NovaSeq 6000 S1 system with pair-ended 2 × 150 bp flow cells	
Sequencing Coverage		
Data Analysis pipelines	HUMAnN3, https://github.com/sudeshna-ghosh/Pathogen-fluctuations	(Beghini et al., 2020; Ghosh et al., 2021)
Annotation Database(s)	UniRef90, the 'unique to genus' pathogen database in https://github.com/sudeshna-ghosh/Pathogen-fluctuations	(Ghosh et al., 2021; Suzek et al., 2007)

5.2.3 Results

Several pathogenic genera, including fecal and non-fecal pathogens, were detected in the RWDSs (Figure 5-1a showing pathogenic genera from two individual samples at the stated conditions). These include fecal pathogens, *Aeromonas, Citrobacter, Enterobacter, Escherichia, Klebsiella, Providencia, Salmonella* and *Vibrio* spp. and non-fecal pathogens, *Acinetobacter, Burkholderia, Chryseobacterium, Elizabethkingia, Haemophilus, Legionella, Pseudomonas, Staphylococcus, Stenotrophomonas* and *Streptococcus* spp. along with mycobacteria, *Mycobacterium, Mycolicibacterium, Mycolicibacter* and *Mycobacteroides* spp. (the new mycobacteria genera are described in Gupta et al., 2018a). An analysis of all the samples collected during the 30°C run found that fecal pathogens are enriched and attenuated under conditions different from several of the non-fecal pathogens (Figure 5-1b). Notable among them were *Legionella* spp. and mycobacteria.

Finally, the potential of metagenome sequencing in estimating changes in absolute abundance was demonstrated. Absolute abundances were computed from relative abundances by normalizing with corresponding relative abundances of 16S rRNA genes from metagenomic samples and then multiplying by 16S rRNA gene copy numbers enumerated via qPCR (Suzuki et al., 2000). The metagenome-derived absolute abundance quantifications were well correlated with qPCR (Nazarian et al., 2008; Radomski et al., 2010) of *Legionella* spp. and mycobacteria (Figure 5-2a-b).

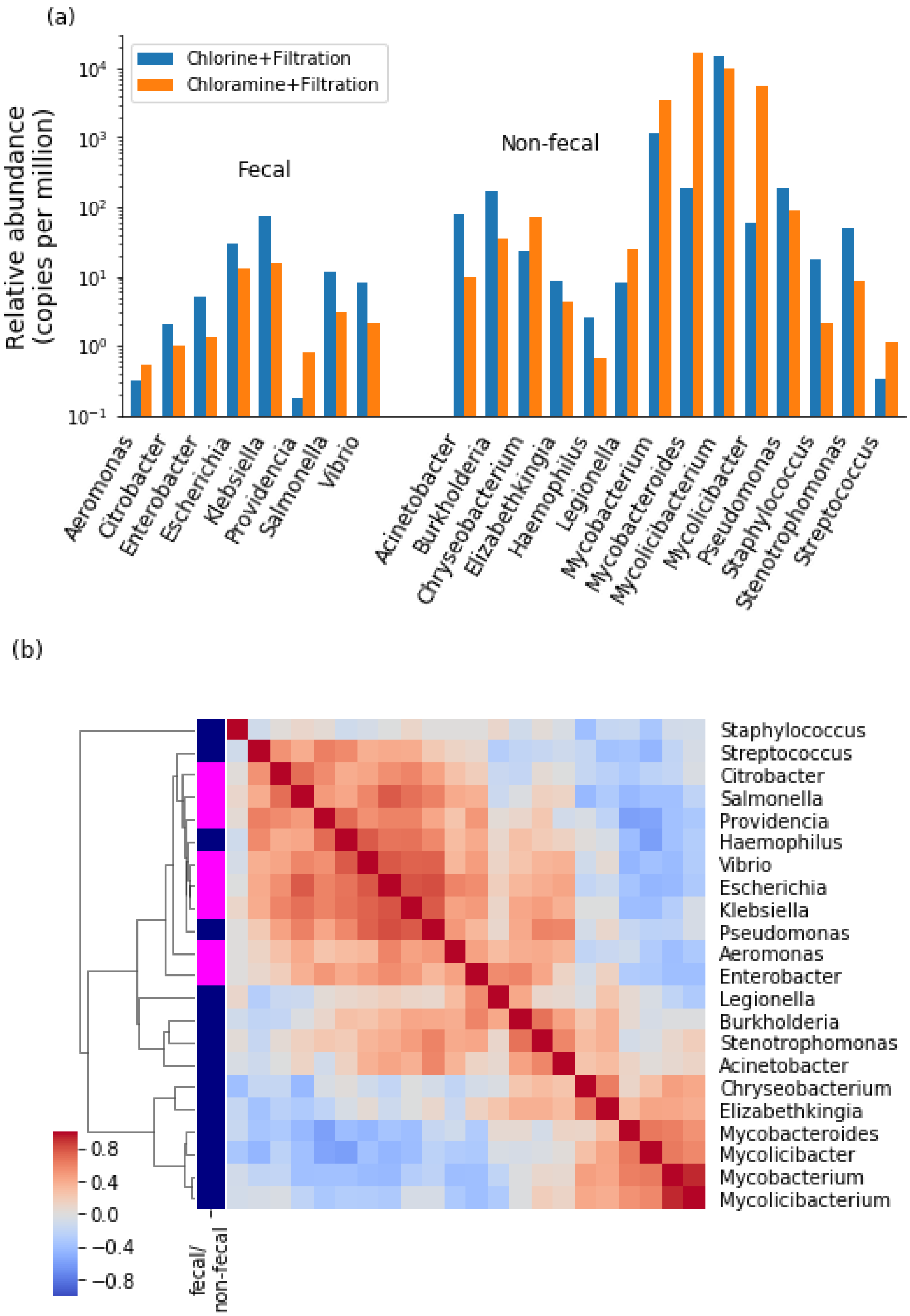


Figure 5-1. Marker Genes for Pathogens Detected via NGS.

(a) Relative abundances (copies per million) of marker genes of pathogens obtained using the in-house metagenome-derived pathogen quantification pipeline leveraging UniRef90 gene-family annotation by HUMAnN3. Two individual samples from the RWDSs with BAC-filtration+secondary chlorine residual treatment and BAC-filtration+secondary chloramine residual treatment are shown. (b) Hierarchical clustering of relative abundances (copies per million) of pathogen marker genes across all 56 samples collected at 30°C. Pathogens were quantifies using the same pipeline. Pink – fecal pathogens. Dark blue – non-fecal pathogens.

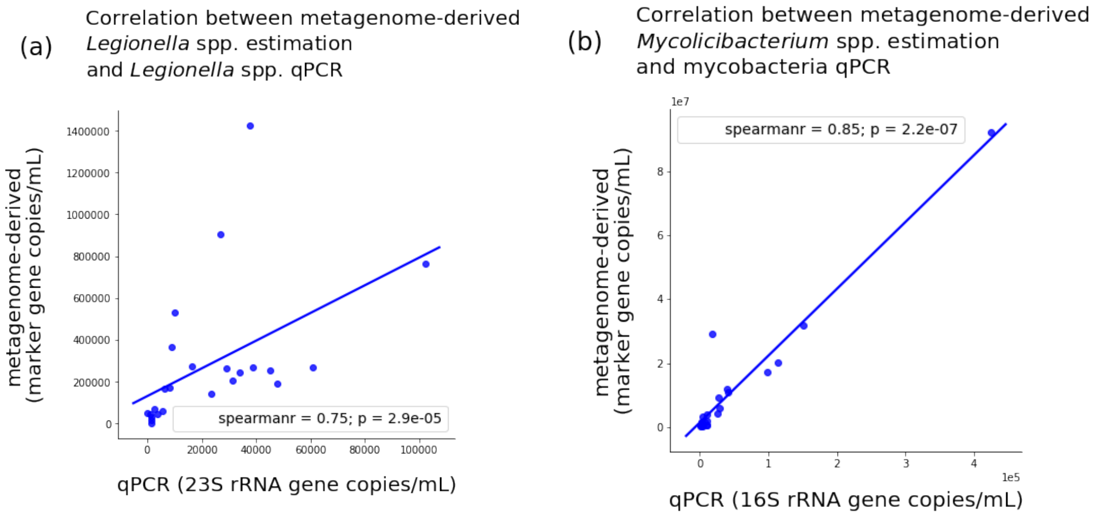


Figure 5-2. Correlation Between Metagenome-derived Quantification of (a) *Legionella* spp. and (b) *Mycolicibacterium* spp. with qPCR of *Legionella* spp. and Mycobacteria, respectively.

Metagenome-derived quantification is based on the in-house pathogen quantification pipeline leveraging UniRef gene-family annotation by HUMAnN. Note that mycobacteria qPCR covered four mycobacteria genera, *Mycobacterium*, *Mycolicibacterium*, *Mycolicibacter* and *Mycobacteroides* spp. with *Mycolicibacterium* spp. dominating the RWDSs.

5.2.4 Conclusion

This case study demonstrates the potential of metagenomic sequencing in detecting and quantifying relative and absolute abundances of pathogenic genera, including fecal and non-fecal microorganisms, from water and biofilm samples from simulated RWDSs. While not validated, the *pathogen quantification pipeline* described here has the potential to estimate pathogen abundances at the species level.

5.3 Case Study #2: Profiling Antimicrobial Resistance in Surface Water

The full manuscript summarized in this case study has been published in Davis et al. (2020a).

Monitoring surface water for the dissemination of antibiotic resistance is essential for curbing the spread of resistance globally. Monitoring single bioindicators as a proxy for the total resistome, though, cannot resolve geographical differences in resistomes nor fully capture all clinically-relevant ARGs. This case study demonstrates that NGS approaches targeting metagenomes paired with publicly-available analysis pipelines such as MetaStorm and NanoARG, can be used to comprehensively track changes in relative abundance of all ARGs in anthropogenically impacted watersheds. This case study further demonstrates the ability of metagenomic assembly to illuminate the genetic context of clinically-relevant ARGs to aid in prioritizing mitigation efforts. The methods described here can be applied broadly to be included as in-depth complementary analyses to surface water monitoring programs globally.

5.3.1 Introduction

Environmental surveillance has been recognized by the World Health Organization as an essential means to understand the development, spread, and circulation of antibiotic resistance between humans, animals, food, and water networks (World Health Organization (WHO), 2015). Surface waters are of interest because they act as both reservoirs, recipients, and pathways for the dissemination of ARGs and antibiotic resistant bacteria (ARB) from human and animal pollution (Amos et al., 2014; Marti et al., 2014). A particular focus has been paid to WWTPs as their variably treated effluents directly contribute to antibiotic resistance in receiving rivers (Bréchet et al., 2014; Cacace et al., 2019; Pruden et al., 2012). As a result, comprehensive watershed monitoring programs are currently being explored and implemented by the U.S. Environmental Protection Agency (Garland et al., 2019) and the Centers for Disease Control (Kirby, 2020) for tracking antibiotic resistance in aquatic environments. These monitoring programs, though, have been designed to implement the detection of indicator ARGs (*sul*1, *int*I1, bla*KPC, tetA)* using qPCR and culture targets (extended-spectrum beta-lactamase (ESBL) *E. coli*) to act as proxies for total resistance levels in anthropogenically impacted environments. However, this approach has significant drawbacks as the abundance and diversity of environmental resistomes (i.e., the total ARGs in an environmental metagenome) is geographically dependent (Hendriksen et al., 2019) and may contain clinically relevant ARGs that will go undetected relying on single target bioindicators.

Targeting the entire metagenome using NGS approaches allows for the simultaneous detection of all ARGs in a resistome as well as the genetic context in which they are present (i.e., their association with MGEs and human pathogens) by sequencing the genetic material from entire environmental samples. This contextualization can allow for the identification of high-priority resistance determinants for prioritization of mitigation efforts (e.g., more advanced wastewater treatment). This comprehensive and high-throughput approach to environmental resistome monitoring also circumvents primer biases introduced by qPCR, as many ARGs are continuously mutating, obsoleting published primers (Crossette et al., 2021). The goal of this case study is to demonstrate the feasibility of using metagenomics to comprehensively estimate the abundance and diversity of surface water resistomes as well as provide genetic context to clinically relevant ARGs.

5.3.2 Methods

5.3.2.1 Sample Collection and Processing

Bulk water samples were taken from three different watersheds with varying degrees of human development in Puerto Rico 6 months after Hurricane Maria (Davis et al., 2020a). The sampling strategy was designed to highlight the gradient of anthropogenic stress being enacted on the rivers as they flowed through residential areas, and eventually were mixed with treated wastewater. Samples were taken from far upstream pristine samples that had little to no human impact, residential samples where urban development was dense, directly downstream of discharged wastewater, and then the WWTP influent and effluents themselves. Water samples were filter concentrated on 0.22 µm filters until clogging and the cells remaining on the filters were DNA extracted using the FastDNA Spin kit for Soil (Table 5-2).

5.3.2.2 Metagenomic Sequencing and Bioinformatics

DNA extracts were sent to Diversigen, Inc. (Houston, TX) where samples were sequenced on an Illumina NovaSeq 6000 with a 2 x 150 bp strategy (Table 1). For short-read quality trimming, read alignment, and assembly, raw sequencing data were uploaded to the publicly available metagenomic analysis platform, MetaStorm (Arango-Argoty et al., 2016). MetaStorm uses a suite of commonly used bioinformatic software such as Trimmomatic for quality filtering and read trimming (Bolger et al., 2014), Vsearch for paired-end read merging (Rognes et al., 2016), DIAMOND for read alignment to protein sequence databases (Buchfink et al., 2014), Bowtie2 for alignment to 16S rRNA databases (Langmead Ben and Steven, 2013), and IDBA-UD for *de-novo* assembly (Peng et al., 2012). The quality filtered short-reads were queried against the Comprehensive Antibiotic Resistance Database (CARD; v2.0.1) using DIAMOND and normalized to the number of 16S rRNA reads per sample (Li et al., 2015b). Assembled metagenomic contigs resulting from IDBA-UD were downloaded from MetaStorm and uploaded to NanoARG (Arango-Argoty et al., 2019), a publicly available platform for long-read and contig annotation. Contigs were annotated against CARD, a MGE database, and assigned taxonomy with Centrifuge (Kim et al., 2016) for ARG contextualization.

Table 5-2. NGS Methodologies Applied in Case Study #2.

Component	Approach/Description	Reference
Source	Surface Water	
Sample Concentration	Membrane Filtration (0.22 µm mixed-cellulose ester filters)	(APHA, 2017)
Nucleic Acid Extraction	FastDNA Spin Kit for Soil (MP Bio)	(Li et al., 2018)
Sequencing Approach	Metagenomic Sequencing	(Garner et al., 2016; Lee et al., 2020)
Library Preparation	NexteraXT DNA Library Prep kit	(Bowers et al., 2015)
Sequencing Platform/Configuration	NovaSeq 6000; 2x150 bp; ~10 Gb/sample	
Sequencing Coverage [mean±standard deviation (range)]	40,722,025±6,521,266 (26,358,963-2,434,282)	
Data Analysis	MetaStorm	(Arango-Argoty et al., 2016)
Annotation Database(s)	CARD, MetaStorm MGEs	(Alcock et al., 2020; Arango-Argoty et al., 2019)

5.3.3 Results

A total of 816 unique ARGs were detected across all samples that encompassed 18 different classes of antibiotics. The most abundant classes were the multidrug, aminoglycosides, beta-lactams, and macrolide-lincosamide-streptogramin (MLS) genes (Figure 5-3). There was a significant difference in total relative abundance depending on the location within each watershed (ANOVA, $p = 6.4e-5$), and a significant increase in abundance between 'Residential' samples and samples directly 'Downstream' of WWTP discharge sites (TukeyHSD, $p<0.05$). This indicates that WWTPs in these catchments are directly responsible for the enrichment of the resistome with emphasis of the aminoglycoside, beta-lactam, and MLS antibiotic classes, many of which are clinically-relevant.

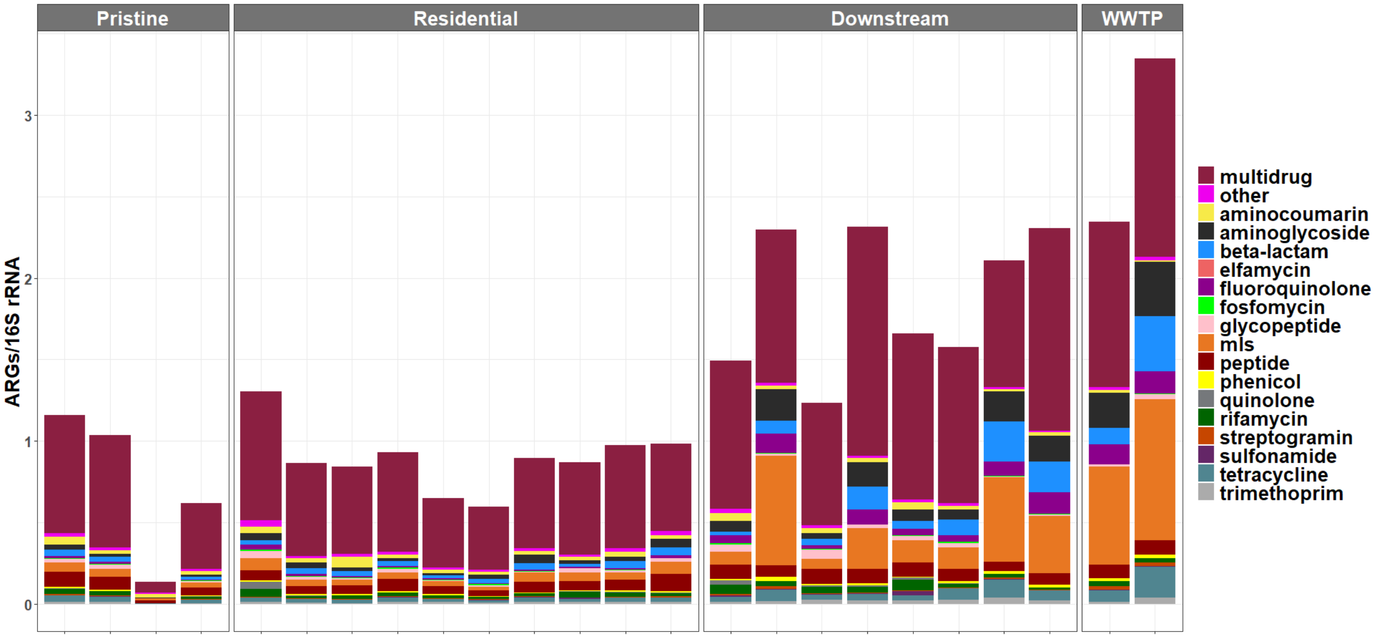


Figure 5-3. Relative Abundance of ARGs across Locations within Each River System
ARGs are Broken Down into their Respective Antibiotic Classes. MLS= Macrolide-Lincosamide-Streptogramin.
Source: Reprinted with permission from Davis, B.C., Riquelme, M.V., Ramirez-toro, G., Bandaragoda, C., Garner, E., Rhoads, W.J., Vikesland, P., Pruden, A., 2020a. Demonstrating an Integrated Antibiotic Resistance Gene Surveillance Approach in Puerto Rican Watersheds Post-Hurricane Maria. Environ. Sci. Technol. https://doi.org/10.1021/acs.est.0c05567. Copyright 2020 American Chemical Society.

Analyzing the assembled contigs that contained an ARG (256,614 contigs) revealed that only a small minority were co-occurring with MGEs but were responsive to anthropogenic stress or pollution. The average mobility incidence was calculated as the number of co-occurrences of ARGs with MGEs on an assembled contig as a percentage of all occurrences of that ARG in the contig library (Ju et al., 2019) (Figure 5-4). The results indicate that mobile beta-lactam and aminoglycoside ARGs were being introduced into rivers not only from WWTPs, but also from surrounding residential areas. These beta-lactam ARGs were of particular concern as they belonged to the ESBL and carbapenamase groups and were taxonomically associated with *Enterobacteriaceae*, a critical pairing according to the WHO's priority list of resistant pathogens. For example, the Klebsiella-pneumoniae-carbapenamase (KPC-2) co-occurs with an ISKpn6-like transposase on a contigs assigned to *Klebsiella pneumoniae* that has been traversing North and South American clinics (Belder et al., 2017) being emitted from WWTPs across Puerto Rico, albeit these represented a small minority of taxonomically assigned ARG contigs (6 out of 62,545 contigs).

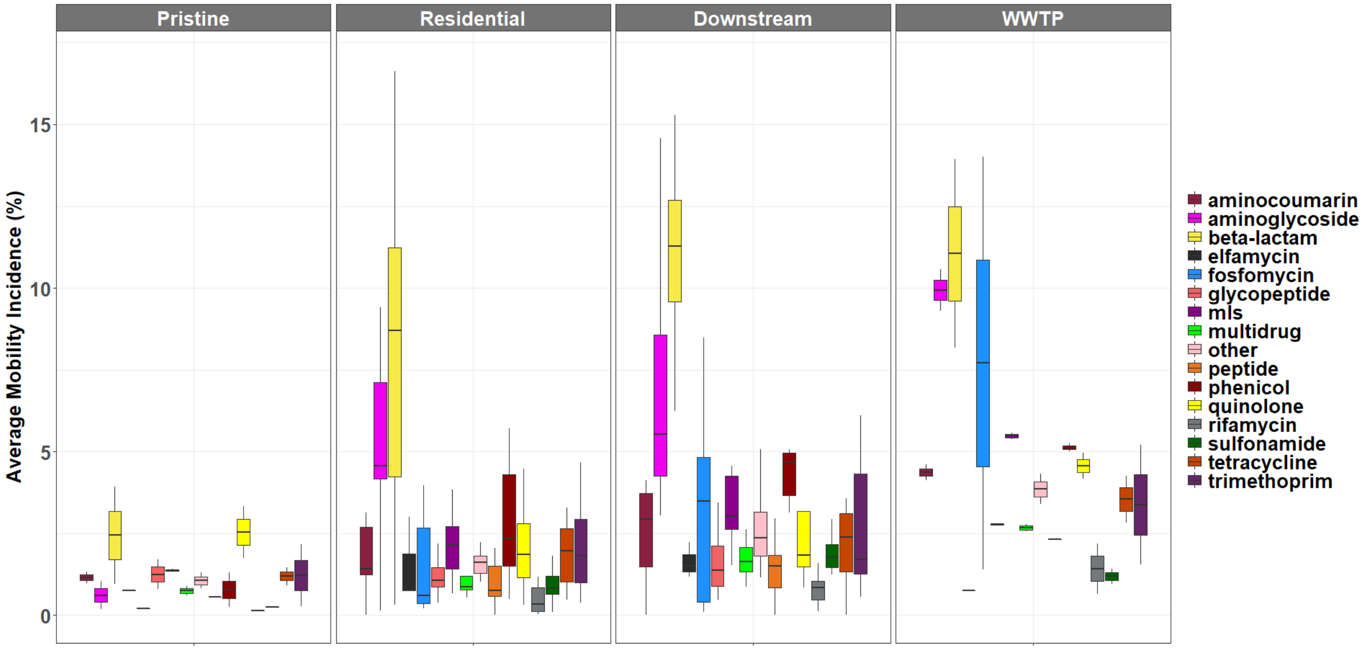


Figure 5-4. Average Mobility Incidence of ARGs Across Locations within Each River System.
Source: Reprinted with permission from Davis, B.C., Riquelme, M.V., Ramirez-toro, G., Bandaragoda, C., Garner, E., Rhoads, W.J., Vikesland, P., Pruden, A., 2020a. Demonstrating an Integrated Antibiotic Resistance Gene Surveillance Approach in Puerto Rican Watersheds Post-Hurricane Maria. Environ. Sci. Technol. https://doi.org/10.1021/acs.est.0c05567. Copyright 2020 American Chemical Society.

5.3.4 Conclusion

This case study demonstrates the ability of metagenomic sequencing and analysis to detect and quantify a large number of ARGs simultaneously to illuminate the impacts that anthropogenic stressors, including WWTP discharge, have on surface water environments. This case study also demonstrates the ability of metagenomic assembly to aid in providing useful contextualization of these ARGs that have direct consequences in clinical medicine. The analysis of contig libraries highlights the novel utility of metagenomic assembly techniques in illuminating otherwise undetectable mobile ARGs here that were unique to the region and responsive to point source and non-point source pollution.

5.4 Case Study #3: Pathogen Screening using 16S rRNA Gene Amplicon Sequencing in Environments Impacted by Extreme Flooding Events

The full manuscript summarized in this case study has been published in (Keenum et al., 2021).

In order to demonstrate the efficacy of utilizing NGS to identify pathogen targets, 16S rRNA sequencing was applied as well as qPCR, PCR, culture, and microscopy methods to six water systems in Puerto Rico. This diverse array of methods enabled us to screen for potential presence of pathogens, quantifying gene markers for *Legionella pneumophila, Mycobacterium avium, Cryptosporidium parvum, Giardia lamblia, Leptospira* spp., *Escherichia coli*, *Salmonella* spp., *Enterococcus*, and Shiga toxin-producing *E. coli*. Detection limits were compared across the different methods. Additionally, 16S rRNA amplicon sequencing and subsequently QIIME2, an open access pipeline for taxonomic profiling of microbial communities, were applied to examine water systems for the entire World Health Organization global list of priority pathogens as well as the overall taxonomy of the systems. Using both NGS and qPCR enabled us to estimate detection limits and the sensitivity of amplicon sequencing as a pathogen screening method.

5.4.1 Introduction

As the frequency and intensity of hurricanes and other storms increase (Balaguru et al., 2016; Lugo, 2000), Puerto Rico and other communities around the world face pressure to efficiently and effectively respond to natural disasters. Natural disasters such as hurricanes and extreme flooding events can cause widespread contamination of drinking water systems (George et al., 2019; Ratnapradipa et al., 2018; Smith et al., 2018). Understanding which pathogens drinking water systems are susceptible to, as well as identifying system characteristics that make them more vulnerable to service disruption and contamination, is critical to prepare for future events. The U.S. CDC has highlighted specific fecal pathogens of concern following hurricanes, such as *Cryptosporidium parvum*, *Giardia lamblia,* and Shiga toxin-producing *Escherichia coli* (CDC, 2019), while *Salmonella typhimurium* has also been identified as a concern for flooded groundwater systems (Yard et al., 2014). The recommended methods of detection for these organisms is culture and microscopy, but these methods are extremely time consuming and can only search for the target organism. The CDC has further identified the OP, *Legionella,* and the zoonotic pathogen, *Leptospira*, to also be of concern following storms (CDC, 2019). Current guidance tends to be focused on fecal pathogens and the extent to which storms should elevate concerns about opportunistic or zoonotic pathogens is not clear (Cassell et al., 2018; Garcia-Vidal et al., 2013; Hicks et al., 2007).

Utilizing a non target approach for detecting pathogens in drinking waters after a hurricane, the objective of this study was to investigate the efficacy of using 16S rRNA amplicon sequencing, to profile bacterial community composition and identify the potential presence of pathogens in hurricane impacted drinking water systems. In order to validate this method of pathogen detection, culture, PCR and microscopy were also applied. The findings aid in identifying vulnerabilities of drinking water systems to pathogen intrusion and proliferation in the face of major disruptive events and also inform improved strategies for preparedness, assessment, response, and recovery for future storms.

5.4.2 Methods

5.4.2.1 Sample Collection for Molecular Analysis

Six small volunteer- operated and one large municipally operated system were sampled in March 2018. Systems were selected to represent a cross section of source waters (labeled S: surface water, or G: groundwater), treatment approaches (C: chlorination, B: both filtration and chlorination, P: PRASA system with coagulation, flocculation, sedimentation, and chlorination), and distinct recovery experiences after the hurricane. Samples collected for molecular analyses, which are less time sensitive and higher throughput than culture and microscopy-based methods, enabled a more thorough source-to-tap investigation. “Untreated” water was collected at the source or after flushing outlets to steady temperature before treatment. First flush "Distribution System" samples were collected from outdoor taps at 1-10 resident homes within each system to represent water delivered to homes. "Post-Chlorination" treated water samples were collected at flushed outlets immediately after water treatment. All water samples were collected in 2-L sterile polypropylene bottles. After mixing and removing an aliquot for water quality analysis, 48 mg of filter-sterilized aqueous sodium thiosulfate solution was added to quench disinfectant residuals. For comparison to microscopic detection of *Cryptosporidium* (described above), a duplicate 100-L ultrafilter sample was collected in parallel. Backwashed eluent was filter-concentrated onto a 1.2-µm filter pore-size mixed-cellulose ester filters (Millipore).

Samples were transported on ice to the lab, where they were filtered onto 0.22-µm pore-size mixed-cellulose ester filters (Millipore, Billerica, MA) until clogging (250 mL – 2 L). Biofilm samples were collected from the inner wall of pipes and/or outlets using sterile cotton swabs and a uniform swabbing pattern at the same locations as the water samples for molecular analysis. Applicator tips were detached and placed directly into DNA extraction lysing matrix tubes and transported back to the lab on ice. Field and trip blank samples consisted of autoclaved water generated in the lab and transported in the field, with the field blank being open during sampling and the trip blank remaining closed. Blank water samples were processed in parallel with the field samples in the lab.

Filters and swabs were stored at -20 °C in 50% high-purity ethanol (Fisher Scientific), 50% ultrapure water (Fisher Scientific), prior to being transported to Virginia Tech (~18 hours) on ice and stored at -20 °C until further processing. DNA extraction for bacterial analyses was carried out on fragmented filters and swabs using the Fast DNA SPIN Kit for Soil (MP Biomedicals, Solon, OH) according to the manufacturer protocol. DNA extraction for ultrafilter samples was carried out using the same commercial kit after lysing cells with five sequential cycles of submersion for 1 minute in liquid nitrogen followed by 1 min in boiling water to maximize recovery of parasite DNA (Guy et al., 2003; Higgins et al., 2001). A DNA extraction blank was carried out on an empty tube for each batch of samples. A summary of NGS methodologies applied in this case study are outlined in Table 5-3 and discussed below.

5.4.2.2 Sample Collection and Analysis for Culture, Filtration and Microscopy

Samples of raw water before treatment ("untreated") and at one outdoor hose-bib sample (flushed "distribution system") were collected from each small system for culture analysis of total coliforms, fecal coliforms and *E.coli* utilizing Standard Method 9222, media and incubation procedures (Rice et al., 2012). One-liter grab samples were collected into sterile bottles pre-dosed with 0.08 mL of a 3% filter-sterilized sodium -thiosulfate solution, after flushing as indicated by steady temperatures (<0.1 °C change/60 seconds). Per this method, if samples did not contain any culturable TC, they were not measured for FC or *E. coli.* Enumeration to log density was determined by presence/absence in 10-fold serial dilutions. Ten additional one-liter samples were collected and analyzed for *Salmonella* spp. by an adaptation of Standard Methods described in Herson et al. (2005) To combine the samples, all 10 bottles were filter-concentrated onto a 0.45-µm pore size Gelman filter and the filter was placed in 100 mL of TET media for enrichment prior to quantification (Minnigh H.A., 2006).

Microscopic analysis for detection of *Cryptosporidium* oocysts and *Giardia* cysts requires ultrafiltration of large volumes of water and thus was applied only to water immediately before treatment and one distribution system point for each site. Approximately 250-L of flushed water was filtered onto a 1-µm ultrafilter (Pall Envirocheck, Ann Arbor, MI). Organisms from the ultrafilter were resuspended in 500 mL of tween80 (Fisher Scientific, Waltham, MA) by backwashing the filter. *Cryptosporidium* and *Giardia* were detected in the backwash eluent after concentration through centrifugation, immunomagnetic separation, and immunofluorescence assay microscopy using EPA method 1623 (U.S. EPA, 2005).

5.4.2.3 16S rRNA Gene Amplicon Sequencing

All DNA extracts were amplified via PCR targeting the V4 and V5 regions of the 16S rRNA gene following the online Earth Microbiome Project protocol using barcoded primers (515F/926R) (Masella et al., 2012). Triplicate PCR products for each sample were composited and purified using a QIAquick PCR Purification Kit (Qiagen, Valencia, CA). Sequencing was performed by the Biocomplexity Institute of Virginia Genomic Sequencing Center (Blacksburg, VA) on an Illumina MiSeq with V3 2x300 paired end cycles. Reads were analyzed using the QIIME2 pipeline (Bolyen et al., 2018). All singleton reads and chimeric sequences were removed using DADA2 and OTU tables were generated at 97% similarity for analysis using VSEARCH (Callahan et al., 2016; Rognes et al., 2016). Taxonomy was classified using the Greengenes (May 2013 release) database (DeSantis et al., 2006). Samples were rarefied to 4,049 reads to minimize the impact of uneven sequencing depth for the purpose of statistical comparison. Field, filtration, DNA extraction blanks, and at least one PCR blank per lane were included in the analysis. Jackknifed beta diversity analysis was performed to calculate unweighted UniFrac distance matrices for the comparison of sample taxonomic similarity (Lozupone et al., 2011). Raw reads have been submitted to NCBIs BioProject PRJNA564042.

5.4.2.4 PCR and qPCR

All water and biofilm samples were analyzed by qPCR for *Legionella* spp. (23S rRNA), *L. pneumophila* (*mip*), *Mycobacterium* spp. (16S rRNA), *M. avium* (16S rRNA), and total bacteria (16S rRNA) gene copies (gc) using previously-reported methods (Juretschko et al., 1998; Lane et al., 1985; Nazarian et al., 2008; Radomski et al., 2010; Wilton and Cousins, 1992). Gene copies were quantified on a CFX96 Real Time System (BioRad, Hercules, CA) from DNA extracts in triplicate reactions with 10-fold serial dilutions of synthetic DNA standards (G-blocks, IDT, Coralville, IA) and non-template controls on each run. The quantification limit for each assay ranged from 10 gc/reaction (16S rRNA, *L. pneumophila*) to 100 gc/reaction (*M. avium, Mycobacterium* spp.*, Legionella* spp.). All qPCR data are reported as $\log_{10}$. Samples were also analyzed by PCR (detect/non-detect) for pathogenic *Leptospira* (*hap1*), *S. typhimurium* (*invA*)*,* Shiga toxins 1 and 2 (stx_1 and stx_2)*, C. parvum* (18S rRNA), and *G. lamblia* (heat shock protein 70) (Abbaszadegan et al., 1993; Johnson et al., 1995). Field, filtration, DNA extraction, and a reaction blank were included in all PCR/qPCR runs. The optimal dilution for minimizing PCR inhibition was identified by performing a dilution series with positive control spikes on every sample. Ideal dilutions ranged from 1:1 – 1:20 for water or biofilm and 1:50 for ultrafiltered samples.

Table 5-3. NGS Methodologies Applied in Case Study #3.

Procedure Component	Approach/Description	Reference
Source	Drinking Water	
Sample Concentration	Membrane Filtration onto 0.22-µm pore-size mixed-cellulose ester filters	
Nucleic Acid Extraction	Fast DNA Spin Kit for Soil	
Sequencing Approach	Amplicon sequencing targeting 16S rRNA genes	
Library Preparation	PCR targeting the V4 and V5 regions of the 16S rRNA gene following the online Earth Microbiome Project protocol using barcoded primers (515F/926R)	(Gilbert et al., 2014; Masella et al., 2012)
Sequencing Platform/Configuration	Illumina MiSeq with V3 2x300 paired end cycles	

Data Analysis	Reads were analyzed using the QIIME2 pipeline. All singleton reads and chimeric sequences were removed using DADA2 and OTU tables were generated at 97% similarity for analysis using VSEARCH.	Qiime2: (Bolyen et al., 2018)
		Dada2:(Callahan et al., 2016)
		Vsearch:(Rognes et al., 2016)

Sequencing Coverage [mean±standard deviation (range)]	95,124±89,277 (3,900-60,3137)	
Annotation Database(s)	Greengenes (May 2013 release)	(DeSantis et al., 2006).

5.4.3 Results

16S rRNA gene amplicon sequencing was highly consistent with qPCR-based methods, proving to be a useful non-target broad level screen for potential pathogens, such as *Legionella* spp., *Mycobacterium* spp., and *Leptospira* spp., yielding results congruent with the more sensitive and pathogen-specific qPCR assays (Figure 5-5). Accordingly, amplicon sequencing identified other potential pathogens of concern that were not directly targeted by the other methods, including *Acinetobacter* spp., *Burkholderia* spp. *Pseudomonas* spp., *Streptococcus* spp*.,* *Staphylococcus* spp., and *Ralstonia* spp. Detection of *Burkholderia* spp. is noteworthy, as this genus is known to contain OPs that are problematic in warmer climates (Inglis et al., 2000; Mayo et al., 2011). However, it is critical to recognize that this method does not have the resolution to confirm these detections truly correspond to pathogens. Taxonomic resolution of amplicon sequencing is at best at the family-, and sometimes genus-, level (Johnson et al., 2019). In such cases that a potential pathogen is identified by amplicon sequencing, a more specific, targeted method would be needed to confirm. In a disaster-response scenario, it would be critical to follow up on such preliminary detections with culture-based testing. While amplicon sequencing could prove useful for screening opportunistic and zoonotic pathogens, it fell short in terms of predicting fecal-associated pathogens.

Detection of the family *Enterobacteriaceae* was consistent with detection of *Salmonella* and *E. coli* by culture-based methods, but resolution was not possible at the genus-level. Given that fecal contamination is a vital concern following major storms, a recommended first step is intensifying traditional TC and FC monitoring, for which methods are widely-available, standardized, and relatively cost-effective.

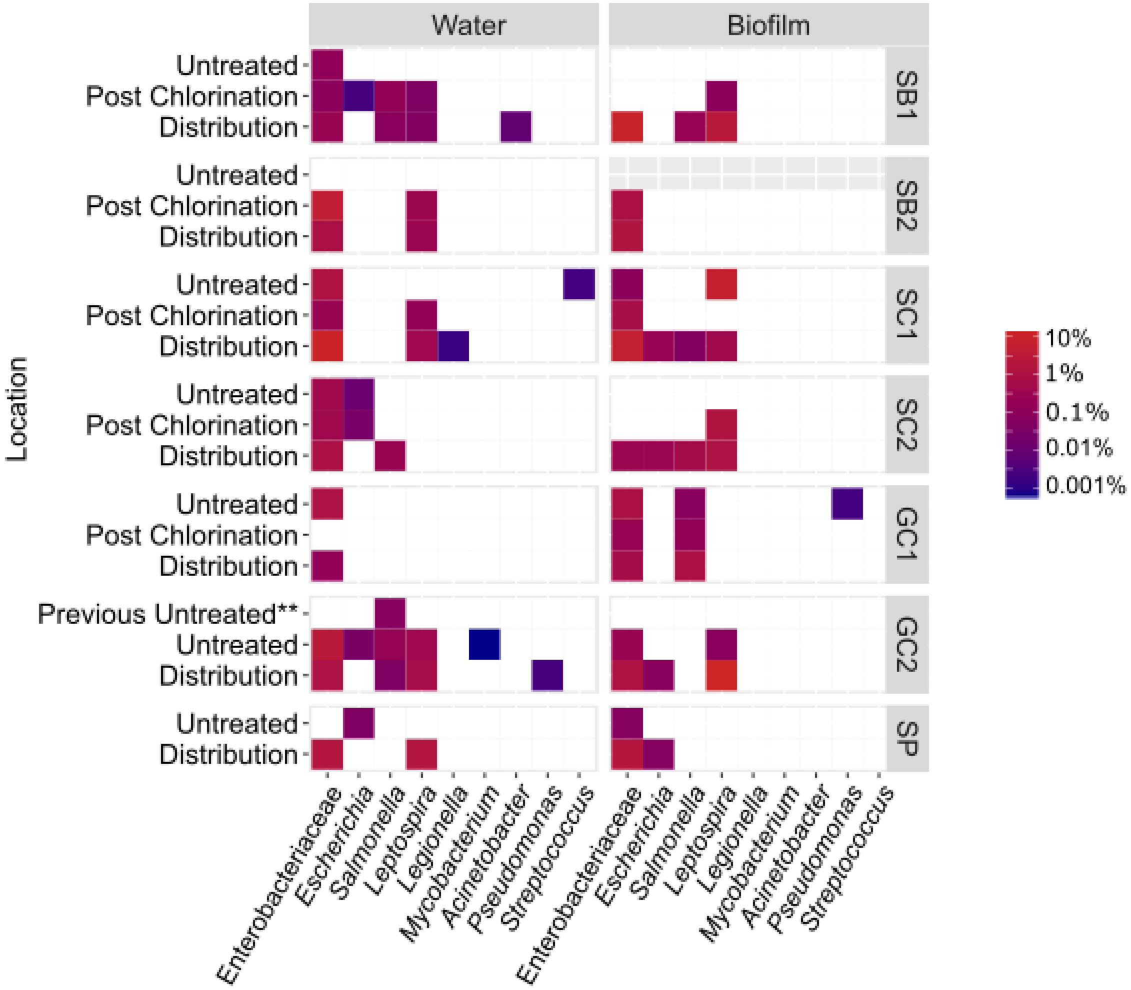


Figure 5-5. Screening 16S rRNA Gene Amplicon Libraries for Detection of Taxonomic Groups Known to Contain Pathogens.

Heat map compares the relative abundance (% total OTUs) at the genus or family level. For this analysis, all amplicon sequencing libraries were pooled among samples representing each location (untreated, post-chlorination, distribution system). Untreated biofilm samples were not be obtained for System SB2 (gray cells). Blank (white) cells indicate non-detection. Based on the rarefied library size of 4,049 reads, the estimated detection limit is 0.025% of the microbial community. **Previous untreated in System GC2 refers to the surface water source that was used immediately after Hurricane Maria before the generator was installed.

Source: Reprinted with permission from Keenum, I., Medina, M.C., Garner, E., Pieper, K.J., Blair, M.F., Milligan, E., Pruden, A., Ramirez-Toro, G., Rhoads, W.J., 2021. Source-to-Tap Assessment of Microbiological Water Quality in Small Rural Drinking Water Systems in Puerto Rico Six Months After Hurricane Maria. Environ. Sci. Technol. acs.est.0c08814. https://doi.org/10.1021/acs.est.0c08814. Copyright 2021 American Chemical Society.

5.4.3.1 A Hierarchy of Complementary Methods for Comprehensive Pathogen Surveillance

Given that the aim of this study was to broadly screen for as many storm-relevant pathogens as possible, a purely culture-based approach would have severely limited the scope of this study. Species-level resolution is adequate to confirm some pathogens (Figure 5-6), such as was the case for the PCR/qPCR assays employed here targeting *L. pneumophila*, *M. avium, G. lamblia, Salmonella* spp., and *C. parvum*, while strain-level resolution is required for others, such as *E. coli.* For this reason, virulence-specific genes of *L. pneumophila* (*mip*)*, E. coli* (*stx*$_1$, *stx*$_2$), *S.*

typhimurium (*invA*) and *Leptospira* (*hap1*) were targeted via qPCR to increase confidence in detection of virulent strains. Overall, there was a higher frequency of target pathogen detection by PCR/qPCR than via amplicon sequencing, which is consistent with the very low detection limit of qPCR.

The comprehensive culture, qPCR, microscopy, and amplicon sequencing approach applied here is not likely practical for rapid testing of the water following a natural disaster. However, the results of this study provide insight into how these tools may be applied in a strategic fashion to first screen, and then zoom in on potential pathogens of concern.

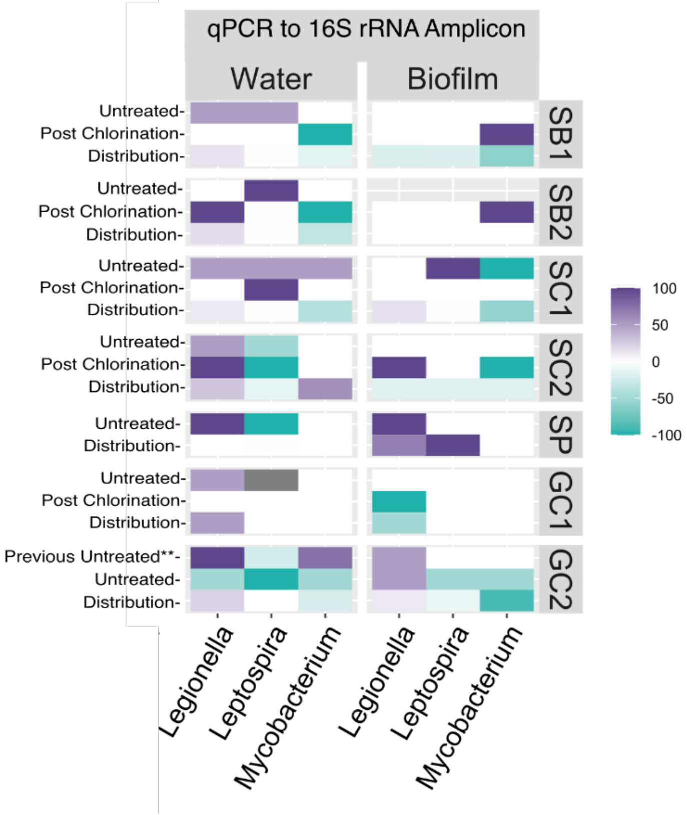


Figure 5-6. 16S rRNA Amplicon Sequencing Results in Rural Water Systems.
Congruence between 16S rRNA amplicon sequencing and qPCR results where 100 indicates full agreement between the two measurements and -100 indicates a discrepancy in results.

5.4.4 Conclusion

Utilizing a NGS approach in addition to culture and qPCR enabled the broad detection of World Health Organization pathogens associated with water as well as the detection of specific pathogens of interest. While 16S rRNA amplicon sequencing does not provide species level identification of pathogens, in many cases this is not needed as entire genera or families contain widespread pathogens. This approach, in conjunction with qPCR and culture methods can ensure that all pathogens in a system have been investigated and should be applied periodically in disturbed systems to assess if the pathogens and taxonomy present are shifting as drinking water systems continue to experience disruptions. The novelty of this approach is that it would enable a high-level screening of pathogens in a drinking water system, thus enabling further assessment into pathogens of interest.

5.5 Case Study #4: Aerosolization of Viruses and Associated Risks at Wastewater Treatment Plants

Metagenomic sequencing of both DNA and RNA viruses will identify all viruses are present in wastewater and all viruses that become aerosolized during treatment. A yearlong sampling campaign will provide insight into seasonal variation in the generation of viral aerosols. Construction of a quantitative microbial risk assessment (QMRA) model based on NGS results will allow operators to input parameters pertaining to the WWTP into the model to assess exposure to viral aerosols on an individual basis.

5.5.1 Introduction

Wastewater treatment is a fundamental aspect of public health protection; WWTPs are designed to reduce the number of pathogens, including viruses, that are released via wastewater effluent into the receiving environment. Untreated wastewater also represents one of the most diverse viral metagenomes that has been studied (Cantalupo et al., 2011). Viral nucleic acids have been detected at multiple points throughout the treatment process, including the effluent (Tamaki et al., 2012). Wastewater produces bioaerosols (microorganisms attached to water or dust that disseminate in air), which can affect the health of WWTP employees (Divizia et al., 2008; Korzeniewska, 2011; Lee et al., 2016). Viruses detected in wastewater include those that are known to infect humans, including human adenovirus, Norwalk virus, human papillomavirus and human polyomavirus (Cantalupo et al., 2011). While it is likely that viral particles pose a public health risk, there are no regulations or protective measures in place.

The purpose of this study is to demonstrate the feasibility of using NGS to identify all viruses that become aerosolized during the wastewater treatment process and potentially pose a risk to those likely to be exposed (i.e., WWTP operators or downwind populations). This study will also demonstrate the use of NGS results in construction of a QMRA model; this model will provide guidelines to allow informed personal protective equipment decisions on an individual basis based on characteristics of a WWTP.

5.5.2 Methods

Simultaneous 24-hour composite samples of wastewater and air will be collected from two locations within the WWTP most likely to generate aerosols (Korzeniewska, 2011), near the influent and aeration basins. Wastewater samples (1 L) will be collected using ISCO 6712 Portable Sampler and air samples will be collected using InnovaPrep ACD-200 Bobcat sampler at 200 L/min (1 min on, 1 min off) for a total of ~288 m^3 air. Samples are to be collected every two weeks for the duration of the year to address seasonal differences in water and air viral communities. Meteorological conditions will also be measured, including wind speed, wind direction, temperature, relative humidity and solar radiation.

This study will investigate both DNA and RNA viral communities via NGS. Based on sequencing results, a QMRA model will be constructed for viruses of interest using number of reads. Construction of the QMRA requires epidemiological data (can be found on QMRAwiki.org), dose-response parameters (also found on QMRAwiki), and exposure parameters (to be determined from this study) to determine the associated risks with viral bioaerosols near the WWTP. A summary of NGS methodologies applied in this case study are outlined in Table 5-4.

Table 5-4. NGS Methodologies Applied in Case Study #4.

Procedure Component	Wastewater Approach	Air Approach	Reference
Source	Wastewater (influent and aeration basin)	Air near influent and aeration basin	
Sample Concentration	Add $MgCl_2$ Adjust pH to 3-4 0.45-µm filtration	Sample onto dry 52 mm electret filter Elute into PBS	(Ahmed et al., 2020) (Raynor et al., 2021)
Nucleic Acid Extraction	Extract viral DNA and RNA using Qiagen UltraSens Viral Kit		
Sequencing Approach	Metagenomic sequencing		
Library Preparation	ScriptSeq (RNA) and NEB DNA Prep		
Sequencing Platform/Configuration	Illumina NextSeq 2x150 bp		
Sequencing Coverage	10M reads/sample		
Data Analysis Annotation Database(s)	Homemade viruses dataset from University of Notre Dame (currently being validated)		

5.5.3 Results

The findings of this case study will indicate all viruses found in wastewater and all viruses that become aerosolized and pose a potential public health risk. NGS results will specify the host distribution of viruses (i.e., bacteria, plant, animal or human); this will be used to determine which viruses infect humans and may pose a direct hazard to human health. Because NGS results indicate relative abundances of viruses, the QMRA model will be constructed based on the number of reads for a particular virus as a means of determining absolute risk. Overall, the QMRA model based on NGS results will provide a better understanding of which viruses become aerosolized during wastewater treatment and their associated risks, allowing the WWTP operators to make more informed decisions on personal protective measures. For example, WWTP operators will be able to assess exposure risks to a particular virus with and without wearing a mask.

5.5.4 Conclusion

Using NGS-based approaches in this case study will provide insight into viruses present during wastewater treatment and which viruses may be preferentially aerosolized due to treatment processes. NGS will provide a more complete picture of viral nucleic acids that are present in wastewater and become aerosolized as compared to more targeted microbiological methods (e.g., qPCR). NGS used in conjunction with QMRA modelling will provide operators at WWTPs the capability to determine exposure risks to aerosolized viruses, allowing operators to make informed decisions about the use of personal protective equipment.

5.6 Case Study #5: Functional Application of Metagenomics in Wastewater

A functional annotation pipeline, developed from the UniProt Knowledgebase Swiss-Prot database, was utilized to characterize, and compare the functional capacities present in the effluents of a WWTP and carbon-based advanced treatment train intended for potable reuse. Functional based metagenomics is a promising approach to better understand the underlying microbial interactions associated with treatment processes present in drinking water and wastewater treatment, especially those designed to either employ biological treatment or disinfection. Of all microbial methods, NGS technologies are best suited to assess complex, environmental communities as holistically as current technology allows. This is especially important when specific molecular markers are unavailable, or when too many exist. Currently, the greatest limitations to the application of these techniques are related to the cost of sequencing, required expertise, and underdeveloped annotation and analysis pipelines. Functional based metagenomics also suffers from a comparatively small amount of fundamental literature, as the technologies themselves are still being rapidly developed, and the extremely numerous potential applications.

5.6.1 Introduction

The utilization of NGS sequencing to survey the functional capabilities of drinking water, wastewater, and water reuse systems remains one of metagenomics' applications with the largest potential for development. Thus far, few studies have fully leveraged functional based metagenomics within the drinking water and wastewater space and those that have commonly suffer from limited sample size (Chao et al., 2016; Reid et al., 2018), scope, and annotation quality. Most studies have primarily focused on activated sludge reactors and anaerobic digesters with very few focused on the interplay between treatment processes or more advanced water reuse trains. Experiments conducted in the Pruden lab focused on functionally assessing advanced water treatment trains intended for potable reuse, published in WRF Report (Pruden et al., 2020), and demonstrated the ability for NGS to shed light on metabolic functions of interests, specifically monooxygenase activity. These results showed enrichment of monooxygenase activity following ozonation with monooxygenase related genes experiencing peak relative abundances during biologically active carbon filtration. Though preliminary, the results speak to the efficacy of developing functionally related metagenomic pipelines for more in-depth surveying of water and wastewater-based systems, especially those employing biologically based treatment where most functional capabilities remain relatively 'blackbox'. Future work looks to expand on these preliminary results by harnessing improvements in database annotation to further characterize important functional capabilities associated with the present microbial communities. Of particular interest are genes associated with nutrient cycling and metabolic functions, especially those linked to CEC degradation and biological removal of trace organic carbon compounds.

The objective of the case study is to demonstrate the differences in select functional capacity between a wastewater's denitrified secondary effluent (indicated as pilot influent) and the effluent of the advanced treatment train prior to disinfection, specifically after GAC contacting with an EBCT of 20 minutes (indicated as GAC low flow).

5.6.2 Methods

Bulk water samples were collected monthly at an advanced treatment train's pilot influent, and GAC contactor effluent. A 1-liter sterile bottle, in triplicate, was filled at each sampling location within the treatment train and immediately stored onsite at 5 degrees Celsius. After all samples had been collected, they were transferred to a cooler and transported back to Virginia Tech on ice for processing. Microbial samples were then concentrated onto 0.22 µm mixed cellulose ester filters that were then folded into quarters, torn, and transferred to DNA extraction tubes and stored at -20 degrees Celsius. Exact sample volumes were recorded at the time of filtering. DNA was extracted for downstream analysis (FastDNA SPIN Kit for Soil, MP Biomedicals, Solon, OH). Samples were then mass normalized in house, before being submitted to the Scripps sequencing core for metagenomic analysis. Samples were sequenced using an Illumina NextSeq 500 High Output (San Diego, CA) after being subjected to the NEB Ultra II DNA Library Prep kit for Illumina. Raw reads where then uploaded to Metastorm (Arango-Argoty et al., 2016) and annotated to the UniProt Knowledgebase Swiss-Prot database (UniProt) using the read-matching pipeline. Annotated sequences where then imported into R-studio (RStudio Team, 2019) and analyzed using several libraries including ggplot (Wickham, 2016), and vegan (Oksanen et al., 2022). Results from Uniprot were filtered for functional genes associated with bacterial communities, functional genes associated with 'oxygenase' activities, and finally any genes specifically associated with the '[GO:0004497] monooxygenase activity'. A summary of NGS methodologies applied in this case study are outlined in Table 5-5.

Table 5-5. NGS Methodologies Applied in Case Study #5.

Procedure Component	Approach/Description	Reference
Source	Wastewater's denitrified secondary effluent (indicated as pilot influent) and the effluent of the advanced treatment train prior to disinfection, specifically after GAC contacting with an EBCT of 20 minutes (indicated as GAC20)	
Sample Concentration	Triplicate, 1-liter bulk water samples. Microbial samples are concentrated onto 0.22 µm mixed cellulose ester filters that are then folded into quarters, torn, and transferred to DNA extraction tubes and stored at -20 degrees Celsius.	
Nucleic Acid Extraction	FastDNA SPIN Kit for Soil	(MP Biomedicals, Solon, OH)
Sequencing Approach	Metagenomic Sequencing	

Library Preparation	NEB Ultra II DNA Library Prep kit for Illumina.	
Sequencing Platform/Configuration	Illumina NextSeq 500 High Output	(San Diego, CA)
Sequencing Coverage	2.1e7 ∓ 9.9e6 (4.4e7 to 1.2e7)	
Data Analysis	Read-matching focused on functional genes	
Annotation Database(s)	UniProt Knowledgebase Swiss-Prot, Metastorm	(Arango-Argoty, 2016; The UnitProt Consortium, 2019)

5.6.3 Results

Figure 5-7 graphically presents the bacterial community's functional capacity, as determined by metagenomic sequencing, for samples that are representative of denitrified wastewater secondary effluent and the effluent of the advanced treatment train utilizing GAC contacting, prior to disinfection. From the NDMS plot, a distinct difference is seen between the two sample groups (ANOSIM, p-value < 0.05, r-stat = 0.994). This indicates that the functional genes present within the bacterial commutates are unique to those treatments, allowing for the utilization of different metabolic pathways, degradation mechanisms, stress responses, and other microbial functions. These unique functional potentials are also related to treatment performance where both sampling locations are known to provide different contaminant removal and intended treatment.

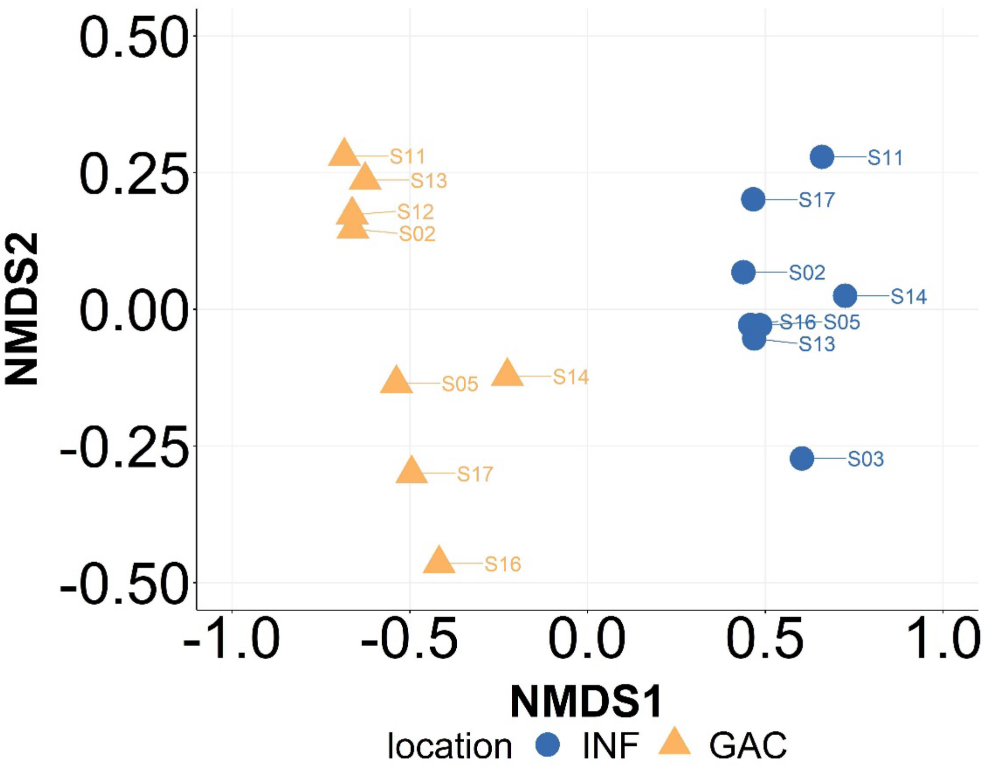


Figure 5-7. NDMS Plot of Bray-Curtis Dissimilarity Matrix for all Bacterially Associated Functional Genes. Samples are derived from a carbon-based advanced water treatment train including wastewater's denitrified secondary effluent (indicated as pilot influent) and the effluent of the advanced treatment train prior to disinfection, specifically after GAC contacting.

Figure provides 16S rRNA normalized abundances of oxygenase related genes subset from the entire functional potential represented in Figure 5-7. Oxygenases are of particular interest due to their ability to facilitate compound degradation (Bai et al., 2013; Mason et al., 2014; Silva et al., 2013). They are comprised of enzymes that incorporate molecular oxygen into various organic and inorganic compounds resulting in energy metabolism, biosynthesis, compound transformation, and degradation of essential metabolites (Lennarz and Lane, 2013). 16S rRNA normalized abundances of oxygenase related gene copies were higher following biologically active filtration and GAC contacting than wastewater treatment. This indicates that biologically active filtration and GAC contacting can select for and stimulate positive selection of compound degrading functional genes. Further, variability within the filters (S02 through S12 vs S13 through S17) seem to indicate that changes in operational conditions or water quality can impact the selection of functional genes.

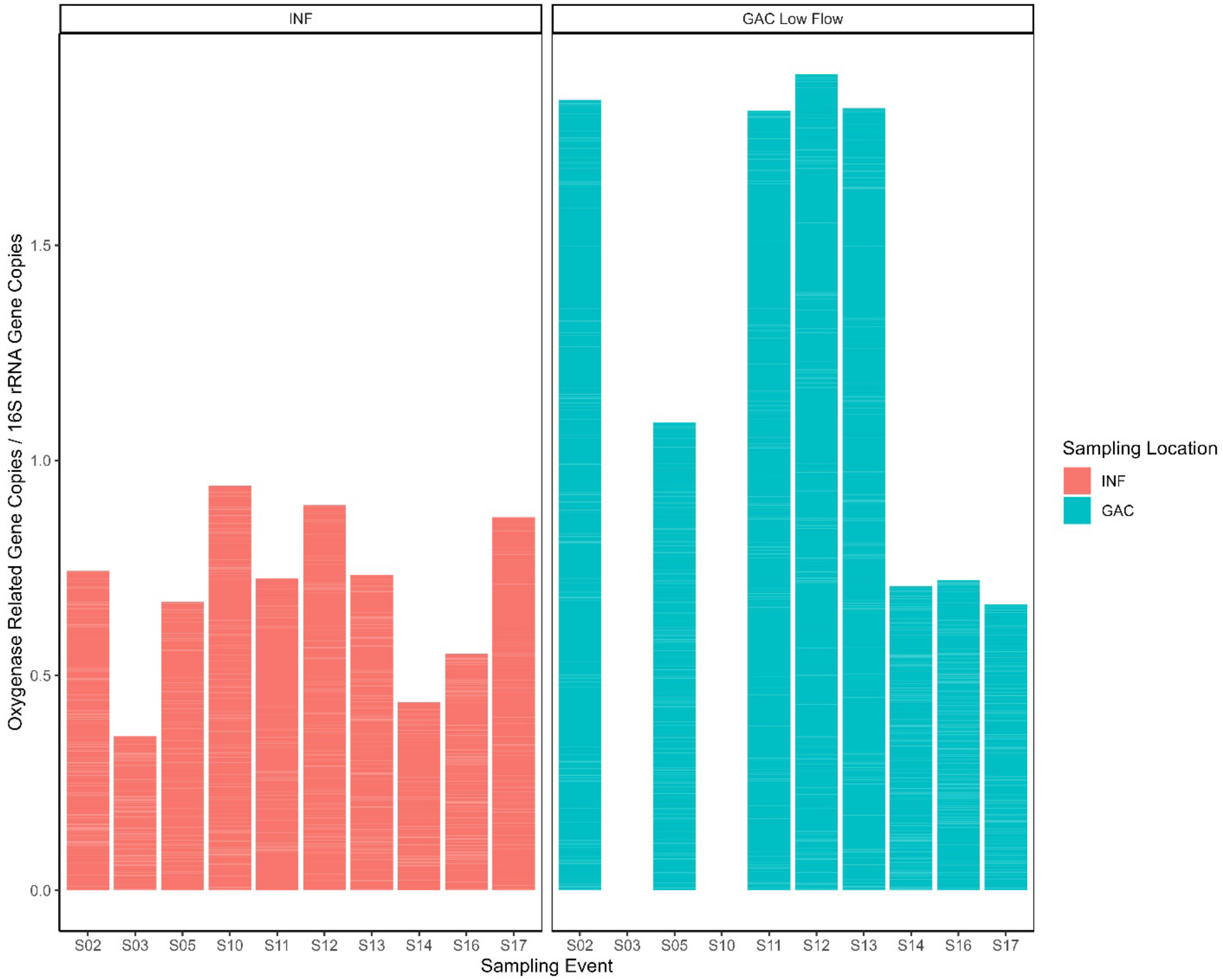


Figure 5-8. Bar Plot of all Oxygenase Related Gene Abundances Normalized to 16S rRNA Gene Abundances. Samples are derived from a carbon-based advanced water treatment train including wastewater's denitrified secondary effluent (indicated as pilot influent) and the effluent of the advanced treatment train prior to disinfection, specifically after GAC contacting.

In addition to identifying characteristics associated with the entire functional profile (Figure 5-7), and target gene groups of interest (Figure 5-8), functionally annotated metagenomics can also identify unique genes present within categories of interest. Figure 5-9 provides a Venn diagram of all unique genes associated with monooxygenase activities (GO:0004497) for each sample location. This data suggests that most monooxygenase related genes are shared between wastewater and GAC effluents. However, the latter does include more unique genes than wastewater effluent, indicating that not only are the 16S rRNA normalized abundances higher, but so is the presence of unique genes responsible for compound degradation and transformation.

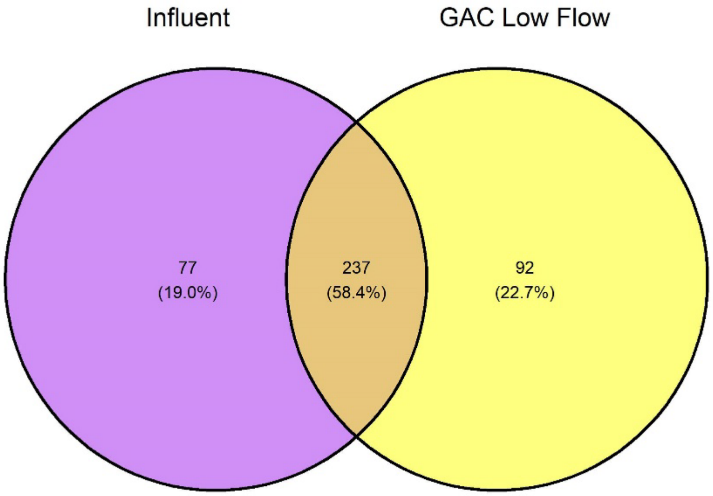


Figure 5-9. Venn Diagram of all Unique Genes Associated with the '[GO:0004497] Monooxygenase Activity' Functional Category at each Sampling Location.

5.6.4 Conclusion

In conclusion, NGS-based approaches applied to functional metagenomics were able to distinguish differences between wastewater and advanced treatment effluents' comprehensive functional capacities. Functional metagenomic analysis was also found to be flexible in its ability to assess groups and individual functional genes of interests. These results show promise in characterizing and optimizing complex biological processes, especially in wastewater and drinking water applications where biological process are often considered 'blackbox'. Currently, conventional molecular methods (both culture and more conventional molecular methods) lack the ability to characterize complex environmental communities while NGS technologies allow for more comprehensive characterizations even in the absence of specific gene targets. Continued development of functionally focused metagenomic pipelines should ultimately allow for increased control and optimization of biological treatment processes in addition to better understanding the microbial dynamics present within them.

CHAPTER 6

Validating Next-Generation Sequencing Techniques for Monitoring Water and Wastewater

6.1 Overview and Approach

A series of pilot-scale validation experiments were conducted to address key knowledge gaps needed to expand the application and consistency of NGS data produced for water and wastewater quality monitoring. The research team worked with utility partners to collect fresh wastewater samples that were promptly processed with uniform biological replication alongside experimental approaches at the forefront of environmental metagenomics research. In all, three individual experiments were performed to illustrate and explore key knowledge gaps in the field that address the application of experimental controls, the sensitivity of Illumina sequencing, and the fundamental effect of DNA extraction techniques on NGS data characteristics. The objectives, knowledge gaps, and experimental approaches that are covered in subsequent sections are outlined in Table 6-1.

Table 6-1. Overview of Objectives and Experimental Approaches for NGS Validation Studies.

Objective Number	Knowledge Gap	Approach
1	Quantitative capacity of shotgun Illumina sequencing	Spike internal DNA reference standards into replicate wastewater DNA extracts; assess the recovery of known reference concentrations
2	Technical limitations of shotgun Illumina sequencing	Serially-dilute internal DNA reference standards and spike into replicate wastewater DNA extracts; assess recovery of known reference concentrations as a function of dilution factor
3	Required sequencing depth per objective (antibiotic resistance, functional characterization, microbiome)	Sequence single samples with ultra-deep approach; randomly subsample large metagenomes to generate rarefaction curves
4	Effect of DNA extraction kit on NGS data (Illumina + Nanopore) produced	Extract identical sets of wastewater samples with high- and low-molecular weight extraction kits; sequence sets on both Illumina and Nanopore platforms
5	Comparability of short and long read data	Sequence two identical sets of samples on both Illumina and Nanopore platforms

6.2 Experiment 1: The Quantitative Capacity and Technical Limitations of Shotgun Illumina Sequencing

6.2.1 Rationale

Quantitative molecular approaches are needed for investigating water and wastewater quality as they allow for sensitive and relatively unbiased enumeration of microbial targets in complex matrices, providing distinct advantages over conventional culture-based assays. The gold standard for quantitative molecular detection of microorganisms for the past two decades has undoubtedly been qPCR. Realistically, however, qPCR is only capable of targeting a handful of targets on a routine basis where there are countless established and emerging microbiological contaminants of interest to the water industry. Recent advances in experimental controls designed for NGS studies are now allowing for the absolute quantification of genetic signatures where the field has primarily relied on semi-quantitative (i.e., relative abundance) metrics that are incompatible with conventional, target "per-volume of sample" datatypes. Here is a demonstration of a benchmarking of such NGS controls termed "sequins" (sequencing spike-ins) for enumerating ARGs in influent (INF), activated sludge (AS), and secondary effluent (SE) wastewater samples to explore the quantitative capacity and technical limits of the Illumina platform (Obj. 1 & 2).

6.2.2 Experimental Approach

Sequins designed for shotgun metagenomics (meta sequins; https://www.sequinstandards.com/metagenome/) are mixtures of 86 unique DNA oligonucleotides of varying lengths (987-9120 bp) and GC content (24-71%) that are present at 16 discrete input proportions, forming a ladder (akin to qPCR standard curves) with a 3.2x104 fold range. Each individual sequin oligo has no known homology to all known prokaryotic and eukaryotic reference sequences past a k-mer length of 25 nt allowing them to be detected reliably from complex environmental matrices (Reis et al., 2020). Meta sequins were spiked into replicate (10x) clean DNA extracts at logarithmically decreasing input mass (20 ng to 2x10-8 ng) to investigate the LOQ and LOD of this quantitative metagenomics (qMeta) experiment. The qMeta approach was then compared to qPCR-derived concentrations of 16S rRNA, sul1 (sulfonamide resistance), and *tet*A (tetracycline resistance) gene copies from the same replicate DNA extracts. A prebuilt bioinformatic package written by the designers of meta sequins, Anaquin (Wong et al., 2017), was used to precisely annotate Illumina reads assigned to the meta sequin ladders and determine their overall sequencing recovery. Table 6-2summarizes the filter volumes, final library mass inputs, input meta sequin masses, and the total recovered Illumina reads recovered per sample.

Table 6-2. Summary of Experimental Design and Results for the qMeta Experiment.

Sample	Filter Volume (mL)	Sample DNA Mass In (ng)	Meta Sequin Mass In (ng)	Reads Passing QA/QC (150 bp)	Base Pairs (Gb)	Total Sequin Ladder Reads	Unique Meta Sequins Detected (out of 86)
1-INF	50	1430	20	535,013,050	80.25	11,040,198	86
2-INF	50	745	2	682,264,862	102.34	1,073,673	83
3-INF	50	950	0.2	626,058,174	93.91	56,592	62
4-INF	50	1090	0.02	852,102,356	127.82	1,942	37
5-INF	50	900	0.002	537,603,870	80.64	332	27
6-INF	50	1250	0.0002	714,076,610	107.11	24	5
7-INF	50	1020	0.00002	484,396,382	72.66	2	2
8-INF	50	1310	0.000002	559,475,636	83.92	1	1

9-INF	50	1120	0.0000002	621,449,370	93.22	2	1
10-INF	50	1090	0.00000002	599,232,550	89.88	2	1
1-AS	10	950	20	663,293,200	99.49	7,551,758	86
2-AS	10	870	2	644,302,050	96.65	503,012	72
3-AS	10	945	0.2	561,812,936	84.27	42,191	57
4-AS	10	1050	0.02	689,547,724	103.43	3,755	41
5-AS	10	1120	0.002	692,374,412	103.86	310	20
6-AS	10	825	0.0002	644,008,238	96.60	8	3
7-AS	10	780	0.00002	644,583,050	96.69	4	1
8-AS	10	835	0.000002	727,957,052	109.19	7	3
9-AS	10	1040	0.0000002	607,594,870	91.14	6	2
10-AS	10	1080	0.00000002	587,974,204	88.20	0	0
1-SE	500	890	20	668,746,476	100.31	7,528,242	86
2-SE	500	1080	2	651,113,730	97.67	398,144	72
3-SE	500	995	0.2	674,983,984	101.25	30,929	55
4-SE	500	1110	0.02	654,267,052	98.14	2,185	41
5-SE	500	905	0.002	538,684,262	80.80	123	18
6-SE	500	875	0.0002	600,034,330	90.01	20	9
7-SE	500	954	0.00002	516,485,834	77.47	0	0
8-SE	500	950	0.000002	675,154,342	101.27	1	1
9-SE	500	950	0.0000002	575,233,546	86.29	1	1
10-SE	500	985	0.00000002	604,263,924	90.64	2	2

6.2.3 Results

6.2.3.1 The Quantitative Capacity of Illumina Sequencing

The linearity of spike-in meta sequin mass to recovered reads was first investigated by plotting the spiked concentration ratio (ng meta sequins/ng total library) against the sequence base ratio (bp meta sequin spiked in/bp reads detected). Strong linearity (Pearson, $R^2 = 0.983$, p<1e-16) was found between known meta sequin spike-in concentrations to detected reads across a wide concentration range (Figure 6-1A). This linearity was maintained with inputs as low as 10^{-9} ng/ng, suggesting that sequins are highly stable and reliably detected at low input abundances. All 86 sequins were detected at the highest input ladder mass across the three sample types and became intermittently identified as the input mass logarithmically decreased (Table 6-2). In general, longer and more GC neutral sequins were detected more frequently at the lowest input concentrations. This is likely due to the nature of Illumina library preparation where longer genomic fragments with stable GC contents generate a greater number of high-quality reads, increasing the likelihood of detection (Bowers et al., 2015). Individual sequin read counts at each input proportion and dilution factor between INF, AS, and SE samples were found to be statistically indistinguishable (paired t-test, p < 1e-10), indicating that the inherent nucleic acid complexity of the DNA extracts (i.e., matrix inhibition) did not influence general ladder recovery. Overall, these features validate the quantitative capacity of the Illumina platform and its ability to detect synthetic DNA oligonucleotides at the correct input concentrations, even at very low inputs.

6.2.3.2 The Technical Limitations of Illumina Sequencing

To explore the technical limitations of shotgun Illumina sequencing, the variability of detecting individual meta sequins at decreasing input concentrations was explored. Because there were minimal differences between sample matrices on ladder recovery, the INF, AS, and SE samples at each respective input mass proportion were treated as technical replicates. Plotting the coefficient of variation (CV; standard deviation/mean) of read counts as a function of input sequin concentrations between the technical replicates revealed a rapid increase in detection uncertainty as spiked concentrations fell below 10^5 copies per µL of DNA extract (Figure 6-1B). Using general recommendations for qPCR experiments, the LOQ can be stringently defined as the lowest input sequin concentration that was detected across all three technical replicates with a read count CV ≤ 0.35. Further, the LOD is defined here as the lowest individual sequin concentration detected across all samples. The LOQ and LOD for this qMeta experiment are defined as 2.7×10^4 copies/µL DNA extract (3.0×10^{-8} ng/ng) and 54 copies/µL DNA extract (1.9×10^{-11} ng/ng), respectively. Sequins that were detected with a single read were found at an input concentration range of 54 – 5500 copies/µL (1.9×10^{-11} – 3.7×10^{-9} ng/ng), which when converted to a per volume of water filtered basis, equates to approximately 1.0 – 243.6 input copies/mL, approaching the LOD of the qPCR experiments (1-10 gc/mL). There was no single instance of a sequin being detected at a calculated input copy number <1, suggesting strong agreement between theoretical stochiometric calculations and the physical presence and detection of individual reference standards.

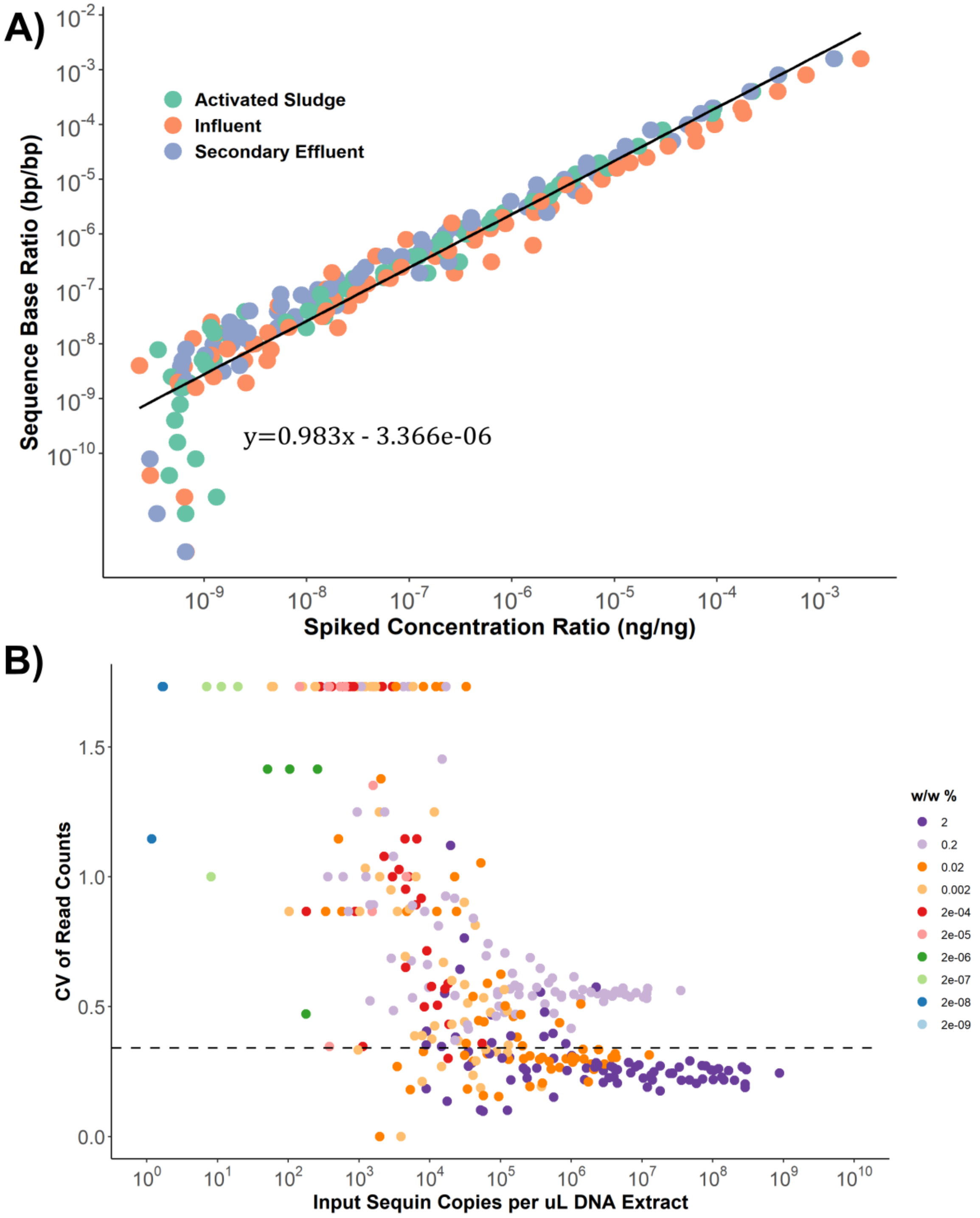


Figure 6-1. Technical Limitations of Shotgun Illumina Sequencing.

A) Spiked concentration ratio (ng/ng) to sequence base ratio (bp/bp) of internal DNA reference standards (meta sequins). B) CV of detected meta sequins as a function of input concentration. The black dotted line represents the threshold of variability for individual sequin detection, defining the LOQ for the experiment.

6.2.3.3 Comparison of qMeta with qPCR using Reference Genes

The qMeta approach was verified by comparing absolute gene quantities (gc per volume of water filtered) to replicate qPCR concentrations of three target genes, 16S rRNA genes, *sul*1, and *tet*A (Figure 6-2). For qMeta and qPCR gene calculations, all ten biological replicates were used. Across all samples, the research team found significant difference between gene quantities derived from qMeta and qPCR for 16S rRNA (paired t-test, $p < 0.001$) and *tet*A (paired t-test, $p < 0.001$), but not for *sul*1 (paired t-test, $p = 0.5981$). Similar discrepancies in gene abundance have been observed across other qMeta studies, where incongruencies between qPCR primer binding sites and reference ARGs used in qMeta calculations result in either under- or over-estimations of gene concentrations across assays. Despite these differences in derived gene target concentrations, the variance between calculated gene concentrations by methodology were insignificant. For the 16S rRNA gene concentrations, the CV of qPCR-derived concentrations (0.365 ± 0.03) was not significantly different than qMeta (0.445 ± 0.147) (paired t-test, $p = 0.41$). The same was observed for *sul*1, where the CV of qPCR concentrations (0.42 ± 0.209) and qMeta (0.472 ± 0.132) showed no significant difference (paired t-test, $p = 0.95$).

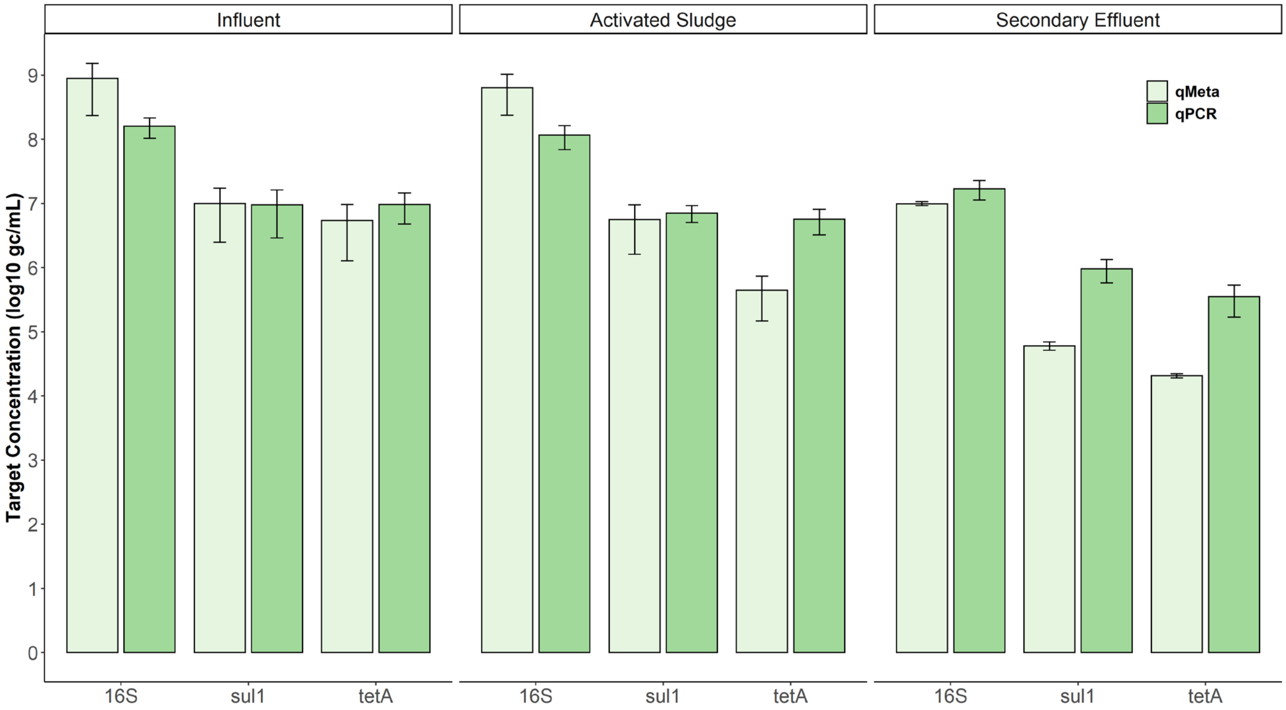


Figure 6-2. Comparison of qMeta and qPCR for Quantifying Three Gene Targets in Representative Wastewater Samples.

Each bar represents 10 biological replicates.

6.2.4 Conclusions and Implications

Here, the immediate quantitative capacity of shotgun Illumina sequencing for the enumeration of gene targets in environmental samples was demonstrated. The technical limits of the technique were also revealed with the LOQ for qMeta defined on the order of magnitude of 10^4

gc/mL, two to three orders of magnitude higher than comparable qPCR assays (LOQ; 10^1-10^2 gc/mL) despite the exorbitant sequencing depths achieved. Due to the differences in assay sensitivity and relative costs, the qMeta approach would be most useful when qPCR throughput and primer design limit the conclusions that can be drawn from complex sample types and where the consistent enumeration of rare gene targets at low abundances are not required. Discrepancies between gene concentrations between qMeta and qPCR were also observed, warranting caution for the immediate implementation of the technique for large scale ARG surveillance projects, for example (Pruden et al., 2021). On the other hand, the addition of internal DNA reference standards to environmental metagenomic studies is a great leap forward in the pursuit of reproducible and universally comparable water and wastewater monitoring data. The principles of qMeta demonstrated here can be readily retooled to enumerate many other targets of interest, for instance, pathogen markers, fecal indicators, and other functional genes.

6.3 Experiment 2: Ultra-deep Sequencing of Wastewater

6.3.1 Rationale

Shotgun metagenomics on the Illumina platform is the current standard for comprehensive investigations of water and wastewater microbiomes by a large margin. The disadvantage of the technique, however, is that it can be relatively less sensitive than other culture- or qPCR-based approaches in detecting individual targets with certainty, as demonstrated in Experiment 1. Limitations that arise during metagenomic experiments can be illustrated by the diversity of the highly studied *Enterobacteriaceae* family of bacteria that comprise many of the human and animal pathogens that concern water quality engineers. In *E. coli* alone, it has been estimated that only 10% of the 18,000 orthologues gene families that comprise its pangenome appear in all strains (Touchon et al., 2009). To further complicate matters, the profile of antibiotic resistance genes that is mediated by the *Enterobacteriaceae* family has already been catalogued at over 2000 distinct ARGs and ARG variants, making the detection rare or novel gene sequences akin to finding a needle in a very large, unknown, and growing haystack (Alcock et al., 2020).

For every discrete environmental sample, however, there is a discrete number of individual genomes, and therefore gene sequences, that comprise its metagenome. Theoretically, every organism within that metagenome can be sequenced and fully characterized, if only the investigator generates sufficient genomic data to cover the unknown regions (i.e., achieve 100% coverage) (Rodriguez-R and Konstantinidis, 2014). Attempting to completely, or near completely, classify every organism and/or gene sequence in an environmental sample is infeasible on a routine basis. Studying individual environmental metagenomes at extreme sequencing depths, however, may provide investigators with an upper-bounded benchmark as to the true diversity of a sample that will be routinely monitored in the future at shallower depths. Here the research team demonstrates an ultra-deep sequencing experiment (~1 Tb per sample) of INF, AS, and SE wastewater samples to explore the required sequencing depths needed to achieve saturated representation of unique protein sequences, ARGs, and bacterial species markers using manually constructed rarefaction curves (Obj 3).

6.3.2 Experimental Approach

The ultra-deep sequencing approach was achieved by leveraging the dataset generated to explore the technical limitations of the Illumina platform in Experiment 1. Because each of the ten samples for the INF, AS, and SE samples are unique biological replicates of the same wastewater samples, their respective sequences can be concatenated together to represent large, extended metagenomic datasets (Table 6-3). Once the large metagenomic datasets were generated, the seqtk tool (Li, 2022) was used to randomly subsample read pairs to discrete depths using a preset seed, preventing resampling of the same read pairs. Each of the three metagenomes were subsampled to 1M, 10M, 50M, 100M, 250M, 500M, 1B, and ~6B reads representing 0.15, 1.5, 7.5, 15, 37.5, 75, 150, and ~1000 Gb of sequencing data, respectively. Each unique subsample was then queried using diamond blastx (Buchfink et al., 2021) against the Comprehensive Antibiotic Resistance Database (Alcock et al., 2020) and SWISS-PROT (Bairoch and Apweiler, 2000) amino acid reference databases to query ARGs and unique protein sequences, respectively. MetaPhlAn3 (Truong et al., 2015), which uses a comprehensive set of marker genes to taxonomically classify all bacteria present in a sample, was also used to investigate the taxonomic diversity present. Rarefaction curves were then generated by plotting the richness of detected targets as a function of the sequencing depth.

Table 6-3. Summary of Sequencing Depths Achieved Across Wastewater Samples.

Sample	Number of Reads	Base Pairs	Giga Base Pairs (Gb)
Influent (INF)	6,211,672,860	9.31×10^{11}	931
Activated Sludge (AS)	6,463,447,736	9.69×10^{11}	969
Secondary Effluent (SE)	6,158,967,480	9.24×10^{11}	924

6.3.3 Results

The concatenation of replicate wastewater samples from Experiment 1 resulted in three ultra-deeply sequenced metagenomes of INF, AS, and SE, representing nearly 3 Tb of total sequencing data. Rarefaction curves generated by plotting the richness of unique genes and marker sequences as a function subsampling depth revealed that ~1Tb of sequencing data was sufficient to achieve near saturation of gene richness in wastewater samples (Figure 6-3). In every instance, the influent sample had the greatest gene richness across all three targets, followed by the secondary effluent, and the activated sludge. For total unique protein sequences, near saturation was achieved at approximately $1.5\text{x}10^5$, $1.0\text{x}10^5$, and $1.25\text{x}10^5$ for INF, AS, and SE respectively. These values ostensibly represent the upper-bound of the total possible detectable genes in each sample. Further, saturation for ARGs was achieved at approximately 1400, 850, and 1000 unique sequences for each respective wastewater compartment, representing 0.01% of total genes per sample on average. For microbiome diversity, only 854, 206, and 447 unique bacterial species marker genes were detected, which is typical for other taxonomic marker genes such as 16S rRNA genes using short-read Illumina data (Gweon et al., 2019). To achieve only 80% target richness across all targets and samples matrices, approximately 150 Gb of sequence data would need to be generated per sample.

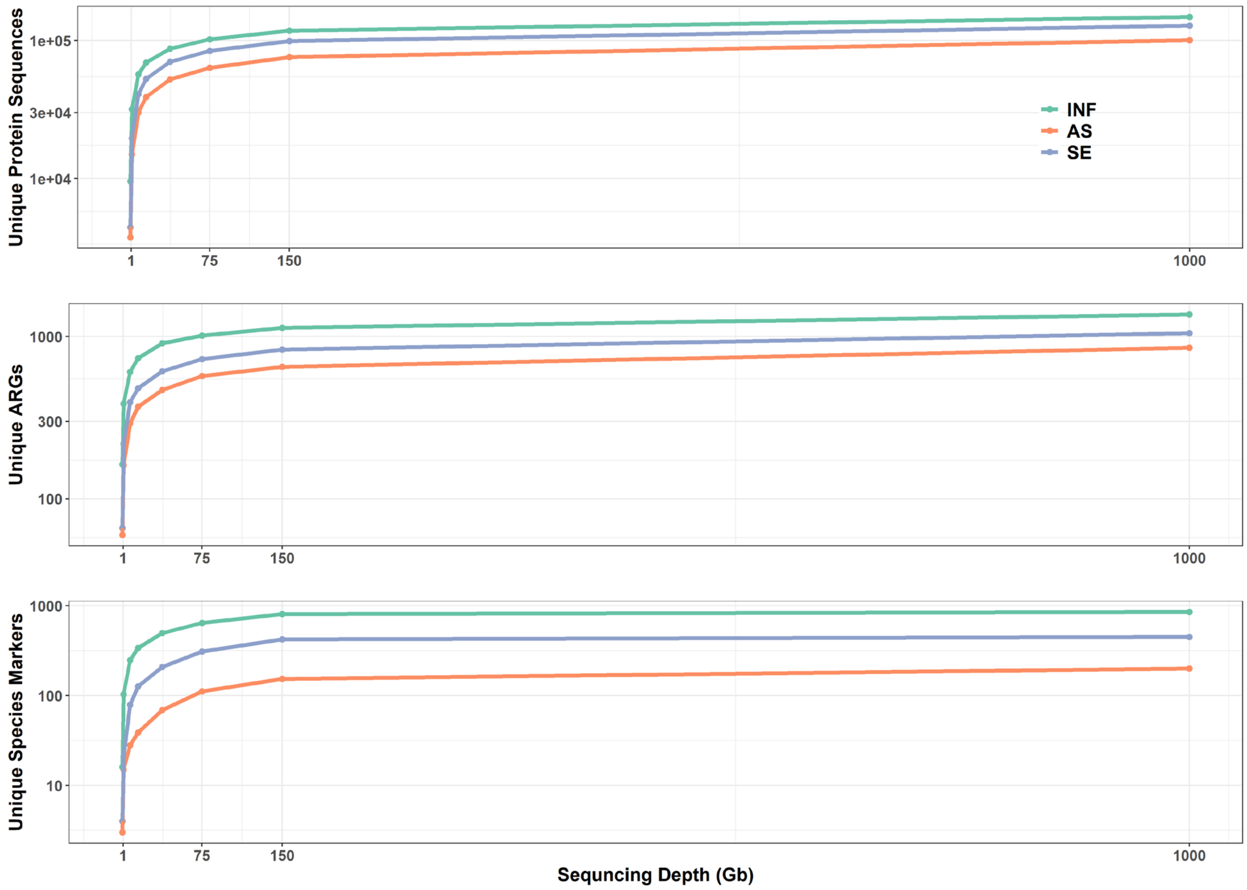


Figure 6-3. Rarefaction Curves for Ultra-deeply Sequenced Influent (INF), Activated Sludge (AS), and Secondary Effluent (SE) Samples.

Unique protein sequences were enumerated using the SWISS-PROT database, ARGs using CARD (v.3.0.3), and species markers using MetaPhlAn3.

6.3.4 Conclusions and Implications

In experiment 2, the upper bound of the total genetic diversity was explored using an ultra-deep sequencing approach and the manual construction of rarefaction curves. Near saturation of genetic diversity was achieved approaching 1 Tb of total sequencing data, with influent samples harboring the greatest diversity of bacterial protein sequences, ARGs, and unique species markers. Researchers and water quality engineers approaching the space of environmental metagenomics must understand that the genomic diversity of environmental microbiomes is vast, and the small fraction of the diversity that is sampled during routine monitoring may be insufficient to fully characterize the “true” nature of a sample.

6.4 Experiment 3: The Effect of DNA Extraction Methodology on NGS Data Produced

6.4.1 Rationale

One of the most critical steps in any NGS study is the appropriate extraction of nucleic acids and substantial method heterogeneity exists across fields. The efficacy of the assay employed will dictate the representativeness of the sampled environment based on the assay's ability to evenly lyse all cells and cell types of interest. The standard DNA extraction approach for environmental matrices is a bead-beating and spin-column kit, where cells are lysed using a combination of chemical dissolution and high shear forces and then freed DNA is retained on a positively charged filter inside a centrifugation column. These assays produce large yields of low molecular weight (LMW) DNA (i.e., short genomic fragments). These methods have been benchmarked for various environmental matrices and difficult organisms, including the most problematic, such as activated sludges and soils and endospores and Gram-positive bacteria (Knudsen et al., 2016; Kuhn et al., 2017; Li et al., 2018). However, as sequencing technologies continue to evolve, current DNA extraction methodologies may be insufficient for optimal data generation on newer and possibly even current sequencing platforms. For example, with the proliferation of long-read sequencing technologies on the Oxford Nanopore and Pacific Biosciences platforms in the last five years, data generation is optimized with long, uninterrupted strands of genomic DNA that conventional LMW bead beating kits cannot achieve.

High molecular weight (HMW) extraction techniques are those that forgo the high shear forces of bead-beating kits and rely on the enzymatic digestion of cellular membranes, preserving the integrity of long genomic fragments. The tradeoff being that they typically have lower yields than LMW approaches, and harder to lyse cellular morphologies may persist through extraction and bias the representation of the sample to primarily Gram-negative bacteria, for example. To date, few studies have directly compared the effect of LMW versus HMW extraction techniques on the quality and representativeness of NGS data produced. Here, the research team benchmarked the direct effect of DNA extraction methodology on the quality of NGS data generated through the reconstruction of mock communities. These experiments will shed light on the efficacy of LMW versus HMW extraction approaches to eventually optimize NGS workflows towards emerging techniques such as *de novo* hybrid assembly generation of MAGs from complex matrices.

6.4.2 Experimental Approach

Preliminary exploratory experiments were performed for the selection of a HMW DNA extraction kit for inclusion in subsequent experiments. Three commercial kits were tested, as well as a modified version of a validated LMW kit, the FastDNA Spin Kit for Soil. The various LMW and HMW kits tested are listed in Figure 6-4A. The research team modified the FastDNA Spin Kit for Soil to avoid mechanical shearing by utilizing the enzymatic digestion approach of the Zymo MagBead Kit, applying the latter half of the FastDNA Spin Kit for Soil's protein precipitation and DNA isolation via silica binding. Briefly, each of the 5 DNA extraction approaches listed were tested on freshy sampled influent, final effluent, and surface water sample replicates (1 biological replicate per kit). After extraction, quantification (Qubit

Fluorometer), and cleanup (Zymo DNA Clean-up and Concentration Kit), samples were sent for analysis on an Agilent TapeStation 4200 to quantitatively assess the DNA fragments produced. The TapeStation produces a histogram of DNA fragment lengths per sample and the peak of that distribution was used to assess assay efficacy (Figure 6-4B) alongside the total yield of double stranded DNA.

After the HMW kit was chosen, Zymo mock community standards (ZymoBIOMICS Microbial Community Standard, CAT# D6300) containing 10 microorganisms at known abundances (Table 6-4) were used to test the efficacy of the LMW (FastDNA Spin Kit for Soil) and HMW assays. Briefly, mock communities were extracted in biological triplicate using both the LMW and HMW kits and sequenced on the Illumina platform. Using MetaPhlan3, each sample was taxonomically classified to assess each extraction kit's ability to reconstruct the mock community at the correct relative proportions within the short-read datasets. To further assess mock community reconstruction, each sample was also *de novo* assembled using MEGAHIT (Li et al., 2015c) and the resultant contigs were binned using MetaBAT2 (Kang et al., 2019) and assigned taxonomy using the Genome Taxonomy Database Toolkit (GTDB-tk) (Chaumeil et al., 2020). The resulting bins (i.e., MAGs) were then assessed based on their completeness and relative abundance using checkM (Parks et al., 2015).

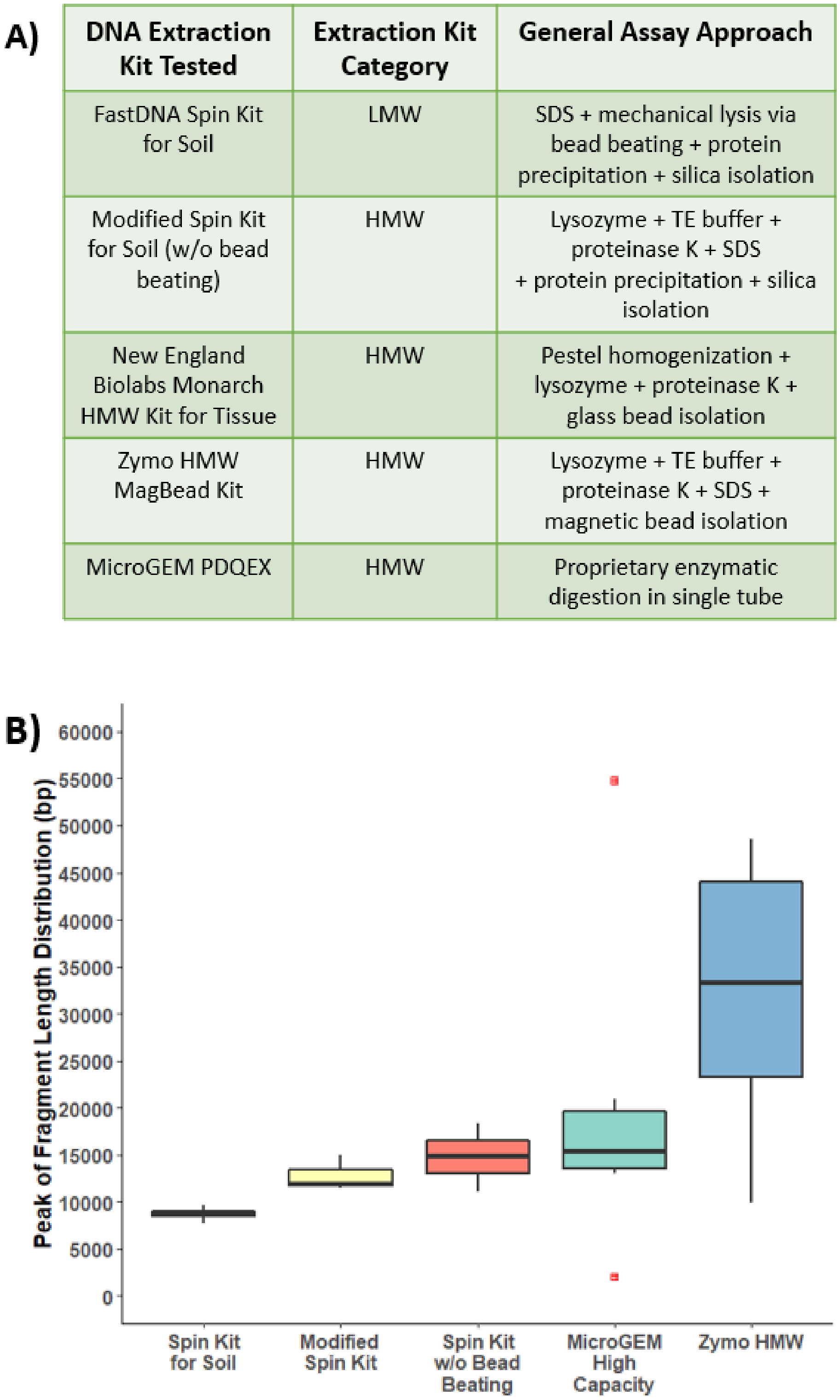
Table 6-4. Summary of Mock Community Composition.

Organism	Domain/ Kingdom	Morphology	Lysis Difficulty	Relative Abundance (%)
Listeria monocytogenes	Bacteria	Gram (+)	Difficult	12
Pseudomonas aeruginosa	Bacteria	Gram (-)	Easy	12
Bacillus subtilis	Bacteria	Gram (+)	Difficult	12
Escherichia coli	Bacteria	Gram (-)	Easy	12
Salmonella enterica	Bacteria	Gram (-)	Easy	12
Lactobacillus fermentum	Bacteria	Gram (+)	Difficult	12
Enterococcus faecalis	Bacteria	Gram (+)	Difficult	12
Staphylococcus aureus	Bacteria	Gram (+)	Difficult	12
Saccharomyces cerevisiae	Fungi	Encapsulated	Most Difficult	2
Cryptococcus neoformans	Fungi	Encapsulated	Most Difficult	2

6.4.3 Results

6.4.3.1 Preliminary High Molecular Weight Extraction Kit Testing

The Zymo HMW MagBead Kit outperformed all kits in producing the longest DNA fragments while also yielding sufficient DNA for library preparation (~1.5 µg). The New England Biolabs Monarch kit failed entirely between 9 different samples tested between two researchers. This kit was designed for pure cultures and tissue samples, and it's hypothesized that matrix inhibition prevented binding to the glass beads, resulting in null DNA recoveries. Although extremely efficient (~1 hr extraction per sample), the MicroGEM PDQEX prototype consistently produced DNA fragments that were well behind the Zymo kit. These data were used as justification for the use of the Zymo HMW MagBead kit in downstream head-to-head experiments.



A)

DNA Extraction Kit Tested	Extraction Kit Category	General Assay Approach
FastDNA Spin Kit for Soil	LMW	SDS + mechanical lysis via bead beating + protein precipitation + silica isolation
Modified Spin Kit for Soil (w/o bead beating)	HMW	Lysozyme + TE buffer + proteinase K + SDS + protein precipitation + silica isolation
New England Biolabs Monarch HMW Kit for Tissue	HMW	Pestel homogenization + lysozyme + proteinase K + glass bead isolation
Zymo HMW MagBead Kit	HMW	Lysozyme + TE buffer + proteinase K + SDS + magnetic bead isolation
MicroGEM PDQEX	HMW	Proprietary enzymatic digestion in single tube

Figure 6-4. Summary of HMW DNA Extraction Kit Results.
A) DNA extraction kits used and summary of the assay approach. B) Boxplots of the peak of the fragment length distributions across extraction replicates. Fragment length distributions were taken from analyses using the Agilent TapeStation 4200. Low molecular weight (LMW), high molecular weight (HMW), sodium dodecyl sulfate (SDS)

6.4.3.2 Reconstructing Mock Communities using Illumina Short Reads

The relative abundance of each taxa that constituted the mock community was calculated on a scale of 0-100 based on the total number of Illumina reads assigned. Remarkably, the proportions of each organism were tightly paired between the biological triplicates of each respective extraction kit, suggesting high reproducibility of the assay from the benchtop through to library preparation and sequencing. Using MetaPhlAn3's dataset of unique species marker genes, there were few instances of contaminant or mislabeled organisms across the triplicate extractions and generally constituted less than 0.0001% of the reads (Figure 6-5; denoted as "other"). Each extraction kit, however, displayed different biases towards different cellular morphologies. For the easy-to-lyse Gram-negative bacteria, the LMW and HMW kits performed similarly, detecting the approximate relative abundances of *Escherichia coli*, *Pseudomonas aeruginosa,* and *Salmonella enterica*. For the difficult-to-lyse Gram-positives, much greater heterogeneity was observed between relative taxa abundance, with *Lactobacillus fermentum* being an obvious outlier. The overrepresentation of *L. fermentum* across extraction approaches was unexpected and difficult to explain. Although annotation with a different pipeline (Kraken2) somewhat reduced the proportion of *L. fermentum*, the overrepresentation across the board suggests that the mock community might not have actually had the target proportions. In several instances, the LMW kit extracted more microbial DNA from the Gram-positive *Enterococcus faecalis*, *Listeria monocytogenes,* and *Staphylococcus aureus* than the HMW, owing to the presence of high mechanical shear forces that "crack open" these recalcitrant cells. For the most difficult cell types, the encapsulated fungi *Cryptococcus neoformans* and *Saccharomyces cerevisiae*, neither kit were able to successfully extract their DNA with appreciable efficiency, although the LMW kit had an average of 0.004% and 0.03% of each organism represented, respectively, whereas the HMW had null recoveries.

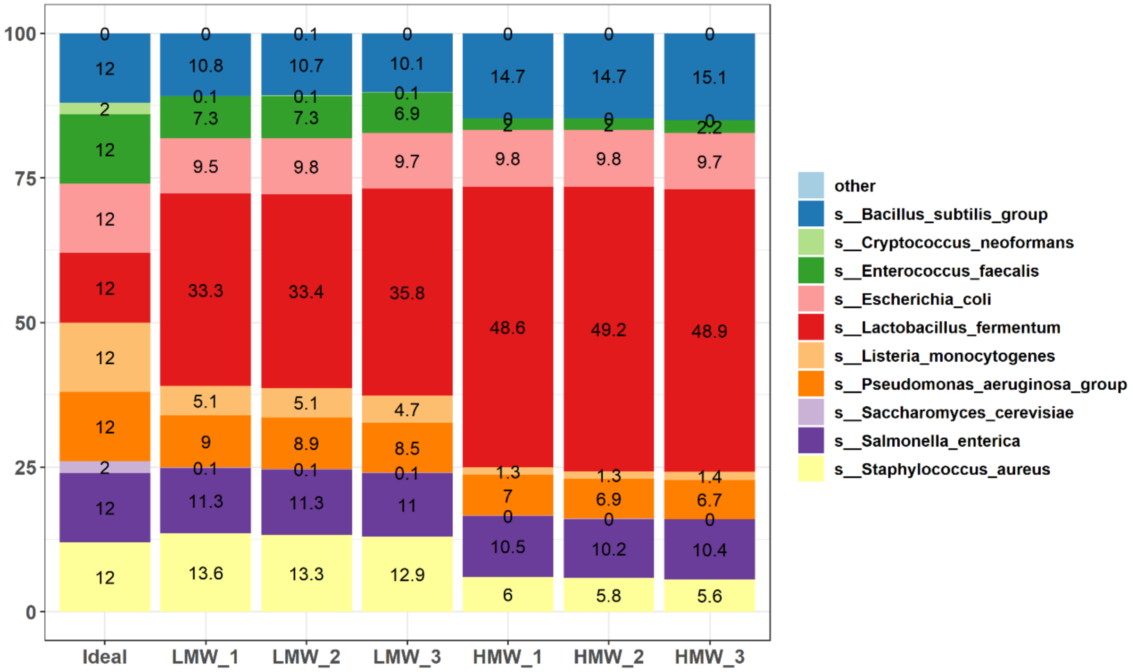


Figure 6-5. Assessment of Mock Community Reconstruction using Illumina Short-reads (MetaPhlAn3).
The "other" category constitutes contaminant or misidentified taxa with relative abundance < 0.01%.

6.4.3.3 The Effect of DNA Extraction on *de novo* Assembly

After *de novo* assembly, each sample was assessed for overall percentage of base pairs represented in the contig library, total number of contigs, N50's, L50's, and the distribution of contig lengths (Figure 6-6). The LMW kit facilitated the assembly of the greatest number of contigs (31,628 ± 2,572) over the HMW kit (12,784 ± 965), however, the bulk of these contigs were shorter than 5,000 bp (Figure 6-6B). Comparing the L50s and N50s of each extraction kit, which measures the number and length of contigs that represent 50% of the library size, the HMW far outperformed the LMW kit (Figure 6-6A and Figure 6-6C). In other words, the HMW approach was more efficient in generating longer contigs that represented a greater breadth of the diversity of the same metagenome using less contigs. Because each dataset was generated using a KAPA HyperPrep plus library preparation kit that utilizes shearing and size selection (~500 bp insert sizes), it's hypothesized that the longer, intact genomic fragments that were used as input for library preparation facilitated more efficient sequencing of adjacent genomic regions concurrently than the LMW kit. This represents a potentially significant advantage when attempting to reconstruct hypervariable or repeated regions of bacterial genomes, for example, features that are common to MGEs that harbor ARGs. Overall, the LMW kit generated a greater percentage of the total metagenome to be assembled in terms of total base pairs represented in the library, 30.2% ± 1.5% versus 14.8% ± 0.06%.

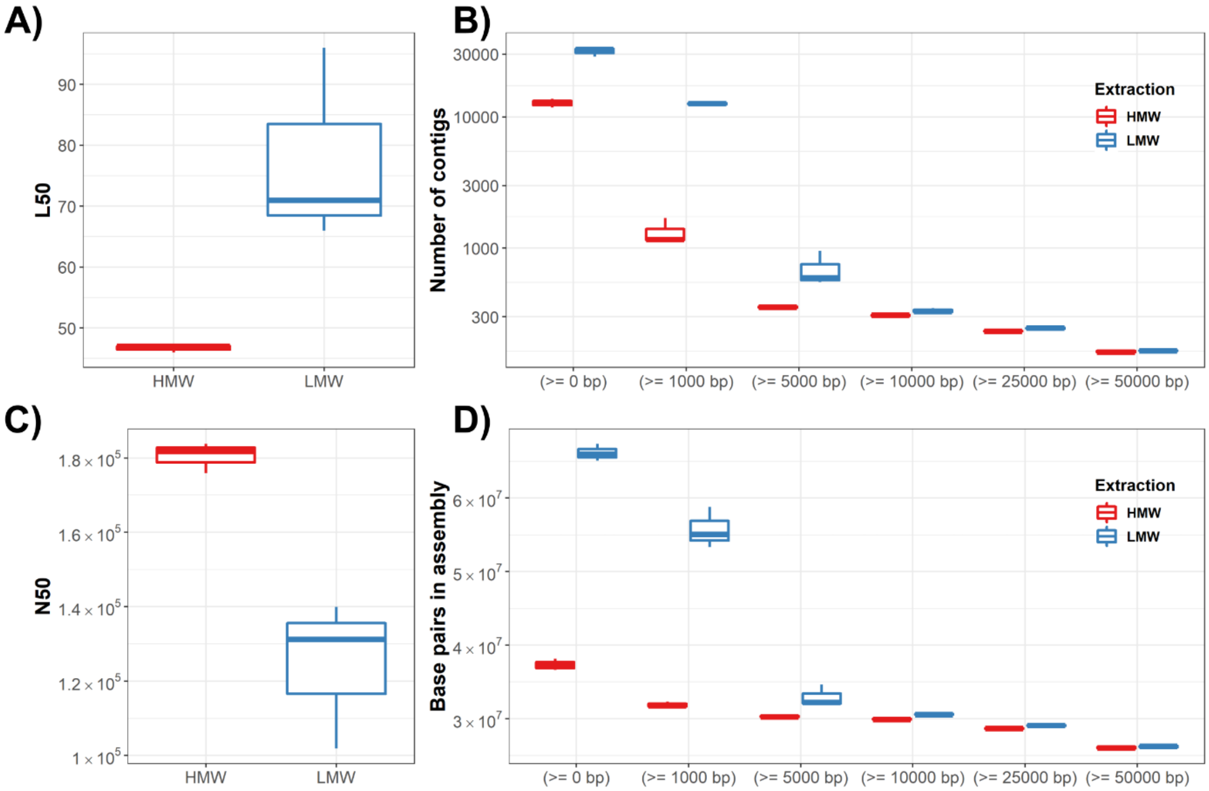


Figure 6-6. Summary Statistics for Assembled Contigs from Mock between Extraction Kits.

6.4.3.4 The Effect of DNA Extraction on the Quality and Completeness of MAGs

To further explore the effect of extraction kits on the mock community data, assembled contigs from each sample were binned by their respective taxonomic signatures to form MAGs. Each MAG was then evaluated for its completeness, degree of contamination (i.e., contigs added to bins erroneously), and relative abundance within each metagenome as function of individual species that represent the mock community. For the LMW and HMW samples, 55 and 54 MAGs were generated of varying completeness and contamination, respectively. Interestingly though, between the two extraction types, there was no statistical difference in the degree of completeness (t-test, p = 0.4616) or contamination (t-test, p = 0.4699) of the generated MAGs (Figure 6-7A and Figure 6-6B). Applying the universal Minimum Information about a Metagenome-Assembled Genome (MIMAG) standards (Bowers et al., 2017), the same number of high-quality (7) and medium-quality (40) MAGs were generated between the two assays. Generally, 50% completeness with less than 10% contamination are the minimum cutoffs for an acceptable MAG. The higher-quality MAGs in the dataset were primarily from Gram-negative bacteria: *P. aeruginosa*, *E. coli*, and *S. enterica* (Figure 6-7C), coinciding with the relative accuracy of the short-read assignments (Figure 6-5). *Lactobacillus fermentum* (denoted in Figure C as *Limosilactobacillus fermentum* due updated taxonomy within GTDB), which was overrepresented in the short read dataset, was inconsistently reconstructed, with a greater level of completeness from the HMW assay (42.9 ± 28.1%) than the LMW assay (29.2% ± 20.8%). Except for the missing two missing fungi and the three erroneous bins assigned to

Rhodococcus erythropolis, all eight bacteria were represented in the MAG dataset, demonstrating the relative accuracy of the *de novo* assembly and binning method.

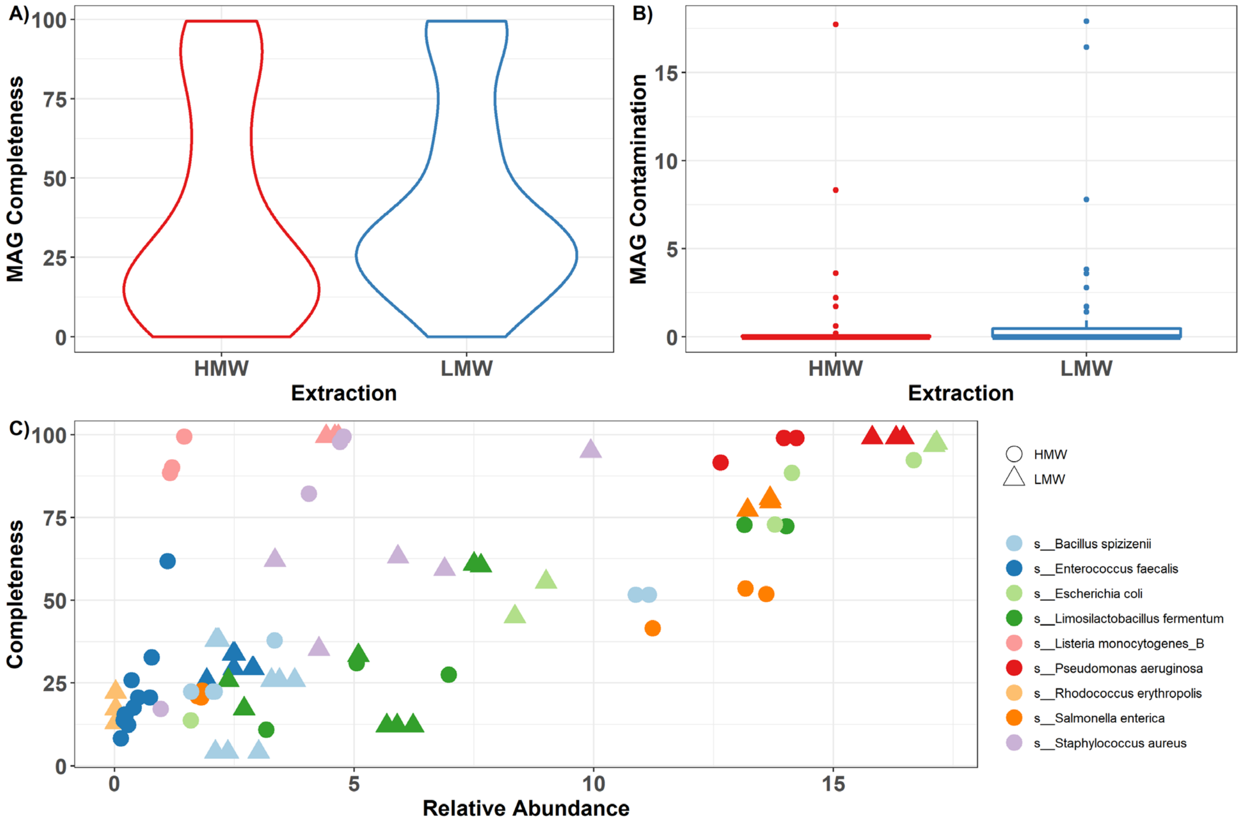


Figure 6-7. Summary Statistics for MAGs of the Zymo Mock Community across Extraction Kits.

6.4.4 Conclusion and Implications

DNA extraction methods remain one of the most critically important steps in any NGS workflow and dictate the representativeness and comparability of resulting metagenomic datasets. In experiment 3, the effect of “vigorous” versus “gentle” DNA extraction techniques on the reconstruction of mock microbial communities was demonstrated to have relatively large consequences on derived relative abundances of short-reads and the fundamental characteristics of *de novo* assembled contigs. At scale, and on real world samples, the effect of LMW versus HMW extraction methodologies would inevitably be exacerbated as individual environmental sample matrices would inhibit effectiveness of enzymatic lysis. The direct benefit of HMW extractions would not be fully realized until long-read sequencing is compared side by side with the present experiment.

CHAPTER 7

Conclusion and Recommendations

There is enormous value that NGS technologies pose to bring to the water industry. A systematic literature review and input from water utility stakeholders helped to inform the comprehensive guidance provided in this report for the application of NGS in the water industry. Promising field applications that advance the understanding of water, wastewater, and water reuse were identified, including pathogen profiling, functional and catabolic gene characterization, AMR profiling, bacterial toxicity potential assessment, monitoring of cyanobacteria and harmful algal blooms, and characterization of viruses.

Results from utility stakeholders clearly indicated that awareness of NGS approaches to water quality monitoring is rising and its usefulness for addressing emerging pathogens and contaminants (e.g., antibiotic resistance) is increasingly being recognized. However, there are many barriers to adoption of NGS, especially cost and lack of a clear entry point for water utilities. Based on the literature review, a detailed explanation of each step of NGS application, "from benchtop to laptop" is provided in this report. Specifically, the options for each step (such as sample concentration or nucleic acid concentration method) are described within a framework of pros and cons for the specific applications of interest.

NGS technologies are rapidly evolving and will likely continue to do so, both at the benchtop and the laptop. Thus, this report can provide timely guidance needed to address key limitations to improve the representativeness, accuracy, reproducibility, and interpretability of applications to water environments. Informed sampling and monitoring programs, including robust experimental controls, can help to advance the current state of NGS technologies to reach their full potential as a powerful tool for investigating various dimensions of water and wastewater systems that were previously inaccessible.

Still, it is important to be aware of the limitations of NGS. Although metagenomics is a powerful "non-target" tool for broadly accessing the genomes of multiple microbes inhabiting an environment, it is important to also be aware of inherent limitations. Validation studies performed here helped to place boundaries on sequencing depths (a major driver of NGS cost) needed for various objectives, but also confirm limitations of metagenomics for detecting rare targets at low abundance. Short read sequencing technologies, such as Illumina, are also inherently limited in the ability to link an identified gene (such as an antibiotic resistance gene) to a host organism. Long-read sequencing circumvents this problem, by providing sequence information of multiple genes along the same segment of DNA. Emerging techniques, such as *de novo* hybrid assembly using both short- and long-read sequencing data, may unlock the strain-resolved specificity needed for wastewater-based epidemiological studies of emerging viral and bacterial pathogens.

Also, despite the advances in the molecular detection of microorganisms, the inability to directly confirm the viability of the microbes detected is an inherent limitation of all nucleic acid based techniques, including NGS. Thus, it is important not to forget the enormous utility of culture assays for enumerating viable organisms with measurable phenotypes. In many cases, the implementation of NGS is best performed in tandem with conventional culture- and qPCR-based assays to provide the most complete understanding of complex microbial systems. In particular, NGS can be applied to pure cultures of organisms to obtain whole genomes, which provides the highest possible strain-level resolution and can be used to differentiate pathogenic from non-pathogenic strains of the same organism. WGS can also be used to track pathogens and other organisms of interest in the environment and link them to sources.

Finally, comprehensive validation studies were conducted to expand the application and consistency of NGS investigations of wastewater. Key knowledge gaps including the absolute quantitative capacity and the LOQ and LOD of shotgun Illumina sequencing, as well as the direct effect of DNA extraction methodologies for reconstruction microbial communities were addressed. Further development and validation would be beneficial to aid these technologies in becoming accessible to water quality professionals on a routine basis, including standardized methodologies, quality assurance controls, prohibitive costs, and manageable data analysis pipelines with readily interpretable, actionable results. The guidance developed here provides a framework for addressing these needs to help advance the application of NGS in a manner that maximizes potential benefit to the water industry.

References

Abbai, N.S., Pillay, B., 2013. Analysis of Hydrocarbon-Contaminated Groundwater Metagenomes as Revealed by High-Throughput Sequencing. Mol. Biotechnol. 54, 900–912. https://doi.org/10.1007/s12033-012-9639-z

Abbaszadegan, M., Huber, M., Pepper, I. and Gerba, C. 1993. Detection of viable Giardia cysts in water samples using polymerase chain reaction, pp. 529-548.

Abia, A.L.K., Alisoltani, A., Keshri, J., Ubomba-Jaswa, E., 2018. Metagenomic analysis of the bacterial communities and their functional profiles in water and sediments of the Apies River, South Africa, as a function of land use. Sci Total Env. 616–617, 326–334. https://doi.org/10.1016/j.scitotenv.2017.10.322

Adriaenssens, E.M., Farkas, K., Harrison, C., Jones, D.L., Allison, H.E., McCarthy, A.J., 2018. Viromic Analysis of Wastewater Input to a River Catchment Reveals a Diverse Assemblage of RNA Viruses. mSystems 3. https://doi.org/10.1128/mSystems.00025-18

Ahmed, W., Angel, N., Edson, J., Bibby, K., Bivins, A., O'Brien, J.W., Choi, P.M., Kitajima, M., Simpson, S.L., Li, J., 2020. First confirmed detection of SARS-CoV-2 in untreated wastewater in Australia: a proof of concept for the wastewater surveillance of COVID-19 in the community. Science of The Total Environment 728, 138764.

Albertsen, M., Hansen, L.B.S., Saunders, A.M., Nielsen, P.H., Nielsen, K.L., 2012. A metagenome of a full-scale microbial community carrying out enhanced biological phosphorus removal. ISME J 6, 1094–1106. https://doi.org/10.1038/ismej.2011.176

Albertsen, M., Karst, S.M., Ziegler, A.S., Kirkegaard, R.H., Nielsen, P.H., 2015. Back to Basics--The Influence of DNA Extraction and Primer Choice on Phylogenetic Analysis of Activated Sludge Communities. PLoS One 10, e0132783. https://doi.org/10.1371/journal.pone.0132783

Alcock, B.P., Raphenya, A.R., Lau, T.T.Y., Tsang, K.K., Bouchard, M., Edalatmand, A., Huynh, W., Nguyen, A.-L.V., Cheng, A.A., Liu, S., Min, S.Y., Miroshnichenko, A., Tran, H.-K., Werfalli, R.E., Nasir, J.A., Oloni, M., Speicher, D.J., Florescu, A., Singh, B., Faltyn, M., Hernandez-Koutoucheva, A., Sharma, A.N., Bordeleau, E., Pawlowski, A.C., Zubyk, H.L., Dooley, D., Griffiths, E., Maguire, F., Winsor, G.L., Beiko, R.G., Brinkman, F.S.L., Hsiao, W.W.L., Domselaar, G.V., McArthur, A.G., 2020. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res. 48, D517–D525. https://doi.org/10.1093/nar/gkz935

Alhamlan, F.S., Ederer, M.M., Brown, C.J., Coats, E.R., Crawford, R.L., 2013. Metagenomics-based analysis of viral communities in dairy lagoon wastewater. J Microbiol Methods 92, 183–188. https://doi.org/10.1016/j.mimet.2012.11.016

Alleron, L., Merlet, N., Lacombe, C., Frère, J., 2008. Long-term survival of Legionella pneumophila in the viable but nonculturable state after monochloramine treatment. Curr Microbiol 57, 497–502. https://doi.org/10.1007/s00284-008-9275-9

Alneberg, J., Karlsson, C.M.G., Divne, A.-M., Bergin, C., Homa, F., Lindh, M.V., Hugerth, L.W., Ettema, T.J.G., Bertilsson, S., Andersson, A.F., Pinhassi, J., 2018. Genomes from uncultivated prokaryotes: a comparison of metagenome-assembled and single-amplified genomes. Microbiome 6, 173. https://doi.org/10.1186/s40168-018-0550-0

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

American Public Health Association (APHA), 2017. Standard Methods for the Examination of Water and Wastewater. American Public Health Association (APHA): Washington, DC, USA.

Amir, A., McDonald, D., Navas-Molina, J.A., Kopylova, E., Morton, J.T., Zech Xu, Z., Kightley, E.P., Thompson, L.R., Hyde, E.R., Gonzalez, A., Knight, R., 2017. Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns. mSystems 2, e00191-16. https://doi.org/10.1128/mSystems.00191-16

Amos, B., Aurrecoechea, C., Barba, M., Barreto, A., Basenko, E.Y., Bażant, W., Belnap, R., Blevins, A.S., Böhme, U., Brestelli, J., Brunk, B.P., Caddick, M., Callan, D., Campbell, L., Christensen, M.B., Christophides, G.K., Crouch, K., Davis, K., DeBarry, J., Doherty, R., Duan, Y., Dunn, M., Falke, D., Fisher, S., Flicek, P., Fox, B., Gajria, B., Giraldo-Calderón, G.I., Harb, O.S., Harper, E., Hertz-Fowler, C., Hickman, M.J., Howington, C., Hu, S., Humphrey, J., Iodice, J., Jones, A., Judkins, J., Kelly, S.A., Kissinger, J.C., Kwon, D.K., Lamoureux, K., Lawson, D., Li, W., Lies, K., Lodha, D., Long, J., MacCallum, R.M., Maslen, G., McDowell, M.A., Nabrzyski, J., Roos, D.S., Rund, S.S.C., Schulman, S.W., Shanmugasundram, A., Sitnik, V., Spruill, D., Starns, D., Stoeckert, C.J., Jr, Tomko, S.S., Wang, H., Warrenfeltz, S., Wieck, R., Wilkinson, P.A., Xu, L., Zheng, J., 2022. VEuPathDB: the eukaryotic pathogen, vector and host bioinformatics resource center. Nucleic Acids Research 50, D898–D911. https://doi.org/10.1093/nar/gkab929

Amos, G.C.A., Zhang, L., Hawkey, P.M., Gaze, W.H., Wellington, E.M., 2014. Functional metagenomic analysis reveals rivers are a reservoir for diverse antibiotic resistance genes. Veterinary Microbiology 171, 441–447. https://doi.org/10.1016/j.vetmic.2014.02.017

Antipov, D., Korobeynikov, A., McLean, J.S., Pevzner, P.A., 2016. hybridSPAdes: an algorithm for hybrid assembly of short and long reads. Bioinformatics 32, 1009–1015. https://doi.org/10.1093/BIOINFORMATICS/BTV688

Arango-Argoty, G., Garner, E., Pruden, A., Heath, L.S., Vikesland, P., Zhang, L., 2018. DeepARG: a deep learning approach for predicting antibiotic resistance genes from metagenomic data. Microbiome 6, 23. https://doi.org/10.1186/s40168-018-0401-z

Arango-Argoty, G., Singh, G., Heath, L.S., Pruden, A., Xiao, W., Zhang, L., 2016. MetaStorm: A public resource for customizable metagenomics annotation. PLoS ONE 11, 1–13. https://doi.org/10.1371/journal.pone.0162442

Arango-Argoty, G.A., Dai, D., Pruden, A., Vikesland, P., Heath, L.S., Zhang, L., 2019. NanoARG: a web service for detecting and contextualizing antimicrobial resistance genes from nanopore-derived metagenomes. Microbiome 7, 88. https://doi.org/10.1186/s40168-019-0703-9

Arango-Argoty, G.A., Guron, G.K.P., Garner, E., Riquelme, M.V., Heath, L.S., Pruden, A., Vikesland, P.J., Zhang, L., 2020. ARGminer: a web platform for the crowdsourcing-based curation of antibiotic resistance genes. Bioinformatics 36, 2966–2973. https://doi.org/10.1093/bioinformatics/btaa095

Ardui, S., Ameur, A., Vermeesch, J.R., Hestand, M.S., 2018. Single molecule real-time (SMRT) sequencing comes of age: Applications and utilities for medical diagnostics. Nucleic Acids Res 46, 2159–2168. https://doi.org/10.1093/nar/gky066

Ayling, M., Clark, M.D., Leggett, R.M., 2019. New approaches for metagenome assembly with short reads. Brief. Bioinform. 21, 584–594. https://doi.org/10.1093/bib/bbz020

Bae, S., Lyons, C., Onstad, N., 2019. A culture-dependent and metagenomic approach of household drinking water from the source to point of use in a developing country. Water Res. X 2, 100026. https://doi.org/10.1016/j.wroa.2019.100026

Bae, S., Wuertz, S., 2009. Discrimination of viable and dead fecal Bacteroidales bacteria by quantitative PCR with propidium monoazide. Appl Env. Microbiol 75, 2940–2944. https://doi.org/10.1128/AEM.01333-08

Bai, Y., Liu, R., Liang, J., Qu, J., 2013. Integrated metagenomic and physiochemical analyses to evaluate the potential role of microbes in the sand filter of a drinking water treatment system. PLoS One 8, e61011. https://doi.org/10.1371/journal.pone.0061011

Bai, Y., Qi, W., Liang, J., Qu, J., 2014. Using high-throughput sequencing to assess the impacts of treated and untreated wastewater discharge on prokaryotic communities in an urban river. Appl. Microbiol. Biotechnol. 98, 1841–1851. https://doi.org/10.1007/s00253-013-5116-2

Bai, Y.; Liu, R.; Liang, J.; Qu, J. Integrated Metagenomic and Physiochemical Analyses to Evaluate the Potential Role of Microbes in the Sand Filter of a Drinking Water Treatment System. PLoS One 2013, 8 (4), e61011. https://doi.org/10.1371/journal.pone.0061011.
Baird, R., Laura Bridgewater, 2017. Standard methods for the examination of water and wastewater, 23rd ed. American Public Health Association, Washington, D.C.

Bairoch, A., Apweiler, R., 2000. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Research 28, 45–48. https://doi.org/10.1093/nar/28.1.45

Bakal, T., Janata, J., Sabova, L., Grabic, R., Zlabek, V., Najmanova, L., 2019. Suitability and setup of next-generation sequencing-based method for taxonomic characterization of aquatic microbial biofilm. Folia Microbiol 64, 9–17. https://doi.org/10.1007/s12223-018-0624-1

Balaguru, K., Judi, D.R. and Leung, L.R. 2016. Future hurricane storm surge risk for the US gulf and Florida coasts based on projections of thermodynamic potential intensity. Climatic Change 138(1-2), 99-110.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19, 455–477. https://doi.org/10.1089/cmb.2012.0021

Baral, D., Dvorak, B.I., Admiraal, D., Jia, S., Zhang, C., Li, X., 2018. Tracking the Sources of Antibiotic Resistance Genes in an Urban Stream during Wet Weather using Shotgun Metagenomic Analyses. Environ. Sci. Technol. 52, 9033–9044. https://doi.org/10.1021/acs.est.8b01219

Bartley, P.B., Ben Zakour, N.L., Stanton-Cook, M., Muguli, R., Prado, L., Garnys, V., Taylor, K., Barnett, T.C., Pinna, G., Robson, J., Paterson, D.L., Walker, M.J., Schembri, M.A., Beatson, S.A., 2016. Hospital-wide eradication of a nosocomial legionella pneumophila serogroup 1 outbreak. Clin Infect Dis 62, 273–279. https://doi.org/10.1093/cid/civ870

Bashiardes, S., Zilberman-Schapira, G., Elinav, E., 2016. Use of Metatranscriptomics in Microbiome Research. Bioinform Biol Insights 10, BBI.S34610. https://doi.org/10.4137/bbi.s34610

Batovska, J., Lynch, S.E., Rodoni, B.C., Sawbridge, T.I., Cogan, N.O., 2017. Metagenomic arbovirus detection using MinION nanopore sequencing. J Virol Methods 249, 79–84. https://doi.org/10.1016/j.jviromet.2017.08.019

Bautista-de Los Santos, Q.M., Schroeder, J.L., Blakemore, O., Moses, J., Haffey, M., Sloan, W., Pinto, A.J., 2016. The impact of sampling, PCR, and sequencing replication on discerning changes in drinking water bacterial community over diurnal time-scales. Water Res. 90, 216–224. https://doi.org/10.1016/j.watres.2015.12.010

Be, N.A., Avila-Herrera, A., Allen, J.E., Singh, N., Checinska Sielaff, A., Jaing, C., Venkateswaran, K., 2017. Whole metagenome profiles of particulates collected from the International Space Station. Microbiome 5, 81. https://doi.org/10.1186/s40168-017-0292-4

Bedoya, K., Coltell, O., Cabarcas, F., Alzate, J.F., 2019. Metagenomic assessment of the microbial community and methanogenic pathways in biosolids from a municipal wastewater treatment plant in Medellín, Colombia. Sci. Total Environ. 648, 572–581. https://doi.org/10.1016/j.scitotenv.2018.08.119

Beghini, F., McIver, L.J., Blanco-Míguez, A., Dubois, L., Asnicar, F., Maharjan, S., Mailyan, A., Thomas, A.M., Manghi, P., Valles-Colomer, M., Weingart, G., Zhang, Y., Zolfo, M., Huttenhower, C., Franzosa, E.A., Segata, N., 2020. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. bioRxiv. https://doi.org/10.1101/2020.11.19.388223

Behzad, H., Gojobori, T., Mineta, K., 2015. Challenges and Opportunities of Airborne Metagenomics. Genome Biol Evol 7, 1216–1226. https://doi.org/10.1093/gbe/evv064

Bekliz, M., Brandani, J., Bourquin, M., Battin, T.J., Peter, H., 2019. Benchmarking protocols for the metagenomic analysis of stream biofilm viromes. PeerJ 7, e8187. https://doi.org/10.7717/peerj.8187

Belder, D. De, Lucero, C., Rapoport, M., Rosato, A., Faccone, D., Petroni, A., Pasteran, F., Albornoz, E., Corso, A., Gomez, S.A., 2017. Genetic Diversity of KPC-Producing *Escherichia coli*, *Klebsiella oxytoca*, *Serratia marcescenes*, and *Citrobacter freundii* isolates from Argentina. Microbial Drug Resistance 00. https://doi.org/10.1089/mdr.2017.0213

Bengtsson-Palme, J., Hammarén, R., Pal, C., Östman, M., Björlenius, B., Flach, C.-F., Fick, J., Kristiansson, E., Tysklind, M., Larsson, D.G.J., 2016. Elucidating selection processes for antibiotic resistance in sewage treatment plants using metagenomics. Sci Total Env. 572, 697–712. https://doi.org/10.1016/j.scitotenv.2016.06.228

Berney, M., Hammes, F., Bosshard, F., Weilenmann, H.-U., Egli, T., 2007. Assessment and interpretation of bacterial viability by using the LIVE/DEAD BacLight Kit in combination with flow cytometry. Appl Env. Microbiol 73, 3283–3290. https://doi.org/10.1128/AEM.02750-06

Bertrand, D., Shaw, J., Kalathiyappan, M., Ng, A.H.Q., Kumar, M.S., Li, C., Dvornicic, M., Soldo, J.P., Koh, J.Y., Tong, C., Ng, O.T., Barkham, T., Young, B., Marimuthu, K., Chng, K.R., Sikic, M., Nagarajan, N., 2019. Hybrid metagenomic assembly enables high-resolution analysis of resistance determinants and mobile elements in human microbiomes. Nat. Biotechnol. 37, 937–944. https://doi.org/10.1038/s41587-019-0191-2

Bharti, R., Grimm, D.G., 2021. Current challenges and best-practice protocols for microbiome analysis. Briefings in Bioinformatics 22, 178–193. https://doi.org/10.1093/bib/bbz155

Bibby, K., Peccia, J., 2013. Identification of viral pathogen diversity in sewage sludge by metagenome analysis. Env. Sci Technol 47, 1945–1951. https://doi.org/10.1021/es305181x

Bibby, K., Viau, E., Peccia, J., 2010. Pyrosequencing of the 16S rRNA gene to reveal bacterial pathogen diversity in biosolids. Water Res 44, 4252–4260. https://doi.org/10.1016/j.watres.2010.05.039

Bibby, K., Viau, E., Peccia, J., 2011. Viral metagenome analysis to guide human pathogen monitoring in environmental samples. Letters in Applied Microbiology 52, 386–392. https://doi.org/10.1111/J.1472-765X.2011.03014.X

Biderre-Petit, C., Taib, N., Gardon, H., Hochart, C., Debroas, D., 2019. New insights into the pelagic microorganisms involved in the methane cycle in the meromictic Lake Pavin through metagenomics. FEMS Microbiol. Ecol. 95. https://doi.org/10.1093/femsec/fiy183

Bier, R.L., Wernegreen, J.J., Vilgalys, R.J., Christopher Ellis, J., Bernhardt, E.S., Grossart, H.-P., Massana, R., McMahon, K., Walsh, D.A., 2020. Subsidized or stressed? Shifts in freshwater benthic microbial metagenomics along a gradient of alkaline coal mine drainage. Limnology and Oceanography 65, S277–S292. https://doi.org/10.1002/LNO.11301

Binh, C.T.T., Tong, T., Gaillard, J.-F., Gray, K.A., Kelly, J.J., 2014. Acute effects of TiO2 nanomaterials on the viability and taxonomic composition of aquatic bacterial communities assessed via high-throughput screening and next generation sequencing. PLoS One 9, e106280. https://doi.org/10.1371/journal.pone.0106280

Birko, S., Dove, E.S., Özdemir, V., 2015. A Delphi Technology Foresight Study: Mapping Social Construction of Scientific Evidence on Metagenomics Tests for Water Safety. PLoS One 10, e0129706. https://doi.org/10.1371/journal.pone.0129706

Bizic, M., Ionescu, D., Karnatak, R., Musseau, C.L., Onandia, G., Berger, S.A., Nejstgaard, J.C., Lischeid, G., Gessner, M.O., Wollrab, S., Grossart, H. -P., 2022. Land-use type temporarily affects active pond community structure but not gene expression patterns. Molecular Ecology. https://doi.org/10.1111/MEC.16348

Bofill-Mas, S., Albinana-Gimenez, N., Clemente-Casares, P., Hundesa, A., Rodriguez-Manzano, J., Allard, A., Calvo, M., Girones, R., 2006. Quantification and stability of human adenoviruses and polyomavirus JCPyV in wastewater matrices. Applied and environmental microbiology 72, 7894–7896.

Bokulich, N.A., Mills, D.A., 2013. Improved selection of internal transcribed spacer-specific primers enables quantitative, ultra-high-throughput profiling of fungal communities. Appl. Environ. Microbiol. 79, 2519–2526. https://doi.org/10.1128/AEM.03870-12

Bokulich, N.A., Ziemski, M., Robeson, M.S., Kaehler, B.D., 2020. Measuring the microbiome: Best practices for developing and benchmarking microbiomics methods. Computational and Structural Biotechnology Journal 18, 4048–4062. https://doi.org/10.1016/j.csbj.2020.11.049

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C., Al-Ghalith, G.A., Alexander, H., Alm, E.J., Arumugam, M. and Asnicar, F. 2018 QIIME 2: Reproducible, interactive, scalable, and extensible microbiome data science, PeerJ Preprints.

Bonk, F., Popp, D., Harms, H., Centler, F., 2018. PCR-based quantification of taxa-specific abundances in microbial communities: Quantifying and avoiding common pitfalls. Journal of microbiological methods 153, 139–147. https://doi.org/10.1016/J.MIMET.2018.09.015

Boonchan, M., Motomura, K., Inoue, K., Ode, H., Chu, P.Y., Lin, M., Iwatani, Y., Ruchusatsawat, K., Guntapong, R., Tacharoenmuang, R., Chantaroj, S., Tatsumi, M., Takeda, N., Sangkitporn, S., 2017. Distribution of norovirus genotypes and subtypes in river water by ultra-deep sequencing-based analysis. Lett Appl Microbiol 65, 98–104. https://doi.org/10.1111/lam.12750

Borchardt, M.A., Boehm, A.B., Salit, M., Spencer, S.K., Wigginton, K.R., Noble, R.T., 2021. The environmental microbiology minimum information (EMMI) guidelines: qPCR and dPCR quality and reporting for environmental microbiology. Environmental Science and Technology 55, 10210–10223. https://doi.org/10.1021/ACS.EST.1C01767/SUPPL_FILE/ES1C01767_SI_001.PDF

Bowers, R.M., Clum, A., Tice, H., Lim, J., Singh, K., Ciobanu, D., Ngan, C.Y., Cheng, J.-F., Tringe, S.G., Woyke, T., 2015. Impact of library preparation protocols and template quantity on the metagenomic reconstruction of a mock microbial community. BMC Genomics 16, 856. https://doi.org/10.1186/s12864-015-2063-6

Bowers, R.M., Kyrpides, N.C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T.B.K., Schulz, F., Jarett, J., Rivers, A.R., Eloe-Fadrosh, E.A., Tringe, S.G., Ivanova, N.N., Copeland, A., Clum, A., Becraft, E.D., Malmstrom, R.R., Birren, B., Podar, M., Bork, P., Weinstock, G.M., Garrity, G.M., Dodsworth, J.A., Yooseph, S., Sutton, G., Glöckner, F.O., Gilbert, J.A., Nelson, W.C., Hallam, S.J., Jungbluth, S.P., Ettema, T.J.G., Tighe, S., Konstantinidis, K.T., Liu, W.-T., Baker, B.J., Rattei, T., Eisen, J.A., Hedlund, B., McMahon, K.D., Fierer, N., Knight, R., Finn, R., Cochrane, G., Karsch-Mizrachi, I., Tyson, G.W., Rinke, C., Lapidus, A., Meyer, F., Yilmaz, P., Parks, D.H., Murat Eren, A., Schriml, L., Banfield, J.F., Hugenholtz, P., Woyke, T., 2017. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat Biotechnol 35, 725–731. https://doi.org/10.1038/nbt.3893

Bradley, I.M., Pinto, A.J., Guest, J.S., 2016. Design and Evaluation of Illumina MiSeq-Compatible, 18S rRNA Gene-Specific Primers for Improved Characterization of Mixed Phototrophic Communities. Appl. Environ. Microbiol. 82, 5878–5891. https://doi.org/10.1128/AEM.01630-16

Bradley, P., Gordon, N.C., Walker, T.M., Dunn, L., Heys, S., Huang, B., Earle, S., Pankhurst, L.J., Anson, L., De Cesare, M., Piazza, P., Votintseva, A.A., Golubchik, T., Wilson, D.J., Wyllie, D.H., Diel, R., Niemann, S., Feuerriegel, S., Kohl, T.A., Ismail, N., Omar, S. V., Smith, E.G., Buck, D., McVean, G., Walker, A.S., Peto, T.E.A., Crook, D.W., Iqbal, Z., 2015. Rapid antibiotic-resistance predictions from genome sequence data for Staphylococcus aureus and Mycobacterium tuberculosis. Nature Communications 6, 10063. https://doi.org/10.1038/ncomms10063

Brandt, C., Bongcam-Rudloff, E., Müller, B., 2020. Abundance Tracking by Long-Read Nanopore Sequencing of Complex Microbial Communities in Samples from 20 Different Biogas/Wastewater Plants. Applied Sciences 2020, Vol. 10, Page 7518 10, 7518–7518. https://doi.org/10.3390/APP10217518

Brandt, J., Albertsen, M., 2018. Investigation of Detection Limits and the Influence of DNA Extraction and Primer Choice on the Observed Microbial Communities in Drinking Water Samples Using 16S rRNA Gene Amplicon Sequencing. Front. Microbiol. 9, 2140. https://doi.org/10.3389/fmicb.2018.02140

Bréchet, C., Plantin, J., Sauget, M., Thouverez, M., Talon, D., Cholley, P., Guyeux, C., Hocquet, D., Bertrand, X., 2014. Wastewater treatment plants release large amounts of extended-spectrum β-lactamase-producing *Escherichia coli* into the environment. Clinical Infectious Diseases 58, 1658–1665. https://doi.org/10.1093/cid/ciu190

Breitbart, M., Salamon, P., Andresen, B., Mahaffy, J.M., Segall, A.M., Mead, D., Azam, F., Rohwer, F., 2002. Genomic analysis of uncultured marine viral communities. Proc Natl Acad Sci U A 99, 14250–14255. https://doi.org/10.1073/pnas.202488399

Brinkman, N.E., Fout, G.S., Keely, S.P., 2017. Retrospective Surveillance of Wastewater To Examine Seasonal Dynamics of Enterovirus Infections. mSphere 2. https://doi.org/10.1128/mSphere.00099-17

Brinkman, N.E., Villegas, E.N., Garland, J.L., Keely, S.P., 2018. Reducing inherent biases introduced during DNA viral metagenome analyses of municipal wastewater. PLoS One 13, e0195350. https://doi.org/10.1371/journal.pone.0195350

Brown, C.L., Keenum, I.M., Dai, D., Zhang, L., Vikesland, P.J., Pruden, A., 2021. Critical evaluation of short, long, and hybrid assembly for contextual analysis of antibiotic resistance genes in complex environmental metagenomes. Scientific Reports 11, 3753. https://doi.org/10.1038/s41598-021-83081-8

Brown C.L., Mullet J., Hindi F., Stoll J.E., Gupta S., Choi M., Keenum I., Vikesland P., Pruden A., Zhang L., 2022. mobileOG-db: a Manually Curated Database of Protein Families Mediating the Life Cycle of Bacterial Mobile Genetic Elements. Appl Environ Microbiol. 29:e0099122. doi: 10.1128/aem.00991-22. Epub ahead of print.

Brumfield, K.D., Hasan, N.A., Leddy, M.B., Cotruvo, J.A., Rashed, S.M., Colwell, R.R., Huq, A., 2020. A comparative analysis of drinking water employing metagenomics. PLoS One 15, e0231210. https://doi.org/10.1371/journal.pone.0231210

Buchfink, B., Reuter, K., Drost, H.-G., 2021. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat Methods 18, 366–368. https://doi.org/10.1038/s41592-021-01101-x

Buchfink, B., Xie, C., Huson, D.H., 2014. Fast and sensitive protein alignment using DIAMOND. Nature Methods 12, 59–60. https://doi.org/10.1038/nmeth.3176

Burke, C., Darling, A., 2016. A method for high precision sequencing of near full-length 16S rRNA genes on an Illumina MiSeq. PeerJ. 4:e2492. https://doi.org/10.7717/peerj.2492.

Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L., Vandesompele, J., Wittwer, C.T., 2009. The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments. Clinical Chemistry 55, 611–622. https://doi.org/10.1373/clinchem.2008.112797

Byrd, J.J., Xu, H.S., Colwell, R.R., 1991. Viable but nonculturable bacteria in drinking water. Appl. Environ. Microbiol. 57, 875–878. https://doi.org/10.1128/aem.57.3.875-878.1991

Cacace, D., Fatta-Kassinos, D., Manaia, C.M., Cytryn, E., Kreuzinger, N., Rizzo, L., Karaolia, P., Schwartz, T., Alexander, J., Merlin, C., Garelick, H., Schmitt, H., de Vries, D., Schwermer, C.U., Meric, S., Ozkal, C.B., Pons, M.N., Kneis, D., Berendonk, T.U., 2019. Antibiotic resistance genes in treated wastewater and in the receiving water bodies: A pan-European survey of urban settings. Water Research 162, 320–330. https://doi.org/10.1016/j.watres.2019.06.039

Cai, L., Yu, K., Yang, Y., Chen, B.-W., Li, X.-D., Zhang, T., 2013. Metagenomic exploration reveals high levels of microbial arsenic metabolism genes in activated sludge and coastal sediments. Appl. Microbiol. Biotechnol. 97, 9579–9588. https://doi.org/10.1007/s00253-012-4678-8

Cai, L., Zhang, T., 2013. Detecting human bacterial pathogens in wastewater treatment plants by a high-throughput shotgun sequencing technique. Environ. Sci. Technol. 47, 5433–5441. https://doi.org/10.1021/es400275r

Cai, M., Wilkins, D., Chen, J., Ng, S.-K., Lu, H., Jia, Y., Lee, P.K.H., 2016. Metagenomic Reconstruction of Key Anaerobic Digestion Pathways in Municipal Sludge and Industrial Wastewater Biogas-Producing Systems. Front. Microbiol. 7, 778. https://doi.org/10.3389/fmicb.2016.00778

Cai, P., Ning, Z., Zhang, N., Zhang, M., Guo, C., Niu, M., Shi, J., 2019. Insights into Biodegradation Related Metabolism in an Abnormally Low Dissolved Inorganic Carbon (DIC) Petroleum-Contaminated Aquifer by Metagenomics Analysis. Microorganisms 7, 412. https://doi.org/10.3390/microorganisms7100412

Callahan, B.J., McMurdie, P.J., Holmes, S.P., 2017. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. ISME J 11, 2639–2643. https://doi.org/10.1038/ismej.2017.119

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., Holmes, S.P., 2016. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods 13, 581–583. https://doi.org/10.1038/nmeth.3869

Cameron, A., Barbieri, R., Read, R., Church, D., Adator, E.H., Zaheer, R., McAllister, T.A., 2019. Functional screening for triclosan resistance in a wastewater metagenome and isolates of Escherichia coli and Enterococcus spp. from a large Canadian healthcare region. PLOS ONE 14, e0211144. https://doi.org/10.1371/journal.pone.0211144

Cantalupo, P.G., Calgua, B., Zhao, G., Hundesa, A., Wier, A.D., Katz, J.P., Grabe, M., Hendrix, R.W., Girones, R., Wang, D., Pipas, J.M., 2011. Raw sewage harbors diverse viral populations. MBio 2. https://doi.org/10.1128/mBio.00180-11

Capo, E., Giguet-Covex, C., Rouillard, A., Nota, K., Heintzman, P.D., Vuillemin, A., Ariztegui, D., Arnaud, F., Belle, S., Bertilsson, S., Bigler, C., Bindler, R., Brown, A.G., Clarke, C.L., Crump, S.E., Debroas, D., Englund, G., Ficetola, G.F., Garner, R.E., Gauthier, J., Gregory-Eaves, I., Heinecke, L., Herzschuh, U., Ibrahim, A., Kisand, V., Kjær, K.H., Lammers, Y., Littlefair, J., Messager, E., Monchamp, M.E., Olajos, F., Orsi, W., Pedersen, M.W., Rijal, D.P., Rydberg, J., Spanbauer, T., Stoof-Leichsenring, K.R., Taberlet, P., Talas, L., Thomas, C., Walsh, D.A., Wang, Y., Willerslev, E., van Woerkom, A., Zimmermann, H.H., Coolen, M.J.L., Epp, L.S., Domaizon, I., Alsos, I.G., Parducci, L., 2021. Lake sedimentary DNA research on past terrestrial and aquatic biodiversity: overview and recommendations. Quaternary 2021, Vol. 4, Page 6 4, 6.

https://doi.org/10.3390/QUAT4010006

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J., Knight, R., 2010. QIIME allows analysis of high-throughput community sequencing data. Nat. Methods 7, 335. https://doi.org/10.1038/nmeth.f.303

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Lozupone, C.A., Turnbaugh, P.J., Fierer, N., Knight, R., 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc. Natl. Acad. Sci. U. S. A. 108 Suppl 1, 4516–4522. https://doi.org/10.1073/pnas.1000080107

Carvalhais, L.C., Dennis, P.G., Tyson, G.W., Schenk, P.M., 2012. Application of metatranscriptomics to soil environments. J. Microbiol. Methods 91, 246–251. https://doi.org/10.1016/j.mimet.2012.08.011

Carvalhais, L.C., Schenk, P.M., 2013. Sample processing and cDNA preparation for microbial metatranscriptomics in complex soil communities. Methods in Enzymology 531, 251–267. https://doi.org/10.1016/B978-0-12-407863-5.00013-7

Caspi, R., Billington R., Keseler, I.M., Kothari, A., Krummenacker, M., Midford, P.E., Ong, W.K., Paley, S., Subhraveti, P., Karp, P.D., 2020. The MetaCyc database of metabolic pathways and enzymes - a 2019 update. Nucleic Acids Res. 8(48), D445-D453. doi: 10.1093/nar/gkz862.

Cassell, K., Gacek, P., Warren, J.L., Raymond, P.A., Cartter, M. and Weinberger, D.M. 2018. Association between sporadic legionellosis and river systems in Connecticut. The Journal of infectious diseases 217(2), 179-187.

Center for Disease Control and Prevention (CDC). 2019. Infectious Disease After a Disaster. https://www.cdc.gov/disasters/disease/infectious.html

Center for Advancing Microbial Risk Assessment. QMRAwiki. 2015. Michigan State University. http://qmrawiki.org (Accessed July 30, 2021)

Chao, Y., Ma, L., Yang, Y., Ju, F., Zhang, X.-X., Wu, W.-M., Zhang, T., 2013. Metagenomic analysis reveals significant changes of microbial compositions and protective functions during drinking water treatment. Sci Rep 3, 3550. https://doi.org/10.1038/srep03550

Chao, Y., Mao, Y., Wang, Z., Zhang, T., 2015. Diversity and functions of bacterial community in drinking water biofilms revealed by high-throughput sequencing. Sci Rep 5, 10044. https://doi.org/10.1038/srep10044

Chao, Y.; Mao, Y.; Yu, K.; Zhang, T. Novel Nitrifiers and Comammox in a Full-Scale Hybrid Biofilm and Activated Sludge Reactor Revealed by Metagenomic Approach. Appl. Microbiol. Biotechnol. 2016, 100 (18), 8225–8237. https://doi.org/10.1007/s00253-016-7655-9.

Chaudhry, R.M., Nelson, K.L., Drewes, J.E., 2015. Mechanisms of pathogenic virus removal in a full-scale membrane bioreactor. Environmental science & technology 49, 2815–2822.

Chaumeil, P.-A., Mussig, A.J., Hugenholtz, P., Parks, D.H., 2020. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics 36, 1925–1927. https://doi.org/10.1093/bioinformatics/btz848

Che, Y., Xia, Y., Liu, L., Li, A.-D., Yang, Y., Zhang, T., 2019. Mobile antibiotic resistome in wastewater treatment plants revealed by Nanopore metagenomic sequencing. Microbiome 7, 44. https://doi.org/10.1186/s40168-019-0663-0

Chen, S., Wang, F., Beaulieu, J.C., Stein, R.E., Ge, B., 2011. Rapid detection of viable salmonellae in produce by coupling propidium monoazide with loop-mediated isothermal amplification. Appl Env. Microbiol 77, 4008–4016. https://doi.org/10.1128/AEM.00354-11

Chen, X., Lang, X.L., Xu, A.L., Song, Z.W., Yang, J., Guo, M.Y., 2019. Seasonal variability in the microbial community and pathogens in wastewater final Effluents. Water 11. https://doi.org/10.3390/w11122586

Chen, Z., Zhang, J., Li, R., Tian, F., Shen, Y., Xie, X., Ge, Q., Lu, Z., 2018. Metatranscriptomics analysis of cyanobacterial aggregates during cyanobacterial bloom period in Lake Taihu, China. Env. Sci Pollut Res Int 25, 4811–4825. https://doi.org/10.1007/s11356-017-0733-4

Chiao, T.-H., Clancy, T.M., Pinto, A., Xi, C., Raskin, L., 2014. Differential resistance of drinking water bacterial populations to monochloramine disinfection. Env. Sci Technol 48, 4038–4047. https://doi.org/10.1021/es4055725

Chopyk, J., Nasko, D.J., Allard, S., Callahan, M.T., Bui, A., Ferelli, A.M.C., Chattopadhyay, S., Mongodin, E.F., Pop, M., Micallef, S.A., Sapkota, A.R., 2020. Metagenomic analysis of bacterial and viral assemblages from a freshwater creek and irrigated field reveals temporal and spatial dynamics. Science of The Total Environment 706, 135395. https://doi.org/10.1016/J.SCITOTENV.2019.135395

Christgen, B., Yang, Y., Ahammad, S., Li, B., Rodriquez, D.C., Zhang, T., Graham, D.W., 2015. Metagenomics shows that low-energy anaerobic− aerobic treatment reactors reduce antibiotic resistance gene levels from domestic wastewater. Environ. Sci. Technol. 49, 2577–2584.

Chu, B.T.T., Petrovich, M.L., Chaudhary, A., Wright, D., Murphy, B., Wells, G., Poretsky, R., 2017. Metagenomics Reveals the Impact of Wastewater Treatment Plants on the Dispersal of Microorganisms and Genes in Aquatic Sediments. Appl. Environ. Microbiol. 84. https://doi.org/10.1128/aem.02168-17

Cleary, D.F.R., Polónia, A.R.M., de Voogd, N.J., 2018. Prokaryote composition and predicted metagenomic content of two Cinachyrella Morphospecies and water from West Papuan Marine Lakes. FEMS Microbiol. Ecol. 94. https://doi.org/10.1093/femsec/fix175

Coenen-Stass, A.M.L., Magen, I., Brooks, T., Ben-Dov, I.Z., Greensmith, L., Hornstein, E., Fratta, P., 2018. Evaluation of methodologies for microRNA biomarker detection by next generation sequencing. RNA Biol 15, 1133–1145. https://doi.org/10.1080/15476286.2018.1514236

Cole, J.R., Wang, Q., Fish, J.A., Chai, B., McGarrell, D.M., Sun, Y., Brown, C.T., Porras-Alfaro, A., Kuske, C.R., Tiedje, J.M., 2014. Ribosomal Database Project: data and tools for high throughput rRNA analysis. Nucleic Acids Res 42, D633-642. https://doi.org/10.1093/nar/gkt1244

Corpuz, M.V.A., Buonerba, A., Vigliotta, G., Zarra, T., Ballesteros, F., Campiglia, P., Belgiorno, V., Korshin, G., Naddeo, V., 2020. Viruses in wastewater: occurrence, abundance and detection methods. Science of The Total Environment 745, 140910. https://doi.org/10.1016/j.scitotenv.2020.140910

Coryell, M.P., Iakiviak, M., Pereira, N., Murugkar, P.P., Rippe, J., Williams, D.B., Heald-Sargent, T., Sanchez-Pinto, L.N., Chavez, J., Hastie, J.L., Sava, R.L., Lien, C.Z., Wang, T.T., Muller, W.J., Fischbach, M.A., Carlson, P.E., 2021. A method for detection of SARS-CoV-2 RNA in healthy human stool: a validation study. The Lancet Microbe 2, e259–e266. https://doi.org/10.1016/S2666-5247(21)00059-8

Costa, P.S., Reis, M.P., Ávila, M.P., Leite, L.R., de Araújo, F.M.G., Salim, A.C.M., Oliveira, G., Barbosa, F., Chartone-Souza, E., Nascimento, A.M.A., 2015. Metagenome of a Microbial Community Inhabiting a Metal-Rich Tropical Stream Sediment. PLOS ONE 10, e0119465–e0119465. https://doi.org/10.1371/journal.pone.0119465

Crossette, E., Gumm, J., Langenfeld, K., Raskin, L., Duhaime, M., Wigginton, K., 2021. Metagenomic Quantification of Genes with Internal Standards. mBio 12. https://doi.org/10.1128/mBio.03173-20

Crovadore, J., Soljan, V., Calmin, G., Chablais, R., Cochard, B., Lefort, F., 2017. Metatranscriptomic and metagenomic description of the bacterial nitrogen metabolism in waste water wet oxidation effluents. Heliyon 3, e00427. https://doi.org/10.1016/j.heliyon.2017.e00427

Cui, Q., Fang, T., Huang, Y., Dong, P., Wang, H., 2017. Evaluation of bacterial pathogen diversity, abundance and health risks in urban recreational water by amplicon next-generation sequencing and quantitative PCR. J Env. Sci 57, 137–149. https://doi.org/10.1016/j.jes.2016.11.008

Cui, Q., Huang, Y., Wang, H., Fang, T., 2019. Diversity and abundance of bacterial pathogens in urban rivers impacted by domestic sewage. Env. Pollut 249, 24–35. https://doi.org/10.1016/j.envpol.2019.02.094

Cydzik-Kwiatkowska, A., Wnuk, M., 2011. Optimization of activated sludge storage before RNA isolation. bta 1, 101–105. https://doi.org/10.5114/bta.2011.46522

D’Costa, V.M., King, C.E., Kalan, L., Morar, M., Sung, W.W.L., Schwarz, C., Froese, D., Zazula, G., Calmels, F., Debruyne, R., Brian Golding, G., Poinar, H.N., Wright, G.D., 2011. Antibiotic resistance is ancient. Nature 477, 457–461. https://doi.org/10.1038/nature10388

Dai, D., Proctor, C.R., Williams, K., Edwards, M.A., Pruden, A., 2018a. Mediation of effects of biofiltration on bacterial regrowth, Legionella pneumophila, and the microbial community structure under hot water plumbing conditions. Environ. Sci. Water Res. Technol. 4, 183–194. https://doi.org/10.1039/c7ew00301c

Dai, D., Rhoads, W.J., Edwards, M.A., Pruden, A., 2018b. Shotgun Metagenomics Reveals Taxonomic and Functional Shifts in Hot Water Microbiome Due to Temperature Setting and Stagnation. Front Microbiol 9, 2695. https://doi.org/10.3389/fmicb.2018.02695

Das, S., Bora, S.S., Yadav, R.N.S., Barooah, M., 2017. A metagenomic approach to decipher the indigenous microbial communities of arsenic contaminated groundwater of Assam. Genom Data 12, 89–96. https://doi.org/10.1016/j.gdata.2017.03.013

Davenport, E.J., Neudeck, M.J., Matson, P.G., Bullerjahn, G.S., Davis, T.W., Wilhelm, S.W., Denney, M.K., Krausfeldt, L.E., Stough, J.M.A., Meyer, K.A., Dick, G.J., Johengen, T.H., Lindquist, E., Tringe, S.G., McKay, R.M.L., 2019. Metatranscriptomic Analyses of Diel Metabolic Functions During a Microcystis Bloom in Western Lake Erie (United States). Front. Microbiol. 10. https://doi.org/10.3389/fmicb.2019.02081

David, S., Afshar, B., Mentasti, M., Ginevra, C., Podglajen, I., Harris, S.R., Chalker, V.J., Jarraud, S., Harrison, T.G., Parkhill, J., 2017. Clinical Infectious Diseases Seeding and Establishment of Legionella pneumophila in Hospitals: Implications for Genomic Investigations of Nosocomial Legionnaires’ Disease. https://doi.org/10.1093/cid/cix153

Davis, B.C.; Calarco, J.; Liguori, K.; Milligan, E.G.; Brown, C.L.; Gupta, S.; Harwood, V.J.; Pruden, A.; Keenum, I.M. 2023. Recommendations for the use of Metagenomics for Routine Monitoring of Antibiotic Resistance in Wastewater and Impacted Aquatic Environments. Current Opinion in Environmental Science and Technology https://doi.org/10.1080/10643389.2023.2181620

Davis, B.C., Riquelme, M.V., Ramirez-toro, G., Bandaragoda, C., Garner, E., Rhoads, W.J., Vikesland, P., Pruden, A., 2020a. Demonstrating an Integrated Antibiotic Resistance Gene Surveillance Approach in Puerto Rican Watersheds Post-Hurricane Maria. Environ. Sci. Technol. https://doi.org/10.1021/acs.est.0c05567

Davis, J.J., Wattam, A.R., Aziz, R.K., Brettin, T., Butler, R., Butler, R.M., Chlenski, P., Conrad, N., Dickerman, A., Dietrich, E.M., Gabbard, J.L., Gerdes, S., Guard, A., Kenyon, R.W., Machi, D., Mao, C., Murphy-Olson, D., Nguyen, M., Nordberg, E.K., Olsen, G.J., Olson, R.D., Overbeek, J.C., Overbeek, R., Parrello, B., Pusch, G.D., Shukla, M., Thomas, C., VanOeffelen, M., Vonstein, V., Warren, A.S., Xia, F., Xie, D., Yoo, H., Stevens, R., 2020b. The PATRIC Bioinformatics Resource

Center: expanding data and analysis capabilities. Nucleic Acids Res. 48, D606–D612. https://doi.org/10.1093/nar/gkz943

Debroas, D., Humbert, J.-F., Enault, F., Bronner, G., Faubladier, M., Cornillot, E., 2009. Metagenomic approach studying the taxonomic and functional diversity of the bacterial community in a mesotrophic lake (Lac du Bourget - France). Environ. Microbiol. 11, 2412–2424. https://doi.org/10.1111/j.1462-2920.2009.01969.x

Delforno, T P, Lacerda, G.V., Jr, Sierra-Garcia, I.N., Okada, D.Y., Macedo, T.Z., Varesche, M.B.A., Oliveira, V.M., 2017bF. Metagenomic analysis of the microbiome in three different bioreactor configurations applied to commercial laundry wastewater treatment. Sci Total Env. 587–588, 389–398. https://doi.org/10.1016/j.scitotenv.2017.02.170

Delforno, T.P., Júnior, G.V.L., Noronha, M.F., Sakamoto, I.K., Varesche, M.B.A., Oliveira, V.M., 2017a. Microbial diversity of a full-scale UASB reactor applied to poultry slaughterhouse wastewater treatment: integration of 16S rRNA gene amplicon and shotgun metagenomic sequencing. MicrobiologyOpen 6, e00443. https://doi.org/10.1002/mbo3.443

Delforno, T.P., Macedo, T.Z., Midoux, C., Lacerda, G.V., Jr, Rué, O., Mariadassou, M., Loux, V., Varesche, M.B.A., Bouchez, T., Bize, A., Oliveira, V.M., 2019. Comparative metatranscriptomic analysis of anaerobic digesters treating anionic surfactant contaminated wastewater. Sci. Total Environ. 649, 482–494. https://doi.org/10.1016/j.scitotenv.2018.08.328

DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P., Andersen, G.L., 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Env. Microbiol 72, 5069–5072. https://doi.org/10.1128/AEM.03006-05

Divizia, M., Cencioni, B., Palombi, L. and Panà, A. 2008. Sewage workers: risk of acquiring enteric virus infections including hepatitis A. New Microbiologica 31 (3): 337–41.

Djikeng, A., Kuzmickas, R., Anderson, N.G., Spiro, D.J., 2009. Metagenomic Analysis of RNA Viruses in a Fresh Water Lake. PLoS ONE 4, e7264. https://doi.org/10.1371/journal.pone.0007264

Djurhuus, A., Port, J., Closek, C.J., Yamahara, K.M., Romero-Maraccini, O., Walz, K.R., Goldsmith, D.B., Michisaki, R., Breitbart, M., Boehm, A.B., Chavez, F.P., 2017. Evaluation of Filtration and DNA Extraction Methods for Environmental DNA Biodiversity Assessments across Multiple Trophic Levels. Front. Mar. Sci. 4. https://doi.org/10.3389/fmars.2017.00314

DNA Sequencing Costs: Data, 2018. URL https://www.genome.gov/27541954/dna-sequencing-costs-data/ (accessed 12.13.18).

Douarre, P.-E., Mallet, L., Radomski, N., Felten, A., Mistou, M.-Y., 2020. Analysis of COMPASS, a New Comprehensive Plasmid Database Revealed Prevalence of Multireplicon and Extensive Diversity of IncF Plasmids. Front Microbiol 11, 483. https://doi.org/10.3389/fmicb.2020.00483

Douterelo, I., Calero-Preciado, C., Soria-Carrasco, V., Boxall, J.B., 2018. Whole metagenome sequencing of chlorinated drinking water distribution systems. Env. Sci Water Res Technol 4, 2080–2091. https://doi.org/10.1039/C8EW00395E

Driscoll, C.B., Otten, T.G., Brown, N.M., Dreher, T.W., 2017. Towards long-read metagenomics: complete assembly of three novel genomes from bacteria dependent on a diazotrophic cyanobacterium in a freshwater lake co-culture. Standards in Genomic Sciences 12. https://doi.org/10.1186/s40793-017-0224-8

Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26, 2460–2461. https://doi.org/10.1093/bioinformatics/btq461

Edgar, R.C., 2013. UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat Methods 10, 996–998. https://doi.org/10.1038/nmeth.2604

Edwardson, C.F., Hollibaugh, J.T., 2017. Metatranscriptomic analysis of prokaryotic communities active in sulfur and arsenic cycling in Mono Lake, California, USA. ISME J 11, 2195–2208. https://doi.org/10.1038/ismej.2017.80

Eiler, A., Drakare, S., Bertilsson, S., Pernthaler, J., Peura, S., Rofner, C., Simek, K., Yang, Y., Znachor, P., Lindström, E.S., 2013. Unveiling distribution patterns of freshwater phytoplankton by a next generation sequencing based approach. PLoS One 8, e53516. https://doi.org/10.1371/journal.pone.0053516

Eisenhofer, R., Minich, J.J., Marotz, C., Cooper, A., Knight, R., Weyrich, L.S., 2019. Contamination in Low Microbial Biomass Microbiome Studies: Issues and Recommendations. Trends in microbiology 27, 105–117. https://doi.org/10.1016/J.TIM.2018.11.003

Ekwanzala, M.D., Dewar, J.B., Kamika, I., Momba, M.N.B., 2019. Tracking the environmental dissemination of carbapenem-resistant Klebsiella pneumoniae using whole genome sequencing. Science of The Total Environment 691, 80–92. https://doi.org/10.1016/j.scitotenv.2019.06.533

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., Qureshi, M., Richardson, L.J., Salazar, G.A., Smart, A., Sonnhammer, E.L.L., Hirsh, L., Paladin, L. Piovesan, D., Tosatto, S.C.E., Finn, R.D., 2019. The Pfam protein families database in 2019: Nucleic Acids Research. doi: 10.1093/nar/gky995

Emmanuel, S.A., Sul, W.J., Seong, H.J., Rhee, C., Ekpheghere, K.I., Kim, I.-S., Kim, H.-G., Koh, S.-C., 2019. Metagenomic analysis of relationships between the denitrification process and carbon metabolism in a bioaugmented full-scale tannery wastewater treatment plant. World J. Microbiol. Biotechnol. 35. https://doi.org/10.1007/s11274-019-2716-8

EPA, 2014a. Sampling and Analysis Plan Guidance and Template, version 4. https://www.epa.gov/quality/sampling-and-analysis-plan-guidance-and-template-v4-general-projects-042014

EPA, 2014b. Method 1680: fecal coliforms in sewage sludge (biosolids) by multiple-tube fermentation using lauryl tryptose broth (LTB) and EC medium. Washington, D.C.

EPA, 2016. Quick Guide To Drinking Water Sample Collection 2nd Edition. Golden, CO. https://www.epa.gov/sites/default/files/2015-11/documents/drinking_water_sample_collection.pdf

EPA, 2017a. Sample collection information document for pathogens companion to selected analytical methods for environmental remediation and recovery (SAM) 2017. Cincinnati, OH.

EPA, 2017b. Procedures for collecting wastewater samples, US Environmental Protection Agency.

Esteve-Codina, A., 2018. Rna-seq data analysis, applications and challenges. Compr Anal Chem. 82, 71-106.

Fahrenfeld, N.L., Delos Reyes, H., Eramo, A., Akob, D.M., Mumford, A.C., Cozzarelli, I.M., 2017. Shifts in microbial community structure and function in surface waters impacted by unconventional oil and gas wastewater revealed by metagenomics. Sci Total Env. 580, 1205–1213. https://doi.org/10.1016/j.scitotenv.2016.12.079

Falkinham, J.O., Pruden, A., Edwards, M., 2015. Opportunistic premise plumbing pathogens: Increasingly important pathogens in drinking water. Pathogens 4, 373–386. https://doi.org/10.3390/pathogens4020373

Fang, H., Cai, L., Yang, Y., Ju, F., Li, X., Yu, Y., Zhang, T., 2014. Metagenomic analysis reveals potential biodegradation pathways of persistent pesticides in freshwater and marine sediments. Sci. Total Environ. 470–471, 983–992. https://doi.org/10.1016/j.scitotenv.2013.10.076

Fang, H., Cai, L., Yu, Y., Zhang, T., 2013. Metagenomic analysis reveals the prevalence of biodegradation genes for organic pollutants in activated sludge. Bioresour. Technol. 129, 209–218. https://doi.org/10.1016/j.biortech.2012.11.054

Fang, H., Zhang, H., Han, L., Mei, J., Ge, Q., Long, Z., Yu, Y., 2018a. Exploring bacterial communities and biodegradation genes in activated sludge from pesticide wastewater treatment plants via metagenomic analysis. Env. Pollut 243, 1206–1216. https://doi.org/10.1016/j.envpol.2018.09.080

Fang, T., Cui, Q., Huang, Y., Dong, P., Wang, H., Liu, W.T., Ye, Q., 2018b. Distribution comparison and risk assessment of free-floating and particle-attached bacterial pathogens in urban recreational water: Implications for water quality management. Sci. Total Environ. 613–614, 428–438. https://doi.org/10.1016/j.scitotenv.2017.09.008

Farhat, M., Shaheed, R.A., Al-Ali, H.H., Al-Ghamdi, A.S., Al-Hamaqi, G.M., Maan, H.S., Al-Mahfoodh, Z.A., Al-Seba, H.Z., 2018. Legionella confirmation in cooling tower water:

Comparison of culture, real-time PCR and next generation sequencing. Saudi Med J 39, 137–141. https://doi.org/10.15537/smj.2018.2.21587

Fernandez-Cassi, X., Timoneda, N., Martínez-Puchol, S., Rusiñol, M., Rodriguez-Manzano, J., Figuerola, N., Bofill-Mas, S., Abril, J.F., Girones, R., 2018. Metagenomics for the study of viruses in urban sewage as a tool for public health surveillance. Sci. Total Environ. 618, 870–880. https://doi.org/10.1016/j.scitotenv.2017.08.249

Fitzhenry, R., Weiss, D., Cimini, D., Balter, S., Boyd, C., Alleyne, L., Stewart, R., McIntosh, N., Econome, A., Lin, Y., Rubinstein, I., Passaretti, T., Kidney, A., Lapierre, P., Kass, D., Varma, J.K., 2017. Legionnaires' disease outbreaks and cooling towers, New York City, New York, USA. Emerg Infect Dis 23, 1769–1776. https://doi.org/10.3201/eid2311.161584

Flekna, G., Stefanic, P., Wagner, M., Smulders, F.J.M., Mozina, S.S., Hein, I., 2007. Insufficient differentiation of live and dead Campylobacter jejuni and Listeria monocytogenes cells by ethidium monoazide (EMA) compromises EMA/real-time PCR. Res Microbiol 158, 405–412. https://doi.org/10.1016/j.resmic.2007.02.008

Fleres, G., Couto, N., Lokate, M., van der Sluis, L.W.M., Ginevra, C., Jarraud, S., Deurenberg, R.H., Rossen, J.W., García-Cobos, S., Friedrich, A.W., 2018. Detection of Legionella Anisa in Water from Hospital Dental Chair Units and Molecular Characterization by Whole-Genome Sequencing. Microorganisms 6. https://doi.org/10.3390/microorganisms6030071

Fodelianakis, S., Washburne, A.D., Bourquin, M., Pramateftaki, P., Kohler, T.J., Styllas, M., Tolosano, M., De Staercke, V., Schön, M., Busi, S.B., Brandani, J., Wilmes, P., Peter, H., Battin, T.J., 2021. Microdiversity characterizes prevalent phylogenetic clades in the glacier-fed stream microbiome. ISME Journal. https://doi.org/10.1038/s41396-021-01106-6

Folch-Mallol, J.L., Zárate, A., Sánchez-Reyes, A., López-Lara, I.M., 2019. Expression, purification, and characterization of a metagenomic thioesterase from activated sludge involved in the degradation of acylCoA-derivatives. Protein Expr. Purif. 159, 49–52. https://doi.org/10.1016/j.pep.2019.03.008

Forsberg, K.J., Patel, S., Gibson, M.K., Lauber, C.L., Knight, R., Fierer, N., Dantas, G., 2014. Bacterial phylogeny structures soil resistomes across habitats. Nature 509, 612–616. https://doi.org/10.1038/nature13377

Franzosa, E.A., McIver, L.J., Rahnavard, G., Thompson, L.R., Schirmer, M., Weingart, G., Lipson, K.S., Knight, R., Caporaso, J.G., Segata, N., Huttenhower, C., 2018. Species-level functional profiling of metagenomes and metatranscriptomes. Nature Methods 15, 962–968. https://doi.org/10.1038/s41592-018-0176-y

Fróes, A.M., da Mota, F.F., Cuadrat, R.R.C., Dávila, A.M.R., 2016. Distribution and Classification of Serine β-Lactamases in Brazilian Hospital Sewage and Other Environmental Metagenomes Deposited in Public Databases. Front. Microbiol. 7. https://doi.org/10.3389/fmicb.2016.01790

Fukasawa, Y., Ermini, L., Wang, H., Carty, K., Cheung, M.S., 2020. LongQC: A Quality Control Tool for Third Generation Sequencing Long Read Data. G3. 10, 1193-1196. https://doi.org/10.1534/g3.119.400864

Fumian, T.M., Fioretti, J.M., Lun, J.H., Dos Santos, I.A.L., White, P.A., Miagostovich, M.P., 2019. Detection of norovirus epidemic genotypes in raw sewage using next generation sequencing. Env. Int 123, 282–291. https://doi.org/10.1016/j.envint.2018.11.054

Galperin, M.Y., Kristensen, D.M., Makarova, K.S., Wolf, Y.I., Koonin, E.V., 2019. Microbial genome analysis: the COG approach. Brief. Bioinform. 20, 1063–1070. https://doi.org/10.1093/bib/bbx117

Garcia-Vidal, C., Labori, M., Viasus, D., Simonetti, A., Garcia-Somoza, D., Dorca, J., Gudiol, F. and Carratalà, J. 2013. Rainfall is a risk factor for sporadic cases of Legionella pneumophila pneumonia. Plos One 8(4), e61036.

Garland, J.L., Ph, D., Brinkman, N., Ph, D., Jahne, M., Ph, D., 2019. Geospatial Distribution of Antimicrobial Resistance Genes in US Rivers and Streams United States Environmental Protection Agency.

Garner, E., Brown, C.L., Schwake, D.O., Rhoads, W.J., Arango-Argoty, G., Zhang, L., Jospin, G., Coil, D.A., Eisen, J.A., Edwards, M.A., Pruden, A., 2019a. Comparison of Whole-Genome Sequences of Legionella pneumophila in Tap Water and in Clinical Strains, Flint, Michigan, USA, 2016. Emerg. Infect. Dis. 25, 2013–2020. https://doi.org/10.3201/eid2511.181032

Garner, E., Chen, C., Xia, K., Bowers, J., Engelthaler, D.M., Mclain, J., Edwards, M.A., Pruden, A., 2018a. Metagenomic Characterization of Antibiotic Resistance Genes in Full- Scale Reclaimed Water Distribution Systems and Corresponding Potable Systems. Environ. Sci. Technol. 52, 6113–6125. https://doi.org/10.1021/acs.est.7b05419

Garner, E., Inyang, M., Garvey, E., Parks, J., Glover, C., Dickenson, E., Sutherland;, J., Salveson, A., Edwards, M.A., Pruden, A., 2019b. Impact of blending for direct potable reuse on premise plumbing microbial ecology and regrowth of opportunistic pathogens and antibiotic resistant bacteria. Water Res. 151, 75–86. https://doi.org/10.1016/j.watres.2018.12.003

Garner, E., McLain, J., Bowers, J., Engelthaler, D.M., Edwards, M.A., Pruden, A., 2018b. Microbial Ecology and Water Chemistry Impact Regrowth of Opportunistic Pathogens in Full-Scale Reclaimed Water Distribution Systems. Env. Sci Technol 52, 9056–9068. https://doi.org/10.1021/acs.est.8b02818

Garner, E., Wallace, J.S., Argoty, G.A., Wilkinson, C., Fahrenfeld, N., Heath, L.S., Zhang, L., Arabi, M., Aga, D.S., Pruden, A., 2016. Metagenomic profiling of historic Colorado Front Range flood impact on distribution of riverine antibiotic resistance genes. Sci Rep 6, 38432. https://doi.org/10.1038/srep38432

Garner, R.E., Gregory-Eaves, I., Walsh, D.A., 2020. Sediment metagenomes as time capsules of lake microbiomes. mSphere 5. https://doi.org/10.1128/MSPHERE.00512-20/SUPPL_FILE/MSPHERE.00512-20-ST005.PDF

Garner, E., B.C. Davis, E. Milligan, M. Blair, I. Keenum, A. Maile-Moskowitz, J. Pan, M. Gnegy, K. Liguori, S. Gupta, A.J. Prussin, L.C., Marr, L.S. Heath, P.J. Vikesland, L. Zhang, and A. Pruden. 2021. "Next Generation Sequencing Approaches to Evaluate Water and Wastewater Quality." *Water Research,* 194(April): 116907. https://doi.org/10.1016/j.watres.2021.116907

Garrido-Cardenas, J.A., Polo-López, M.I., Oller-Alberola, I., 2017. Advanced microbial analysis for wastewater quality monitoring: metagenomics trend. Appl. Microbiol. Biotechnol. 101, 7445–7458. https://doi.org/10.1007/s00253-017-8490-3

George, A., Gray, K., Wait, K., Kriss, R., Edwards, M. and Pieper, K. 2019. Citizen Scientists and University Researchers Assess Post-Hurricane Well Water Contamination in North Carolina Environmental Justice Communities. AGUFM 2019, ED32A-02.

Ghai, R., Mizuno, C.M., Picazo, A., Camacho, A., Rodriguez-Valera, F., 2014. Key roles for freshwater Actinobacteria revealed by deep metagenomic sequencing. Mol Ecol 23, 6073–6090. https://doi.org/10.1111/mec.12985

Ghosh, S., Zhu, N., Milligan, E., Falkinham III, J., Pruden, A., Edwards, M., 2021. Mapping the terrain for pathogen persistence and proliferation in reclaimed water distribution systems: Interactive effects of biofiltration, disinfection, and water age. Environ. Sci. Technol. 2021, 55, 18, 12561–12573.

Ghurye, J.S., Cepeda-Espinoza, V., Pop, M., 2016. Focus: Microbiome: Metagenomic Assembly: Overview, Challenges and Applications. Yale J Biol Med 89, 353.

Gibson, M.K., Forsberg, K.J., Dantas, G., 2015. Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. ISME J. 9, 207–216. https://doi.org/10.1038/ismej.2014.106

Gilbert, J.A., Jansson, J.K. and Knight, R. 2014. The Earth Microbiome project: successes and aspirations. BMC Biology 12(1), 69.

Gomez-Alvarez, V., Pfaller, S., Pressman, J.G., Wahman, D.G., Revetta, R.P., 2016. Resilience of microbial communities in a simulated drinking water distribution system subjected to disturbances: role of conditionally rare taxa and potential implications for antibiotic-resistant bacteria. Environ. Sci. Water Res. Technol. 2, 645–657. https://doi.org/10.1039/c6ew00053c

Gomez-Alvarez, V., Revetta, R.P., Santo Domingo, J.W., 2012. Metagenome analyses of corroded concrete wastewater pipe biofilms reveal a complex microbial system. BMC Microbiol 12, 122. https://doi.org/10.1186/1471-2180-12-122

Gong, C., Zhang, W., Zhou, X., Wang, H., Sun, G., Xiao, J., Pan, Y., Yan, S., Wang, Y., 2016. Novel Virophages Discovered in a Freshwater Lake in China. Front. Microbiol. 7.

https://doi.org/10.3389/fmicb.2016.00005

Gonzalez, R., Curtis, K., Bivins, A., Bibby, K., Weir, M.H., Yetka, K., Thompson, H., Keeling, D., Mitchell, J., Gonzalez, D., 2020. COVID-19 surveillance in Southeastern Virginia using wastewater-based epidemiology. Water Research 186, 116296. https://doi.org/10.1016/j.watres.2020.116296

Gonzales-Gustavson, E., Cárdenas-Youngs, Y., Calvo, M., Marques da Silva, M.F., Hundesa, A., Amorós, I., Moreno, Y., Moreno-Mesonero, L., Rosell, R., Ganges, L., Araujo, R., Girones, R., 2017. Characterization of the efficiency and uncertainty of skimmed milk flocculation for the simultaneous concentration and quantification of water-borne viruses, bacteria and protozoa. Journal of Microbiological Methods, 134:46-53, https://doi.org/10.1016/j.mimet.2017.01.006

Graham, K.E., Loeb, S.K., Wolfe, M.K., Catoe, D., Sinnott-Armstrong, N., Kim, S., Yamahara, K.M., Sassoubre, L.M., Mendoza Grijalva, L.M., Roldan-Hernandez, L., 2020. SARS-CoV-2 RNA in wastewater settled solids is associated with COVID-19 cases in a large urban sewershed. Environmental science & technology 55, 488–498.

Graham, R.M.A., Doyle, C.J., Jennison, A.V., 2014. Real-time investigation of a Legionella pneumophila outbreak using whole genome sequencing. Epidemiol Infect 142, 2347–2351. https://doi.org/10.1017/S0950268814000375

Greay, T.L., Gofton, A.W., Zahedi, A., Paparini, A., Linge, K.L., Joll, C.A., Ryan, U.M., 2019. Evaluation of 16S next-generation sequencing of hypervariable region 4 in wastewater samples: An unsuitable approach for bacterial enteric pathogen identification. Sci Total Env. 670, 1111–1124. https://doi.org/10.1016/j.scitotenv.2019.03.278

Green, J.C., Rahman, F., Saxton, M.A., Williamson, K.E., 2015. Metagenomic assessment of viral diversity in Lake Matoaka, a temperate, eutrophic freshwater lake in southeastern Virginia, USA. Aquat. Microb. Ecol. 75, 117–128. https://doi.org/10.3354/ame01752

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J., Knight, R., 2010. QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7, 335. https://doi.org/10.1038/nmeth.f.303

Gu, X., Tay, Q.X.M., Te, S.H., Saeidi, N., Goh, S.G., Kushmaro, A., Thompson, J.R., Gin, K.Y.-H., 2018. Geospatial distribution of viromes in tropical freshwater ecosystems. Water Res 137, 220–232. https://doi.org/10.1016/j.watres.2018.03.017

Guo, F., Ju, F., Cai, L., Zhang, T., 2013. Taxonomic Precision of Different Hypervariable Regions of 16S rRNA Gene and Annotation Methods for Functional Bacterial Groups in Biological Wastewater Treatment. PLoS One 8. https://doi.org/10.1371/journal.pone.0076185

Guo, F., Zhang, T., 2012. Profiling bulking and foaming bacteria in activated sludge by high throughput sequencing. Water Res 46, 2772–2782. https://doi.org/10.1016/j.watres.2012.02.039

Guo, F., Zhang, T., 2013. Biases during DNA extraction of activated sludge samples revealed by high throughput sequencing. Appl Microbiol Biotechnol 97, 4607–4616. https://doi.org/10.1007/s00253-012-4244-4

Guo, J., Li, J., Chen, H., Bond, P.L., Yuan, Z., 2017. Metagenomic analysis reveals wastewater treatment plants as hotspots of antibiotic resistance genes and mobile genetic elements. Water Res. 123, 468–478. https://doi.org/10.1016/j.watres.2017.07.002

Guo, J., Peng, Y., Ni, B.-J., Han, X., Fan, L., Yuan, Z., 2015. Dissecting microbial community structure and methane-producing pathways of a full-scale anaerobic reactor digesting activated sludge from wastewater treatment by metagenomic sequencing. Microb. Cell Factories 14. https://doi.org/10.1186/s12934-015-0218-4

Gupta, R.S., Lo, B., Son, J., 2018a. Phylogenomics and comparative genomic studies robustly support division of the genus Mycobacterium into an emended genus Mycobacterium and four novel genera. Frontiers in Microbiology 9. https://doi.org/10.3389/fmicb.2018.00067

Gupta, S.K., Shin, H., Han, D., Hur, H.-G., Unno, T., 2018b. Metagenomic analysis reveals the prevalence and persistence of antibiotic- and heavy metal-resistance genes in wastewater treatment plant. J. Microbiol. 56, 408–415. https://doi.org/10.1007/s12275-018-8195-z

Guy, R.A., Payment, P., Krull, U.J. and Horgen, P.A. 2003. Real-Time PCR for Quantification of Giardia and Cryptosporidium in Environmental Water Samples and Sewage. Appl Environ Microb 69, 5178-5185.

Gweon, H. Soon, Shaw, Liam P., Swann, Jeremy, De Maio, Nicola, AbuOun, M., Niehus, R., Hubbard, Alasdair T. M., Bowes, Mike J., Bailey, Mark J., Peto, Tim E. A., Hoosdally, S.J., Walker, A. Sarah, Sebra, R.P., Crook, Derrick W., Anjum, M.F., Read, Daniel S., Stoesser, N., Abuoun, M., Anjum, M., Bailey, M. J., Barker, L., Brett, H., Bowes, M. J., Chau, K., Crook, D. W., De Maio, N., Gilson, D., Gweon, H. S., Hubbard, A. T. M., Hoosdally, S., Kavanagh, J., Jones, H., Peto, T. E. A., Read, D. S., Sebra, R., Shaw, L. P., Sheppard, A.E., Smith, R., Stubberfield, E., Swann, J., Walker, A. S., Woodford, N., on behalf of the REHAB consortium, 2019. The impact of sequencing depth on the inferred taxonomic composition and AMR gene content of metagenomic samples. Environmental Microbiome 14, 7. https://doi.org/10.1186/s40793-019-0347-1

Haig, S.-J., Kotlarz, N., LiPuma, J.J., Raskin, L., 2018. A High-Throughput Approach for Identification of Nontuberculous Mycobacteria in Drinking Water Reveals Relationship between Water Age and. MBio 9. https://doi.org/10.1128/mBio.02354-17

Hall, M., Beiko, R.G., 2018. 16S rRNA Gene Analysis with QIIME2, in: Beiko, R.G., Hsiao, W., Parkinson, J. (Eds.), Microbiome Analysis: Methods and Protocols, Methods in Molecular Biology. Springer, New York, NY, pp. 113–129. https://doi.org/10.1007/978-1-4939-8728-3_8

Hamilton, J.J., Garcia, S.L., Brown, B.S., Oyserman, B.O., Moya-Flores, F., Bertilsson, S., Malmstrom, R.R., Forest, K.T., McMahon, K.D., 2017. Metabolic Network Analysis and Metatranscriptomics Reveal Auxotrophies and Nutrient Sources of the Cosmopolitan Freshwater Microbial Lineage acI. mSystems 2. https://doi.org/10.1128/mSystems.00091-17

Hammes, F., Berney, M., Wang, Y., Vital, M., Köster, O., Egli, T., 2008. Flow-cytometric total bacterial cell counts as a descriptive microbiological parameter for drinking water treatment processes. Water Res 42, 269–277. https://doi.org/10.1016/j.watres.2007.07.009

Hamner, S., Brown, B.L., Hasan, N.A., Franklin, M.J., Doyle, J., Eggers, M.J., Colwell, R.R., Ford, T.E., 2019. Metagenomic profiling of microbial pathogens in the little bighorn river, Montana. Int J Env. Res Public Health 16. https://doi.org/10.3390/ijerph16071097

Hampel, J.J., McCarthy, M.J., Neudeck, M., Bullerjahn, G.S., McKay, R.M.L., Newell, S.E., 2019. Ammonium recycling supports toxic Planktothrix blooms in Sandusky Bay, Lake Erie: Evidence from stable isotope and metatranscriptome data. Harmful Algae 81, 42–52. https://doi.org/10.1016/j.hal.2018.11.011

Handley, K.M., Bartels, D., O'Loughlin, E.J., Williams, K.H., Trimble, W.L., Skinner, K., Gilbert, J.A., Desai, N., Glass, E.M., Paczian, T., Wilke, A., Antonopoulos, D., Kemner, K.M., Meyer, F., 2014. The complete genome sequence for putative H2- and S-oxidizerCandidatusSulfuricurvum sp., assembledde novofrom an aquifer-derived metagenome. Environ. Microbiol. 16, 3443–3462. https://doi.org/10.1111/1462-2920.12453

Haramoto, E., Katayama, H., Utagawa, E., Ohgaki, S., 2008. Development of sample storage methods for detecting enteric viruses in environmental water. Journal of Virological Methods 151, 1–6. https://doi.org/10.1016/j.jviromet.2008.04.006

Harb, M., Hong, P.Y., 2017. Molecular-based detection of potentially pathogenic bacteria in membrane bioreactor (MBR) systems treating municipal wastewater: a case study. Env. Sci Pollut Res 24, 5370–5380. https://doi.org/10.1007/s11356-016-8211-y

Hardwick, S.A., Deveson, I.W., Mercer, T.R., 2017. Reference standards for next-generation sequencing. Nat Rev Genet 18, 473–484. https://doi.org/10.1038/nrg.2017.44

Harke, M.J., Davis, T.W., Watson, S.B., Gobler, C.J., 2016. Nutrient-Controlled Niche Differentiation of Western Lake Erie Cyanobacterial Populations Revealed via Metatranscriptomic Surveys. Env. Sci Technol 50, 604–615. https://doi.org/10.1021/acs.est.5b03931

Harrison, J.G., Randolph, G.D., Buerkle, C.A., 2021. Characterizing Microbiomes via Sequencing of Marker Loci: Techniques To Improve Throughput, Account for Cross-Contamination, and Reduce Cost. mSystems 6, e00294-21. https://doi.org/10.1128/mSystems.00294-21

Harwood, V.J., Levine, A.D., Scott, T.M., Chivukula, V., Lukasik, J., Farrah, S.R., Rose, J.B., 2005. Validity of the indicator organism paradigm for pathogen reduction in reclaimed water and public health protection. Applied and Environmental Microbiology 71, 3163–3170. https://doi.org/10.1128/AEM.71.6.3163-3170.2005

Hassoun-Kheir, N., Stabholz, Y., Kreft, J.-U., de la Cruz, R., Dechesne, A., Smets, B.F., Romalde, J.L., Lema, A., Balboa, S., García-Riestra, C., Torres-Sangiao, E., Neuberger, A., Graham, D., Quintela-Baluja, M., Stekel, D.J., Graham, J., Pruden, A., Nesme, J., Sørensen, S.J., Hough, R., Paul, M., 2021. EMBRACE-WATERS statement: Recommendations for reporting of studies on antimicrobial resistance in wastewater and related aquatic environments. One Health 13, 100339. https://doi.org/10.1016/j.onehlt.2021.100339

Hata, A., Hanamoto, S., Ihara, M., Shirasaka, Y., Yamashita, N., Tanaka, H., 2018a. Comprehensive Study on Enteric Viruses and Indicators in Surface Water in Kyoto, Japan, During 2014–2015 Season. Food Environ. Virol. 10, 353–364. https://doi.org/10.1007/s12560-018-9355-3

Hata, A., Kitajima, M., Haramoto, E., Lee, S., Ihara, M., Gerba, C.P., Tanaka, H., 2018b. Next-generation amplicon sequencing identifies genetically diverse human astroviruses, including recombinant strains, in environmental waters. Sci Rep 8, 11837. https://doi.org/10.1038/s41598-018-30217-y

He, S., Wurtzel, O., Singh, K., Froula, J.L., Yilmaz, S., Tringe, S.G., Wang, Z., Chen, F., Lindquist, E.A., Sorek, R., Hugenholtz, P., 2010. Validation of two ribosomal RNA removal methods for microbial metatranscriptomics. Nat Methods 7, 807–812. https://doi.org/10.1038/nmeth.1507

Hellein, K.N., Kennedy, E.M., Harwood, V.J., Gordon, K.V., Wang, S.Y., Lepo, J.E., 2012. A filter-based propidium monoazide technique to distinguish live from membrane-compromised microorganisms using quantitative PCR. J Microbiol Methods 89, 76–78. https://doi.org/10.1016/j.mimet.2012.01.015

Hembach, N., Alexander, J., Hiller, C., Wieland, A., Schwartz, T., 2019. Dissemination prevention of antibiotic resistant and facultative pathogenic bacteria by ultrafiltration and ozone treatment at an urban wastewater treatment plant. Sci Rep 9, 1–12. https://doi.org/10.1038/s41598-019-49263-1

Hemme, C.L., Deng, Y., Gentry, T.J., Fields, M.W., Wu, L., Barua, S., Barry, K., Tringe, S.G., Watson, D.B., He, Z., Hazen, T.C., Tiedje, J.M., Rubin, E.M., Zhou, J., 2010. Metagenomic insights into evolution of a heavy metal-contaminated groundwater microbial community. ISME J 4, 660–672. https://doi.org/10.1038/ismej.2009.154

Hemme, C.L., Tu, Q., Shi, Z., Qin, Y., Gao, W., Deng, Y., Van Nostrand, J.D., Wu, L., He, Z., Chain, P.S.G., Tringe, S.G., Fields, M.W., Rubin, E.M., Tiedje, J.M., Hazen, T.C., Arkin, A.P., Zhou, J., 2015. Comparative metagenomics reveals impact of contaminants on groundwater microbiomes. Front Microbiol 6, 1205. https://doi.org/10.3389/fmicb.2015.01205

Hendriksen, Rene S, The Global Sewage Surveillance project consortium, Munk, P., Njage, P., van Bunnik, B., McNally, L., Lukjancenko, O., Röder, T., Nieuwenhuijse, D., Pedersen, S.K., Kjeldgaard, J., Kaas, R.S., Philip Thomas Lanken, Vogt, J.K., Leekitcharoenphon, P., van de Schans, M.G.M., Zuidema, T., de Roda Husman, A.M., Rasmussen, S., Petersen, B., Amid, C., Cochrane, G., Sicheritz-Ponten, T., Schmitt, H., Alvarez, J.R.M., Aidara-Kane, A., Pamp, S.J., Lund, O., Hald, T., Woolhouse, M., Koopmans, M.P., Vigre, H., Petersen, T.N., Aarestrup, F.M., 2019. Global monitoring of antimicrobial resistance based on metagenomics analyses of urban sewage. Nature Communications 10. https://doi.org/10.1038/s41467-019-08853-3

Herson, D., Nichols, C., Verville, K., Stromberg, T., Bright, A., Ramírez Toro, G., Martinez, L., Ramirez, Z. and Minnigh, H. 2005 Occurrence of Salmonella spp.

Hess, J.F., Kohl, T.A., Kotrová, M., Rönsch, K., Paprotka, T., Mohr, V., Hutzenlaub, T., Brüggemann, M., Zengerle, R., Niemann, S., Paust, N., 2020. Library preparation for next generation sequencing: A review of automation strategies. Biotechnol Adv 41, 107537. https://doi.org/10.1016/j.biotechadv.2020.107537

Hewson, I., Barbosa, J.G., Brown, J.M., Donelan, R.P., Eaglesham, J.B., Eggleston, E.M., LaBarre, B.A., 2012. Temporal dynamics and decay of putatively allochthonous and autochthonous viral genotypes in contrasting freshwater lakes. Appl Env. Microbiol 78, 6583–6591. https://doi.org/10.1128/AEM.01705-12

Hewson, I., Bistolas, K.S.I., Button, J.B., Jackson, E.W., 2018. Occurrence and seasonal dynamics of RNA viral genotypes in three contrasting temperate lakes. PLoS One 13, e0194419. https://doi.org/10.1371/journal.pone.0194419

Hicks, L.A., Rose, C., Fields, B., Drees, M., Engel, J., Jenkins, P., Rouse, B., Blythe, D., Khalifah, A. and Feikin, D. 2007. Increased rainfall is associated with increased risk for legionellosis. Epidemiology & Infection 135(5), 811-817.

Higgins, J.A., Jenkins, M.C., Shelton, D.R., Fayer, R. and Karns, J.S. 2001. Rapid Extraction of DNA From Escherichia coli and Cryptosporidium parvum for Use in PCR. Appl Environ Microb 67(11), 5321-5324.

Highlander, S., 2013. Mock Community Analysis, in: Nelson, K.E. (Ed.), Encyclopedia of Metagenomics. Springer, New York, NY, pp. 1–7. https://doi.org/10.1007/978-1-4614-6418-1_54-1

Hill, V.R., Narayanan, J., Gallen, R.R., Ferdinand, K.L., Cromeans, T., Vinjé, J., 2015. Development of a Nucleic Acid Extraction Procedure for Simultaneous Recovery of DNA and RNA from Diverse Microbes in Water. Pathogens 4, 335–335. https://doi.org/10.3390/PATHOGENS4020335

Hinlo, R., Gleeson, D., Lintermans, M., Furlan, E., 2017. Methods to maximise recovery of environmental DNA from water samples. PLoS ONE 12, e0179251. https://doi.org/10.1371/journal.pone.0179251

Hjelmsø, M.H., Hellmér, M., Fernandez-Cassi, X., Timoneda, N., Lukjancenko, O., Seidel, M., Elsässer, D., Aarestrup, F.M., Löfström, C., Bofill-Mas, S., Abril, J.F., Girones, R., Schultz, A.C.,

2017. Evaluation of Methods for the Concentration and Extraction of Viruses from Sewage in the Context of Metagenomic Sequencing. PLoS One 12, e0170199. https://doi.org/10.1371/journal.pone.0170199

Hoefel, D., Grooby, W.L., Monis, P.T., Andrews, S., Saint, C.P., 2003. Enumeration of water-borne bacteria using viability assays and flow cytometry: a comparison to culture-based techniques. J Microbiol Methods 55, 585–597. https://doi.org/10.1016/s0167-7012(03)00201-x

Hokajärvi, A.-M., Rytkönen, A., Tiwari, A., Kauppinen, A., Oikarinen, S., Lehto, K.-M., Kankaanpää, A., Gunnar, T., Al-Hello, H., Blomqvist, S., Miettinen, I.T., Savolainen-Kopra, C., Pitkänen, T., 2021. The detection and stability of the SARS-CoV-2 RNA biomarkers in wastewater influent in Helsinki, Finland. Science of The Total Environment 770, 145274. https://doi.org/10.1016/j.scitotenv.2021.145274

Hornstra, L.M., da Silva, T.R., Blankert, B., Heijnen, L., Beerendonk, E., Cornelissen, E.R., Medema, G., 2019. Monitoring the integrity of reverse osmosis membranes using novel indigenous freshwater viruses and bacteriophages. Environ. Sci. Water Res. Technol. 5, 1535–1544. https://doi.org/10.1039/c9ew00318e

Hornung, B.V.H., Zwittink, R.D., Kuijper, E.J., 2019. Issues and current standards of controls in microbiome research. FEMS Microbiology Ecology 95, 45. https://doi.org/10.1093/FEMSEC/FIZ045

Hou, Q., Fang, Z., Zhu, Q., Dong, H., 2019. Microbial diversity in Huguangyan Maar Lake of China revealed by high–throughput sequencing. J. Oceanol. Limnol. 37, 1245–1257. https://doi.org/10.1007/s00343-019-8016-1

Huang, K., Mao, Y., Zhao, F., Zhang, X.X., Ju, F., Ye, L., Wang, Y., Li, B., Ren, H., Zhang, T., 2018. Free-living bacteria and potential bacterial pathogens in sewage treatment plants. Appl Microbiol Biotechnol 102, 2455–2464. https://doi.org/10.1007/s00253-018-8796-9

Huang, K., Zhang, X.-X., Shi, P., Wu, B., Ren, H., 2014. A comprehensive insight into bacterial virulence in drinking water using 454 pyrosequencing and Illumina high-throughput sequencing. Ecotoxicol Env. Saf 109, 15–21. https://doi.org/10.1016/j.ecoenv.2014.07.029

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J., von Mering, C., Bork, P., 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 47, D309–D314. https://doi.org/10.1093/nar/gky1085

Huggett, J.F., Foy, C.A., Benes, V., Emslie, K., Garson, J.A., Haynes, R., Hellemans, J., Kubista, M., Mueller, R.D., Nolan, T., Pfaffl, M.W., Shipley, G.L., Vandesompele, J., Wittwer, C.T., Bustin, S.A., 2013. The digital MIQE guidelines: Minimum information for publication of quantitative digital PCR experiments. Clinical Chemistry 59, 892–902. https://doi.org/10.1373/clinchem.2013.206375

Hull, N.M., Holinger, E.P., Ross, K.A., Robertson, C.E., Harris, J.K., Stevens, M.J., Pace, N.R., 2017. Longitudinal and Source-to-Tap New Orleans, LA, U.S.A. Drinking Water Microbiology. Environ. Sci. Technol. 51, 4220–4229. https://doi.org/10.1021/acs.est.6b06064

Huson, D.H., Auch, A.F., Qi, J., Schuster, S.C., 2007. MEGAN analysis of metagenomic data. Genome Res. 17, 377–386. https://doi.org/10.1101/gr.5969107

Hwang, C., Ling, F., Andersen, G.L., LeChevallier, M.W., Liu, W.-T., 2012. Evaluation of Methods for the Extraction of DNA from Drinking Water Distribution System Biofilms. Microbes Environ. 27, 9–18. https://doi.org/10.1264/jsme2.me11132

Iaconelli, M., Valdazo-González, B., Equestre, M., Ciccaglione, A.R., Marcantonio, C., Della Libera, S., La Rosa, G., 2017. Molecular characterization of human adenoviruses in urban wastewaters using next generation and Sanger sequencing. Water Res. 121, 240–247. https://doi.org/10.1016/j.watres.2017.05.039

Iker, B.C., Bright, K.R., Pepper, I.L., Gerba, C.P., Kitajima, M., 2013. Evaluation of commercial kits for the extraction and purification of viral nucleic acids from environmental and fecal samples. Journal of virological methods 191, 24–30. https://doi.org/10.1016/J.JVIROMET.2013.03.011

Illumina Inc., 2010. TruSeqTM DNA Sample Preparation Guide.

Illumina Inc., 2019. Nextera XT DNA Library Prep Reference Guide (No. 15031942 v05).

Inglis, T., Garrow, S., Henderson, M., Clair, A., Sampson, J., O'Reilly, L. and Cameron, B. 2000. Burkholderia pseudomallei traced to water treatment plant in Australia. Emerging Infectious Diseases 6(1), 56.

Jacquiod, S., Cyriaque, V., Riber, L., Al-Soud, W.A., Gillan, D.C., Wattiez, R., Sørensen, S.J., 2018. Long-term industrial metal contamination unexpectedly shaped diversity and activity response of sediment microbiome. Journal of Hazardous Materials 344, 299–307. https://doi.org/10.1016/j.jhazmat.2017.09.046

Jadeja, N.B., More, R.P., Purohit, H.J., Kapley, A., 2014. Metagenomic analysis of oxygenases from activated sludge. Bioresour. Technol. 165, 250–256. https://doi.org/10.1016/j.biortech.2014.02.045

Jadeja, N.B., Purohit, H.J., Kapley, A., 2019. Decoding microbial community intelligence through metagenomics for efficient wastewater treatment. Funct. Integr. Genomics 19, 839–851. https://doi.org/10.1007/s10142-019-00681-4

Jewell, T.N.M., Karaoz, U., Brodie, E.L., Williams, K.H., Beller, H.R., 2016. Metatranscriptomic evidence of pervasive and diverse chemolithoautotrophy relevant to C, S, N and Fe cycling in a shallow alluvial aquifer. ISME J. 10, 2106–2117. https://doi.org/10.1038/ismej.2016.25

Ji, P., Parks, J., Edwards, M.A., Pruden, A., 2015. Impact of Water Chemistry, Pipe Material and

Stagnation on the Building Plumbing Microbiome. PLoS One 10, e0141087. https://doi.org/10.1371/journal.pone.0141087

Jia, S., Wu, J., Ye, L., Zhao, F., Li, T., Zhang, X.-X., 2019. Metagenomic assembly provides a deep insight into the antibiotic resistome alteration induced by drinking water chlorination and its correlations with bacterial host changes. J. Hazard. Mater. 379, 120841. https://doi.org/10.1016/j.jhazmat.2019.120841

Jiang, X., Cui, X., Xu, H., Liu, W., Tao, F., Shao, T., Pan, X., Zheng, B., 2019. Whole Genome Sequencing of Extended-Spectrum Beta-Lactamase (ESBL)-Producing Escherichia coli Isolated From a Wastewater Treatment Plant in China. Front. Microbiol. 10. https://doi.org/10.3389/fmicb.2019.01797

Jin, D., Kong, X., Cui, B., Jin, S., Xie, Y., Wang, X., Deng, Y., 2018. Bacterial communities and potential waterborne pathogens within the typical urban surface waters. Sci Rep 8, 1–9. https://doi.org/10.1038/s41598-018-31706-w

Johnson, D.W., Pieniazek, N.J., Griffin, D.W., Misener, L. and Rose, J.B. 1995. Development of a PCR protocol for sensitive detection of Cryptosporidium oocysts in water samples. Appl Environ Microb 61(11), 3849-3855.

Johnson, J.S., Spakowicz, D.J., Hong, B.-Y., Petersen, L.M., Demkowicz, P., Chen, L., Leopold, S.R., Hanson, B.M., Agresta, H.O. and Gerstein, M. 2019. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. Nature communications 10(1), 1-11.

Ju, F., Beck, K., Yin, X., Maccagnan, A., McArdell, C.S., Singer, H.P., Johnson, D.R., Zhang, T., Bürgmann, H., 2019. Wastewater treatment plant resistomes are shaped by bacterial composition, genetic exchange, and upregulated expression in the effluent microbiomes: Supplementary material. ISME Journal 1–39.

Ju, F., Guo, F., Ye, L., Xia, Y., Zhang, T., 2014. Metagenomic analysis on seasonal microbial variations of activated sludge from a full-scale wastewater treatment plant over 4 years. Environ. Microbiol. Rep. 6, 80–89. https://doi.org/10.1111/1758-2229.12110

Judge, K., Harris, S.R., Reuter, S., Parkhill, J., Peacock, S.J., 2015. Early insights into the potential of the Oxford Nanopore MinION for the detection of antimicrobial resistance genes. J Antimicrob Chemother 70, 2775–2778. https://doi.org/10.1093/jac/dkv206

Julian, T.R., Schwab, K.J., 2012. Challenges in environmental detection of human viral pathogens. Current opinion in virology 2, 78–83.

Juretschko, S., Timmermann, G., Schmid, M., Schleifer, K.-H., Pommerening-Röser, A., Koops, H.-P. and Wagner, M. 1998. Combined Molecular and Conventional Analyses of Nitrifying Bacterium Diversity in Activated Sludge: Nitrosococcus mobilisand Nitrospira-Like Bacteria as Dominant Populations. Appl Environ Microb 64(8), 3042-3051.

Kaas, L., Ogorzaly, L., Lecellier, G., Berteaux-Lecellier, V., Cauchie, H.-M., Langlet, J., 2019. Detection of Human Enteric Viruses in French Polynesian Wastewaters, Environmental Waters and Giant Clams. Food Env. Virol 11, 52–64. https://doi.org/10.1007/s12560-018-9358-0

Kahlisch, L., Henne, K., Groebe, L., Draheim, J., Höfle, M.G., Brettar, I., 2010. Molecular analysis of the bacterial drinking water community with respect to live/dead status. Water Sci Technol 61, 9–14. https://doi.org/10.2166/wst.2010.773

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., Morishima, K., 2017. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res 45, D353–D361. https://doi.org/10.1093/nar/gkw1092

Kanehisa, M., Goto, S., 2000. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 28, 27–30. https://doi.org/10.1093/nar/28.1.27

Kang, D.D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., Wang, Z., 2019. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ 7, e7359. https://doi.org/10.7717/peerj.7359

KAPA Biosystems, 2019. KAPA HyperPrep Kit Technical Data Sheet (No. KR0961 – v7.19).

Kapustina, Ž., Medžiūnė, J., Alzbutas, G., Rokaitis, I., Matjošaitis, K., Mackevičius, G., Žeimytė, S., Karpus, L., Lubys, A. 2021, 2021. High-resolution microbiome analysis enabled by linking of 16S rRNA gene sequences with adjacent genomic contexts. Microbial Genomics 7, 000624. https://doi.org/10.1099/mgen.0.000624

Karp, P.D., Billington, R., Caspi, R., Fulcher, C.A., Latendresse, M., Kothari, A., Keseler, I.M., Krummenacker, M., Midford, P.E., Ong, Q., Ong, W.K., Paley, S.M., Subhraveti P., 2019. The BioCyc collection of microbial genomes and metabolic pathways, Briefings in Bioinformatics, 20(4), 1085–1093. https://doi.org/10.1093/bib/bbx085

Kavagutti, V.S., Andrei, A.-Ş., Mehrshad, M., Salcher, M.M., Ghai, R., 2019. Phage-centric ecological interactions in aquatic ecosystems revealed through ultra-deep metagenomics. Microbiome 7, 135. https://doi.org/10.1186/s40168-019-0752-0

Kaya, D., Niemeier, D., Ahmed, W., Kjellerup, B.V., 2022. Evaluation of multiple analytical methods for SARS-CoV-2 surveillance in wastewater samples. Science of The Total Environment 808, 152033. https://doi.org/10.1016/j.scitotenv.2021.152033

Keenum, I., Medina, M.C., Garner, E., Pieper, K.J., Blair, M.F., Milligan, E., Pruden, A., Ramirez-Toro, G., Rhoads, W.J., 2021. Source-to-Tap Assessment of Microbiological Water Quality in Small Rural Drinking Water Systems in Puerto Rico Six Months After Hurricane Maria. Environ. Sci. Technol. acs.est.0c08814. https://doi.org/10.1021/acs.est.0c08814

Keller, A.H., Schleinitz, K.M., Starke, R., Bertilsson, S., Vogt, C., Kleinsteuber, S., 2015. Metagenome-Based Metabolic Reconstruction Reveals the Ecophysiological Function of

Epsilonproteobacteria in a Hydrocarbon-Contaminated Sulfidic Aquifer. Front. Microbiol. 6. https://doi.org/10.3389/fmicb.2015.01396

Kim, D., Song, L., Breitwieser, F.P., Salzberg, S.L., 2016. Centrifuge: rapid and accurate classification of metagenomic sequences. bioRxiv 26, 054965. https://doi.org/10.1101/054965

Kirby, A., 2020. Antibiotic Resistance Monitoring Using Extended Spectrum Beta-Lactamase-Producing E. coli as an Indicator Organism, in: AWWA International Symposium on Potable Reuse. Atalanta, GA.

Kirtane, A., Atkinson, J.D., Sassoubre, L., 2020. Design and Validation of Passive Environmental DNA Samplers Using Granular Activated Carbon and Montmorillonite Clay. Environmental science & technology 54. https://doi.org/10.1021/ACS.EST.0C01863

Kiseleva, L., Garushyants, S.K., Ma, H., Simpson, D.J.W., Fedorovich, V., Cohen, M.F., Goryanin, I., 2015. Taxonomic and functional metagenomic analysis of anodic communities in two pilot-scale microbial fuel cells treating different industrial wastewaters. J. Integr. Bioinforma. 12, 1–15. https://doi.org/10.1515/jib-2015-273

Klempay, B., Arandia-Gorostidi, N., Dekas, A.E., Bartlett, D.H., Carr, C.E., Doran, P.T., Dutta, A., Erazo, N., Fisher, L.A., Glass, J.B., 2021. Microbial diversity and activity in Southern California salterns and bitterns: analogues for remnant ocean worlds. Environmental Microbiology.

Knudsen, B.E., Bergmark, L., Munk, P., Lukjancenko, O., Priemé, A., Aarestrup, F.M., Pamp, S.J., 2016. Impact of Sample Type and DNA Isolation Procedure on Genomic Inference of Microbiome Composition. mSystems 1. https://doi.org/10.1128/MSYSTEMS.00095-16

Kobayashi, H., Oethinger, M., Tuohy, M.J., Hall, G.S., Bauer, T.W., 2009. Unsuitable distinction between viable and deadStaphylococcus aureusandStaphylococcus epidermidisby ethidium bromide monoazide. Lett Appl Microbiol 48, 633–638. https://doi.org/10.1111/j.1472-765x.2009.02585.x

Kohl, C., Wegener, M., Nitsche, A., Kurth, A., 2017. Use of RNALater® Preservation for Virome Sequencing in Outbreak Settings. Front Microbiol 8:1888. https://doi.org/10.3389/fmicb.2017.01888

Kolmogorov, M., Bickhart, D.M., Behsaz, B., Gurevich, A., Rayko, M., Shin, S.B., Kuhn, K., Yuan, J., Polevikov, E., Smith, T.P.L., Pevzner, P.A., 2020. metaFlye: scalable long-read metagenome assembly using repeat graphs. Nature Methods 2020 17:11 17, 1103–1110. https://doi.org/10.1038/s41592-020-00971-x

Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., Phillippy, A.M., 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Research 27, gr.215087.116. https://doi.org/10.1101/GR.215087.116

Kori, J.A., Mahar, R.B., Vistro, M.R., Tariq, H., Khan, I.A., Goel, R., 2019. Metagenomic analysis of drinking water samples collected from treatment plants of Hyderabad City and Mehran University Employees Cooperative Housing Society. Env. Sci Pollut Res Int 26, 29052–29064. https://doi.org/10.1007/s11356-019-05859-8

Korzeniewska, Ewa. 2011. “Emission of bacteria and fungi in the air from wastewater treatment plants - a review.” Front Biosci (Schol Ed) 2011 Jan 1 (3) 393-407.

Kranz, A., Vogel, A., Degner, U., Kiefler, I., Bott, M., Usadel, B., Polen, T., 2017. High precision genome sequencing of engineered Gluconobacter oxydans 621H by combining long nanopore and short accurate Illumina reads. J Biotechnol 258, 197–205. https://doi.org/10.1016/j.jbiotec.2017.04.016

Kuhn, R., Böllmann, J., Krahl, K., Bryant, I.M., Martienssen, M., 2017. Comparison of ten different DNA extraction procedures with respect to their suitability for environmental samples. Journal of Microbiological Methods 143, 78–86. https://doi.org/10.1016/j.mimet.2017.10.007

Kumar, G., Eble, J.E., Gaither, M.R., 2020. A practical guide to sample preservation and pre-PCR processing of aquatic environmental DNA. Mol Ecol Resour 20, 29–39. https://doi.org/10.1111/1755-0998.13107

Kumar, P.S., Suganya, S., Varjani, S.J., 2018. Evaluation of Next-Generation Sequencing Technologies for Environmental Monitoring in Wastewater Abatement, in: Bioremediation: Applications for Environmental Protection and Management. Springer, pp. 29–52.

Kumaraswamy, R., Amha, Y.M., Anwar, M.Z., Henschel, A., Rodríguez, J., Ahmad, F., 2014. Molecular Analysis for Screening Human Bacterial Pathogens in Municipal Wastewater Treatment and Reuse. Environmental Science & Technology 48, 11610–11619. https://doi.org/10.1021/es502546t

Kuo, H.-W., Chen, L.-Z., Shih, M.-H., 2015. High prevalence of type 41 and high sequence diversity of partial hexon gene of human adenoviruses in municipal raw sewage and activated sludge. J Appl Microbiol 119, 1181–1195. https://doi.org/10.1111/jam.12907

Kurilkina, M.I., Zakharova, Y.R., Galachyants, Y.P., Petrova, D.P., Bukin, Y.S., Domysheva, V.M., Blinov, V.V., Likhoshway, Y.V., 2016. Bacterial community composition in the water column of the deepest freshwater Lake Baikal as determined by next-generation sequencing. FEMS Microbiol Ecol 92. https://doi.org/10.1093/femsec/fiw094

Kurobe, T., Lehman, P.W., Hammock, B.G., Bolotaolo, M.B., Lesmeister, S., Teh, S.J., 2018. Biodiversity of cyanobacteria and other aquatic microorganisms across a freshwater to brackish water gradient determined by shotgun metagenomic sequencing analysis in the San Francisco Estuary, USA. PLoS One 13, e0203953. https://doi.org/10.1371/journal.pone.0203953

Labonte, J., 2016. Peer Review #1 of “Viral recombination blurs taxonomic lines: examination of single-stranded DNA viruses in a wastewater treatment plant (v0.1).” https://doi.org/10.7287/peerj.2585v0.1/reviews/1

Land, M., Hauser, L., Jun, S.R., Nookaew, I., Leuze, M.R., Ahn, T.H., Karpinets, T., Lund, O., Kora, G., Wassenaar, T., Poudel, S., Ussery, D.W., 2015. Insights from 20 years of bacterial genome sequencing. Functional & Integrative Genomics 15, 141. https://doi.org/10.1007/S10142-015-0433-4

Lande, L., Alexander, D.C., Wallace, R.J., Kwait, R., Iakhiaeva, E., Williams, M., Cameron, A.D.S., Olshefsky, S., Devon, R., Vasireddy, R., Peterson, D.D., Falkinham, J.O., 2019. Mycobacterium avium in community and household water, suburban Philadelphia, Pennsylvania, USA, 2010-2012. Emerg Infect Dis 25, 473–481. https://doi.org/10.3201/eid2503.180336

Lane, D.J., Pace, B., Olsen, G.J., Stahl, D.A., Sogin, M.L. and Pace, N.R. 1985. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. Proceedings of the National Academy of Sciences 82(20), 6955-6959.

Langille, M.G.I., Zaneveld, J., Caporaso, J.G., McDonald, D., Knights, D., Reyes, J.A., Clement, J.C., Burkepile, D.E., Vega Thurber, R. L., Knight, R., Beiko, R. G., Huttenhower, C. 2013. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nat Biotechnol 31, 814-821. https://doi.org/10.1038/nbt.2676

Langmead Ben, Steven, S., 2013. Fast gapped-read alignment with Bowtie 2. Nature methods 9, 357–359. https://doi.org/10.1038/nmeth.1923.Fast

Lanza, V.F., Baquero, F., Martínez, J.L., Ramos-Ruíz, R., González-Zorn, B., Andremont, A., Sánchez-Valenzuela, A., Ehrlich, S.D., Kennedy, S., Ruppé, E., van Schaik, W., Willems, R.J., de la Cruz, F., Coque, T.M., 2018. In-depth resistome analysis by targeted metagenomics. Microbiome 6, 11. https://doi.org/10.1186/s40168-017-0387-y

Lautenschlager, K., Hwang, C., Ling, F., Liu, W.-T., Boon, N., Köster, O., Egli, T., Hammes, F., 2014. Abundance and composition of indigenous bacterial communities in a multi-step biofiltration-based drinking water treatment plant. Water Res. 62, 40–52. https://doi.org/10.1016/j.watres.2014.05.035

Layton, A.C., Chauhan, A., Williams, D.E., Mailloux, B., Knappett, P.S.K., Ferguson, A.S., McKay, L.D., Alam, M.J., Matin Ahmed, K., van Geen, A., Sayler, G.S., 2014. Metagenomes of microbial communities in arsenic- and pathogen-contaminated well and surface water from bangladesh. Genome Announc 2. https://doi.org/10.1128/genomeA.01170-14

LeBrun, E.S., King, R.S., Back, J.A., Kang, S., 2018. A Metagenome-Based Investigation of Gene Relationships for Non-Substrate-Associated Microbial Phosphorus Cycling in the Water Column of Streams and Rivers. Microb Ecol 76, 856–865. https://doi.org/10.1007/s00248-018-1178-0

LeChevallier, M.W., Mansfield, T.J., Gibson, J.M., 2020. Protecting wastewater workers from disease risks: Personal protective equipment guidelines. Water Environment Research 92, 524–533. https://doi.org/10.1002/wer.1249

Leddy, M.B., Hasan, N.A., Subramanian, P., Heberling, C., Cotruvo, J., Colwell, R.R., 2017.

Characterization of Microbial Signatures From Advanced Treated Wastewater Biofilms. J. - Am. Water Works Assoc. 109, E503–E512. https://doi.org/10.5942/jawwa.2017.109.0116

Lee, E., Khurana, M.S., Whiteley, A.S., Monis, P.T., Bath, A., Gordon, C., Ryan, U.M., Paparini, A., 2017. Novel Primer Sets for Next Generation Sequencing-Based Analyses of Water Quality. PLoS One 12, e0170008. https://doi.org/10.1371/journal.pone.0170008

Lee, K., Kim, D., Lee, D., Kim, Y., Bu, J., Cha, J., Thawng, C.N., Hwang, E., Seong, H.J., Sul, W.J., Wellington, E.M.H., Quince, C., Cha, C., 2020. Mobile resistome of human gut and pathogen drives anthropogenic bloom of antibiotic resistance. Microbiome 1–14.

Lee, M.T., Pruden, A., and Marr, L.C. 2016. Partitioning of viruses in wastewater systems and potential for aerosolization. Environmental Science and Technology Letters 3 (5): 210–15.

Lekunberri, I., Balcázar, J.L., Borrego, C.M., 2018. Metagenomic exploration reveals a marked change in the river resistome and mobilome after treated wastewater discharges. Environ. Pollut. 234, 538–542. https://doi.org/10.1016/j.envpol.2017.12.001

Lemarchand, K., Berthiaume, F., Maynard, C., Harel, J., Payment, P., Bayardelle, P., Masson, L., Brousseau, R., 2005. Optimization of microbial DNA extraction and purification from raw wastewater samples for downstream pathogen detection by microarrays. Journal of microbiological methods 63, 115–126. https://doi.org/10.1016/J.MIMET.2005.02.021

Lennarz, W. J., & Lane, M. D. (2013). Encyclopedia of biological chemistry. Academic Press.
Leplae, R., Lima-Mendez, G., Toussaint, A., 2010. ACLAME: a CLAssification of Mobile genetic Elements, update 2010. Nucleic Acids Res. 38, D57-61. https://doi.org/10.1093/nar/gkp938

Leplae, R., Lima-Mendez, G., Toussaint, A., 2010. ACLAME: a CLAssification of Mobile genetic Elements, update 2010. Nucleic Acids Res 38, D57-61. https://doi.org/10.1093/nar/gkp938

Lévesque, S., Lalancette, C., Bernard, K., Pacheco, A.L., Dion, R., Longtin, J., Tremblay, C., 2016. Molecular typing of Legionella pneumophila isolates in the province of Quebec from 2005 to 2015. PLoS One 11, 1–12. https://doi.org/10.1371/journal.pone.0163818

Lévesque, S., Plante, P.L., Mendis, N., Cantin, P., Marchand, G., Charest, H., Raymond, F., Huot, C., Goupil-Sormany, I., Desbiens, F., Faucher, S.P., Corbeil, J., Tremblay, C., 2014. Genomic characterization of a large outbreak of Legionella pneumophila serogroup 1 strains in Quebec City, 2012. PLoS One 9. https://doi.org/10.1371/journal.pone.0103852

Li, A.-D., Ma, L., Jiang, X.-T., Zhang, T., 2017. Cultivation-dependent and high-throughput sequencing approaches studying the co-occurrence of antibiotic resistance genes in municipal sewage system. Appl. Microbiol. Biotechnol. 101, 8197–8207. https://doi.org/10.1007/s00253-017-8573-1

Li, A.D., Metch, J.W., Wang, Y., Garner, E., Zhang, A.N., Riquelme, M. V., Vikesland, P.J., Pruden, A., Zhang, T., 2018. Effects of sample preservation and DNA extraction on enumeration of antibiotic resistance genes in wastewater. FEMS microbiology ecology 94, 1–11. https://doi.org/10.1093/femsec/fix189

Li, B., Ju, F., Cai, L., Zhang, T., 2015a. Profile and Fate of Bacterial Pathogens in Sewage Treatment Plants Revealed by High-Throughput Metagenomic Approach. Environ. Sci. Technol. 49, 10492–10502. https://doi.org/10.1021/acs.est.5b02345

Li, B., Saingam, P., Ishii, S., Yan, T., 2019a. Multiplex PCR coupled with direct amplicon sequencing for simultaneous detection of numerous waterborne pathogens. Applied Microbiology and Biotechnology 103, 953–961. https://doi.org/10.1007/s00253-018-9498-z

Li, B., Yang, Y., Ma, L., Ju, F., Guo, F., Tiedje, J.M., Zhang, T., 2015b. Metagenomic and network analysis reveal wide distribution and co-occurrence of environmental antibiotic resistance genes. ISME Journal 9, 2490–2502. https://doi.org/10.1038/ismej.2015.59

Li, D., Liu, C.-M., Luo, R., Sadakane, K., Lam, T.-W., 2015c. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31, 1674–1676. https://doi.org/10.1093/bioinformatics/btv033

Li, D., Luo, R., Liu, C.M., Leung, C.M., Ting, H.F., Sadakane, K., Yamashita, H., Lam, T.W., 2016. MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. Methods (San Diego, Calif.) 102, 3–11. https://doi.org/10.1016/J.YMETH.2016.02.020

Li, H., 2022. seqtk: Toolkit for processing sequences in FASTA/Q formats. https://github.com/lh3/seqtk

Li, T., Guo, F., Lin, Y., Li, Y., Wu, G., 2019b. Metagenomic analysis of quorum sensing systems in activated sludge and membrane biofilm of a full-scale membrane bioreactor. J. Water Process Eng. 32, 100952. https://doi.org/10.1016/j.jwpe.2019.100952

Li, X., Harwood, V.J., Nayak, B., Staley, C., Sadowsky, M.J., Weidhaas, J., 2015d. A Novel Microbial Source Tracking Microarray for Pathogen Detection and Fecal Source Identification in Environmental Systems. Environmental Science and Technology 49, 7319–7329. https://doi.org/10.1021/acs.est.5b00980

Lim, M.Y., Song, E.J., Kim, S.H., Lee, J., Nam, Y.D., 2018. Comparison of DNA extraction methods for human gut microbial community profiling. Systematic and applied microbiology 41, 151–157. https://doi.org/10.1016/J.SYAPM.2017.11.008

Lim, N.Y.N., Roco, C.A., Frostegård, Å., 2016. Transparent DNA/RNA co-extraction workflow protocol suitable for inhibitor-rich environmental samples that focuses on complete DNA removal for transcriptomic analyses. Frontiers in Microbiology 7, 1588. https://doi.org/10.3389/FMICB.2016.01588/BIBTEX

Lin, Y., Li, D., Zeng, S., He, M., 2016. Changes of microbial composition during wastewater reclamation and distribution systems revealed by high-throughput sequencing analyses. Front. Environ. Sci. Eng. 10, 539–547. https://doi.org/10.1007/s11783-016-0830-5

Ling, F., Hwang, C., LeChevallier, M.W., Andersen, G.L., Liu, W.T., 2015. Core-satellite populations and seasonality of water meter biofilms in a metropolitan drinking water distribution system. The ISME Journal 2016 10:3 10, 582–595. https://doi.org/10.1038/ismej.2015.136

Liu, G., Zhang, Y., Liu, X., Hammes, F., Liu, W.T., Medema, G., Wessels, P., Van Der Meer, W., 2020. 360-degree distribution of biofilm quantity and community in an operational unchlorinated drinking water distribution pipe. Environmental Science and Technology 54, 5619–5628. https://doi.org/10.1021/ACS.EST.9B06603/SUPPL_FILE/ES9B06603_SI_001.PDF

Liu, G., Zhang, Y., van der Mark, E., Magic-Knezev, A., Pinto, A., van den Bogert, B., Liu, W., van der Meer, W., Medema, G., 2018. Assessing the origin of bacteria in tap water and distribution system in an unchlorinated drinking water system by SourceTracker using microbial community fingerprints. Water Res 138, 86–96. https://doi.org/10.1016/j.watres.2018.03.043

Liu, M., Li, X., Xie, Y., Bi, D., Sun, J., Li, J., Tai, C., Deng, Z., Ou, H.-Y., 2019a. ICEberg 2.0: an updated database of bacterial integrative and conjugative elements. Nucleic Acids Res 47, D660–D665. https://doi.org/10.1093/nar/gky1123

Liu, X., Yang, X., Hu, X., He, Q., Zhai, J., Chen, Y., Xiong, Q., Vymazal, J., 2019b. Comprehensive metagenomic analysis reveals the effects of silver nanoparticles on nitrogen transformation in constructed wetlands. Chem. Eng. J. 358, 1552–1560. https://doi.org/10.1016/j.cej.2018.10.151

Liu, Z., Klümper, U., Liu, Y., Yang, Y., Wei, Q., Lin, J.-G., Gu, J.-D., Li, M., 2019c. Metagenomic and metatranscriptomic analyses reveal activity and hosts of antibiotic resistance genes in activated sludge. Environment International 129, 208–220. https://doi.org/10.1016/j.envint.2019.05.036

Llewellyn, A.C., Lucas, C.E., Roberts, S.E., Brown, E.W., Nayak, B.S., Raphael, B.H., Winchell, J.M., 2017. Distribution of Legionella and bacterial community composition among regionally diverse US cooling towers. PLoS One 12, 1–16. https://doi.org/10.1371/journal.pone.0189937

Llorens, C., Futami, R., Covelli, L., Domínguez-Escribá, L., Viu, J.M., Tamarit, D., Aguilar-Rodríguez, J., Vicente-Ripolles, M., Fuster, G., Bernet, G.P., Maumus, F., Munoz-Pomer, A., Sempere, J.M., Latorre, A., Moya, A., 2011. The Gypsy Database (GyDB) of mobile genetic elements: release 2.0. Nucleic Acids Res. 39, D70-4. https://doi.org/10.1093/nar/gkq1061

Loman, N., Rowe, W., Rambaut, A., 2020. nCoV-2019 novel coronavirus bioinformatics protocol.

Lombard, V., Ramulu, H.G., Drula, E., Coutinho, P.M., Henrissat, B., 2013. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 42, D490–D495. https://doi.org/10.1093/nar/gkt1178

Lozupone, C., Lladser, M.E., Knights, D., Stombaugh, J. and Knight, R. 2011. UniFrac: an effective distance metric for microbial community comparison. The ISME journal 5(2), 169-172.

Lu, X., Zhang, X.X., Wang, Z., Huang, K., Wang, Y., Liang, W., Tan, Y., Liu, B., Tang, J., 2015. Bacterial pathogens and community composition in advanced sewage treatment systems revealed by metagenomics analysis based on high-throughput sequencing. PLoS One 10. https://doi.org/10.1371/journal.pone.0125549

Lu, X.-M., Chen, C., Zheng, T.-L., 2017. Metagenomic Insights into Effects of Chemical Pollutants on Microbial Community Composition and Function in Estuarine Sediments Receiving Polluted River Water. Microb Ecol 73, 791–800. https://doi.org/10.1007/s00248-016-0868-8

Lu, X.-M., Lu, P.-Z., 2014. Characterization of bacterial communities in sediments receiving various wastewater effluents with high-throughput sequencing analysis. Microb Ecol 67, 612–623. https://doi.org/10.1007/s00248-014-0370-0

Ludington, W.B., Seher, T.D., Applegate, O., Li, X., Kliegman, J.I., Langelier, C., Atwill, E.R., Harter, T., DeRisi, J.L., 2017. Assessing biosynthetic potential of agricultural groundwater through metagenomic sequencing: A diverse anammox community dominates nitrate-rich groundwater. PLoS One 12, e0174930. https://doi.org/10.1371/journal.pone.0174930

Lugo, A.E. 2000. Effects and outcomes of Caribbean hurricanes in a climate change scenario. Science of the Total Environment 262(3), 243-251.

Ma, K.-L., Li, X.-K., Bao, L.-L., 2019. Influence of organic loading rate on purified terephthalic acid wastewater treatment in a temperature staged anaerobic treatment (TSAT) system: Performance and metagenomic characteristics. Chemosphere 220, 1091–1099. https://doi.org/10.1016/j.chemosphere.2019.01.028

Ma, L., Li, B., Zhang, T., 2019. New insights into antibiotic resistome in drinking water and management perspectives: A metagenomic based study of small-sized microbes. WATER RESEARCH 152, 191–201. https://doi.org/10.1016/j.watres.2018.12.069

Maghini, D.G., Moss, E.L., Vance, S.E., Bhatt, A.S., 2021. Improved high-molecular-weight DNA extraction, nanopore sequencing and metagenomic assembly from the human gut microbiome. Nature protocols 16, 458–471. https://doi.org/10.1038/S41596-020-00424-X

Majaneva, M., Diserud, O.H., Eagle, S.H.C., Boström, E., Hajibabaei, M., Ekrem, T., 2018. Environmental DNA filtration techniques affect recovered biodiversity. Sci Rep 8, 4682. https://doi.org/10.1038/s41598-018-23052-8

Majumdar, M., Klapsa, D., Wilton, T., Akello, J., Anscombe, C., Allen, D., Mee, E.T., Minor, P.D., Martin, J., 2018. Isolation of Vaccine-Like Poliovirus Strains in Sewage Samples From the United Kingdom. J Infect Dis 217, 1222–1230. https://doi.org/10.1093/infdis/jix667

Malki, K., Kula, A., Bruder, K., Sible, E., Hatzopoulos, T., Steidel, S., Watkins, S.C., Putonti, C., 2015. Bacteriophages isolated from Lake Michigan demonstrate broad host-range across several bacterial phyla. Virol J 12, 164. https://doi.org/10.1186/s12985-015-0395-0

Mancini, P., Bonanno Ferraro, G., Iaconelli, M., Suffredini, E., Valdazo-González, B., Della Libera, S., Divizia, M., La Rosa, G., 2019. Molecular characterization of human Sapovirus in untreated sewage in Italy by amplicon-based Sanger and next-generation sequencing. J Appl Microbiol 126, 324–331. https://doi.org/10.1111/jam.14129

Mansfeldt, C., Achermann, S., Men, Y., Walser, J.-C., Villez, K., Joss, A., Johnson, D.R., Fenner, K., 2019. Microbial residence time is a controlling parameter of the taxonomic composition and functional profile of microbial communities. ISME J. 13, 1589–1601. https://doi.org/10.1038/s41396-019-0371-6

Manz, W., Wagner, M., Amann, R., Schleifer, K.-H., 1994. In situ characterization of the microbial consortia active in two wastewater treatment plants. Water Research 28, 1715–1723. https://doi.org/10.1016/0043-1354(94)90243-7

Manzari, C., Oranger, A., Fosso, B., Piancone, E., Pesole, G., D'Erchia, A.M., 2020. Accurate quantification of bacterial abundance in metagenomic DNAs accounting for variable DNA integrity levels. Microb Genom 6, mgen000417. https://doi.org/10.1099/mgen.0.000417

Mao, D., Yu, S., Rysz, M., Luo, Y., Yang, F., Li, F., Hou, J., Mu, Q., Alvarez, P.J.J., 2015. Prevalence and proliferation of antibiotic resistance genes in two municipal wastewater treatment plants. Water Research 85, 458–466. https://doi.org/10.1016/j.watres.2015.09.010

Markowitz, V.M., Chen, I.-M.A., Palaniappan, K., Chu, K., Szeto, E., Grechkin, Y., Ratner, A., Jacob, B., Huang, J., Williams, P., Huntemann, M., Anderson, I., Mavromatis, K., Ivanova, N.N., Kyrpides, N.C., 2012. IMG: the Integrated Microbial Genomes database and comparative analysis system. Nucleic Acids Res. 40, D115-122. https://doi.org/10.1093/nar/gkr1044

Markt, R., Mayr, M., Peer, E., Wagner, A.O., Lackner, N., Insam, H., 2021. Detection and Stability of SARS-CoV-2 Fragments in Wastewater: Impact of Storage Temperature. Pathogens 10, 1215. https://doi.org/10.3390/pathogens10091215

Marotz, C., Amir, A., Humphrey, G., Gaffney, J., Gogul, G., Knight, R., 2017. DNA extraction for streamlined metagenomics of diverse environmental samples. BioTechniques 62, 290–293. https://doi.org/10.2144/000114559

Marti, E., Variatza, E., Balcazar, J.L., 2014. The role of aquatic ecosystems as reservoirs of antibiotic resistance. Trends in Microbiology 22, 36–41. https://doi.org/10.1016/j.tim.2013.11.001

Masella, A.P., Bartram, A.K., Truszkowski, J.M., Brown, D.G. and Neufeld, J.D. 2012. PANDAseq: paired-end assembler for illumina sequences. BMC Bioinformatics 13(1), 31.

Mason, O. U.; Scott, N. M.; Gonzalez, A.; Robbins-Pianka, A.; Bælum, J.; Kimbrel, J.; Bouskill, N. J.; Prestat, E.; Borglin, S.; Joyner, D. C.; Fortney, J. L.; Jurelevicius, D.; Stringfellow, W. T.;

Alvarez-Cohen, L.; Hazen, T. C.; Knight, R.; Gilbert, J. A.; Jansson, J. K. Metagenomics Reveals Sediment Microbial Community Response to Deepwater Horizon Oil Spill. ISME J. 2014, 8 (7), 1464–1475. https://doi.org/10.1038/ismej.2013.254.

Mason, O.U., Scott, N.M., Gonzalez, A., Robbins-Pianka, A., Bælum, J., Kimbrel, J., Bouskill, N.J., Prestat, E., Borglin, S., Joyner, D.C., Fortney, J.L., Jurelevicius, D., Stringfellow, W.T., Alvarez-Cohen, L., Hazen, T.C., Knight, R., Gilbert, J.A., Jansson, J.K., 2014. Metagenomics reveals sediment microbial community response to Deepwater Horizon oil spill. ISME J. 8, 1464–1475. https://doi.org/10.1038/ismej.2013.254

Mayo, M., Kaestli, M., Harrington, G., Cheng, A.C., Ward, L., Karp, D., Jolly, P., Godoy, D., Spratt, B.G. and Currie, B.J. 2011. Burkholderia pseudomallei in unchlorinated domestic bore water, Tropical Northern Australia. Emerging Infectious Diseases 17(7), 1283.

McMurdie, P.J., Holmes, S., 2014. Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. PLoS Comput Biol 10, e1003531. https://doi.org/10.1371/journal.pcbi.1003531

Medeiros, J.D., Cantão, M.E., Cesar, D.E., Nicolás, M.F., Diniz, C.G., Silva, V.L., Vasconcelos, A.T.R. de, Coelho, C.M., 2016. Comparative metagenome of a stream impacted by the urbanization phenomenon. Braz J Microbiol 47, 835–845. https://doi.org/10.1016/j.bjm.2016.06.011

Meng, Y., Huang, L.-N., Meng, F., 2019. Metagenomics Response of Anaerobic Ammonium Oxidation (anammox) Bacteria to Bio-Refractory Humic Substances in Wastewater. Water 11, 365. https://doi.org/10.3390/w11020365

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E.M., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R., Wilke, A., Wilkening, J., Edwards, R.A., 2008. The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics 9, 386. https://doi.org/10.1186/1471-2105-9-386

Minnigh H.A., R.T.G.I., Hunter P.R., Herson D., Verville K. 2006. Concurrence of standard microbial indicators of potable water quality, frank pathogens and human disease in small potable water systems. . PR002207. XXX Congreso Internacional de Ingeniería Sanitaria y Ambiental. AIDIS. .

Mizusawa, N., Reza, M.S., Oikawa, C., Kuga, S., Iijima, M., Kobiyama, A., Yamada, Y., Ikeda, Y., Ikeda, D., Ikeo, K., Sato, S., Ogata, T., Kudo, T., Jimbo, M., Yasumoto, K., Urano, N., Watabe, S., 2021. Diversity and functions of bacterial communities in water and sediment from the watershed of the Tama River flowing a highly urbanized area. Fisheries Science 87. https://doi.org/10.1007/s12562-021-01543-4

Mohan, A.M., Bibby, K.J., Lipus, D., Hammack, R.W., Gregory, K.B., 2014. The functional potential of microbial communities in hydraulic fracturing source water and produced water from natural gas extraction characterized by metagenomic sequencing. PLoS One 9, e107682. https://doi.org/10.1371/journal.pone.0107682

Mohiuddin, M., Schellhorn, H.E., 2015. Spatial and temporal dynamics of virus occurrence in two freshwater lakes captured through metagenomic analysis. Front Microbiol 6, 960. https://doi.org/10.3389/fmicb.2015.00960

Moran, M.A., 2009. Metatranscriptomics: Eavesdropping on Complex Microbial Communities. Microbe Mag. 4, 329–335. https://doi.org/10.1128/microbe.4.329.1

More, R.P., Mitra, S., Raju, S.C., Kapley, A., Purohit, H.J., 2014. Mining and assessment of catabolic pathways in the metagenome of a common effluent treatment plant to induce the degradative capacity of biomass. Bioresour. Technol. 153, 137–146. https://doi.org/10.1016/j.biortech.2013.11.065

Morvay, A.A., Decun, M., Scurtu, M., Sala, C., Morar, A., Sarandan, M., 2011. Biofilm formation on materials commonly used in household drinking water systems. Water Sci Technol Water Supply 11, 252–257.

Moss, E.L., Maghini, D.G., Bhatt, A.S., 2020. Complete, closed bacterial genomes from microbiomes using nanopore sequencing. Nature Biotechnology 2020 38:6 38, 701–707. https://doi.org/10.1038/s41587-020-0422-6

Mou, X., Lu, X., Jacob, J., Sun, S., Heath, R., 2013. Metagenomic identification of bacterioplankton taxa and pathways involved in microcystin degradation in lake erie. PLoS One 8, e61890. https://doi.org/10.1371/journal.pone.0061890

Moura, A., Soares, M., Pereira, C., Leitão, N., Henriques, I., Correia, A., 2009. INTEGRALL: a database and search engine for integrons, integrases and gene cassettes. Bioinformatics 25, 1096–1098. https://doi.org/10.1093/bioinformatics/btp105

Mousazadeh, M., Ashoori, R., Paital, B., Kabdaşlı, I., Frontistis, Z., Hashemi, M., Sandoval, M.A., Sherchan, S., Das, K., Emamjomeh, M.M., 2021. Wastewater Based Epidemiology Perspective as a Faster Protocol for Detecting Coronavirus RNA in Human Populations: A Review with Specific Reference to SARS-CoV-2 Virus. Pathogens 10, 1008. https://doi.org/10.3390/pathogens10081008

Mthethwa, N. P., Amoah, I.D., Reddy, P., Bux, F., Kumari, S., 2021. A review on application of next-generation sequencing methods for profiling of protozoan parasites in water: Current methodologies, challenges, and perspectives. Journal of Microbiological Methods. https://doi.org/10.1016/j.mimet.2021.106269

Mukherjee, N., Bartelli, D., Patra, C., Chauhan, B.V., Dowd, S.E., Banerjee, P., 2016. Microbial diversity of source and point-of-use water in rural Haiti - A pyrosequencing-based metagenomic survey. PLoS One 11, 1–16. https://doi.org/10.1371/journal.pone.0167353

Mutter, G.L., Zahrieh, D., Liu, C., Neuberg, D., Finkelstein, D., Baker, H.E., Warrington, J.A., 2004. Comparison of frozen and RNALater solid tissue storage methods for use in RNA expression microarrays. BMC Genomics 5, 88. https://doi.org/10.1186/1471-2164-5-88

Nadya, S., Delaquis, P., Chen, J., Allen, K., Johnson, R.P., Ziebell, K., Laing, C., Gannon, V., Bach, S., Topp, E., 2016. Phenotypic and Genotypic Characteristics of Shiga Toxin-Producing Escherichia coli Isolated from Surface Waters and Sediments in a Canadian Urban-Agricultural Landscape. Front. Cell. Infect. Microbiol. 6. https://doi.org/10.3389/fcimb.2016.00036

Nakanishi, N., Nomoto, R., Tanaka, S., Arikawa, K., Iwamoto, T., 2019. Analysis of genetic characterization and clonality of legionella pneumophila isolated from cooling towers in japan. Int J Env. Res Public Health 16. https://doi.org/10.3390/ijerph16091664

National Center for Biotechnology Information, n.d. NCBI Pathogen Detection [WWW Document]. URL https://www.ncbi.nlm.nih.gov/pathogens/

Nayfach, S., Pollard, K.S., 2016. Toward Accurate and Quantitative Comparative Metagenomics. Cell 166, 1103–1116. https://doi.org/10.1016/j.cell.2016.08.007

Nazarian, E.J., Bopp, D.J., Saylors, A., Limberger, R.J., Musser, K.A., 2008. Design and implementation of a protocol for the detection of Legionella in clinical and environmental samples. Diagnostic Microbiology and Infectious Disease 62, 125–132. https://doi.org/10.1016/j.diagmicrobio.2008.05.004

Nemudryi, A., Nemudraia, A., Surya, K., Wiegand, T., Buyukyoruk, M., Wilkinson, R., Wiedenheft, B., 2020. Temporal detection and phylogenetic assessment of SARS-CoV-2 in municipal wastewater. medRxiv 2020.04.15.20066746. https://doi.org/10.1101/2020.04.15.20066746

New England Biolabs,Inc., 2020. NEBNext® UltraTM II DNA Library Prep Kit for Illumina®.

Ng, C., Tay, M., Tan, B., Le, T.-H., Haller, L., Chen, H., Koh, T.H., Barkham, T.M.S., Thompson, J.R., -H. Gin, K.Y., 2017. Characterization of Metagenomes in Urban Aquatic Compartments Reveals High Prevalence of Clinically Relevant Antibiotic Resistance Genes in Wastewaters. Front. Microbiol. 8. https://doi.org/10.3389/fmicb.2017.02200

Ng, P.C., Kirkness, E.F., 2010. Whole Genome Sequencing, in: Methods in Molecular Biology. pp. 215–226. https://doi.org/10.1007/978-1-60327-367-1_12

Ng, T.F.F., Marine, R., Wang, C., Simmonds, P., Kapusinszky, B., Bodhidatta, L., Oderinde, B.S., Wommack, K.E., Delwart, E., 2012. High variety of known and new RNA and DNA viruses of diverse origins in untreated sewage. J Virol 86, 12161–12175. https://doi.org/10.1128/JVI.00869-12

Niedringhaus, T.P., Milanova, D., Kerby, M.B., Snyder, M.P., Barron, A.E., 2011. Landscape of next-generation sequencing technologies. Anal Chem 83, 4327–4341. https://doi.org/10.1021/ac2010857

Niestepski, S., Harnisz, M., Korzeniewska, E., Guadalupe Aguilera-Arreola, M., Contreras-Rodriguez, A., Filipkowska, Z., Osinska, A., 2019. The emergence of antimicrobial resistance in environmental strains of the Bacteroides fragilis group. Environ. International 124, 408–419. https://doi.org/10.1016/j.envint.2018.12.056

Niestępski, S., Harnisz, M., Korzeniewska, E., Osińska, A., 2018. The occurrence of specific markers of Bacteroides fragilis group, B. dorei and antibiotic-resistance genes in the wastewater treatment plants. E3S Web of Conferences 44. https://doi.org/10.1051/E3SCONF/20184400124

Nocker, A., Cheung, C.-Y., Camper, A.K., 2006. Comparison of propidium monoazide with ethidium monoazide for differentiation of live vs. dead bacteria by selective removal of DNA from dead cells. J Microbiol Methods 67, 310–320. https://doi.org/10.1016/j.mimet.2006.04.015

Nurk, S., Meleshko, D., Korobeynikov, A., Pevzner, P.A., 2017. MetaSPAdes: A new versatile metagenomic assembler. Genome Research 27, 824–834. https://doi.org/10.1101/GR.213959.116/-/DC1

O'Brien, E., Nakyazze, J., Wu, H., Kiwanuka, N., Cunningham, W., Kaneene, J.B., Xagoraraki, I., 2017. Viral diversity and abundance in polluted waters in Kampala, Uganda. Water Res 127, 41–49. https://doi.org/10.1016/j.watres.2017.09.063

O'Brien, M., Rundell, Z.C., Nemec, M.D., Langan, L.M., Back, J.A., Lugo, J.N., 2021. A comparison of four commercially available RNA extraction kits for wastewater surveillance of SARS-CoV-2 in a college population. Science of The Total Environment 801, 149595. https://doi.org/10.1016/j.scitotenv.2021.149595

O'Leary, N.A., Wright, M.W., Brister, J.R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cox, E., Ermolaeva, O., Farrell, C.M., Goldfarb, T., Gupta, T., Haft, D., Hatcher, E., Hlavina, W., Joardar, V.S., Kodali, V.K., Li, W., Maglott, D., Masterson, P., McGarvey, K.M., Murphy, M.R., O'Neill, K., Pujar, S., Rangwala, S.H., Rausch, D., Riddick, L.D., Schoch, C., Shkeda, A., Storz, S.S., Sun, H., Thibaud-Nissen, F., Tolstoy, I., Tully, R.E., Vatsan, A.R., Wallin, C., Webb, D., Wu, W., Landrum, M.J., Kimchi, A., Tatusova, T., DiCuccio, M., Kitts, P., Murphy, T.D., Pruitt, K.D., 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 44, D733-745. https://doi.org/10.1093/nar/gkv1189

Obayomi, O., Ghazaryan, L., Ben-Hur, M., Edelstein, M., Vonshak, A., Safi, J., Bernstein, N., Gillor, O., 2019. The fate of pathogens in treated wastewater-soil-crops continuum and the effect of physical barriers. Sci Total Env. 681, 339–349. https://doi.org/10.1016/j.scitotenv.2019.04.378

Ogorzaly, L., Walczak, C., Galloux, M., Etienne, S., Gassilloud, B., Cauchie, H.-M., 2015. Human Adenovirus Diversity in Water Samples Using a Next-Generation Amplicon Sequencing Approach. Food Env. Virol. https://doi.org/10.1007/s12560-015-9194-4

Okazaki, Y., Nishimura, Y., Yoshida, T., Ogata, H., Nakano, S.-I., 2019. Genome-resolved viral and cellular metagenomes revealed potential key virus-host interactions in a deep freshwater lake. https://doi.org/10.1101/655167

Oksanen J, Simpson G, Blanchet F, Kindt R, Legendre P, Minchin P, O'Hara R, Solymos P, Stevens M, Szoecs E, Wagner H, Barbour M, Bedward M, Bolker B, Borcard D, Carvalho G, Chirico M, De Caceres M, Durand S, Evangelista H, FitzJohn R, Friendly M, Furneaux B, Hannigan G, Hill M, Lahti L, McGlinn D, Ouellette M, Ribeiro Cunha E, Smith T, Stier A, Ter Braak C, Weedon J (2022). _vegan: Community Ecology Package. R package version 2.6-2, <https://CRAN.R-project.org/package=vegan>.

Oshiki, M., Miura, T., Kazama, S., Segawa, T., Ishii, S., Hatamoto, M., Yamaguchi, T., Kubota, K., Iguchi, A., Tagawa, T., Okubo, T., Uemura, S., Harada, H., Kobayashi, N., Araki, N., Sano, D., 2018. Microfluidic PCR Amplification and MiSeq Amplicon Sequencing Techniques for High-Throughput Detection and Genotyping of Human Pathogenic RNA Viruses in Human Feces, Sewage, and Oysters. Front Microbiol 9, 830. https://doi.org/10.3389/fmicb.2018.00830

Osunmakinde, C.O., Selvarajan, R., Mamba, B.B., Msagati, T.A.M., 2019. Profiling bacterial diversity and potential pathogens in wastewater treatment plants using high-throughput sequencing analysis. Microorganisms 7. https://doi.org/10.3390/microorganisms7110506

Otten, T.G., Graham, J.L., Harris, T.D., Dreher, T.W., 2016. Elucidation of Taste- and Odor-Producing Bacteria and Toxigenic Cyanobacteria in a Midwestern Drinking Water Supply Reservoir by Shotgun Metagenomic Analysis. Appl. Environ. Microbiol. 82, 5410–5420. https://doi.org/10.1128/aem.01334-16

Overbeek, R., 2005. The Subsystems Approach to Genome Annotation and its Use in the Project to Annotate 1000 Genomes. Nucleic Acids Res. 33, 5691–5702. https://doi.org/10.1093/nar/gki866

Oxford Nanopore Technologies, 2019a. Nanopore Protocol, Genomic DNA by Ligation (No. SQK-LSK109).

Oxford Nanopore Technologies, 2019b. Nanopore Protocol, Rapid Sequencing (No. SQK-RAD004).

Oxford Nanopore Technologies, 2020. “At NCM, announcements include single-read accuracy of 99.1% on new chemistry and sequencing a record 10 Tb in a single PromethION run.” https://nanoporetech.com/about-us/news/ncm-announcements-include-single-read-accuracy-991-new-chemistry-and-sequencing

Oxford Nanopore Technologies, 2022. Nanopore Protocol, Rapid sequencing gDNA - low input PCR (No. SQK-PSK004).

PacBio, 2022. Preparing whole genome and metagenome libraries using SMRTbell prep kit 3.0.

Pal, C., Bengtsson-Palme, J., Kristiansson, E., Larsson, D.G.J., 2016. The structure and diversity of human, animal and environmental resistomes. Microbiome 4, 54. https://doi.org/10.1186/s40168-016-0199-5

Pal, C., Bengtsson-Palme, J., Rensing, C., Kristiansson, E., Larsson, D.G.J., 2014. BacMet: antibacterial biocide and metal resistance genes database. Nucleic Acids Res. 42, D737-43. https://doi.org/10.1093/nar/gkt1252

Palermo, C.N., Fulthorpe, R.R., Saati, R., Short, S.M., 2019. Metagenomic Analysis of Virus Diversity and Relative Abundance in a Eutrophic Freshwater Harbour. Viruses 11, 792. https://doi.org/10.3390/v11090792

Palomo, A., Jane Fowler, S., Gülay, A., Rasmussen, S., Sicheritz-Ponten, T., Smets, B.F., 2016. Metagenomic analysis of rapid gravity sand filter microbial communities suggests novel physiology of Nitrospira spp. ISME J 10, 2569–2581. https://doi.org/10.1038/ismej.2016.63

Paranjape, K., Bédard, É., Whyte, L.G., Ronholm, J., Prévost, M., Faucher, S.P., 2020. Presence of *Legionella* spp. in cooling towers: the role of microbial diversity, *Pseudomonas*, and continuous chlorine application. Water Res 169. https://doi.org/10.1016/j.watres.2019.115252

Park, S.-C., Won, S., 2018. Evaluation of 16S rRNA Databases for Taxonomic Assignments Using a Mock Community. Genomics Inform 16, e24. https://doi.org/10.5808/GI.2018.16.4.e24

Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., Tyson, G.W., 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 25, 1043–1055. https://doi.org/10.1101/gr.186072.114

Parra-Guardado, A.L., Sweeney, C.L., Hayes, E.K., Trueman, B.F., Huang, Y., Jamieson, R.C., Rand, J.L., Gagnon, G.A., Stoddart, A.K., 2021. Development of a rapid pre-concentration protocol and a magnetic beads-based RNA extraction method for SARS-CoV-2 detection in raw municipal wastewater. Environmental Science: Water Research & Technology 8, 47–61. https://doi.org/10.1039/D1EW00539A

Parsley, L.C., Consuegra, E.J., Kakirde, K.S., Land, A.M., Harper, W.F., Jr, Liles, M.R., 2010. Identification of diverse antimicrobial resistance determinants carried on bacterial, plasmid, or viral metagenomes from an activated sludge microbial assemblage. Appl Env. Microbiol 76, 3753–3757. https://doi.org/10.1128/AEM.03080-09

Pascault, N., Loux, V., Derozier, S., Martin, V., Debroas, D., Maloufi, S., Humbert, J.-F., Leloup, J., 2014. Technical challenges in metatranscriptomic studies applied to the bacterial communities of freshwater ecosystems. Genetica 143, 157–167. https://doi.org/10.1007/s10709-014-9783-4

Patnaik, A., Lepene, B., Barclay, A., 2020. Protocol for capture and concentration of viruses from wastewater samples (up to 50 mL) using Magnetic Nanotrap® particles [WWW Document]. URL https://www.protocols.io/view/protocol-for-capture-and-concentration-of-viruses-bkauksew

Pecson, B.M., Darby, E., Haas, C.N., Amha, Y.M., Bartolo, M., Danielson, R., Dearborn, Y., Di

Giovanni, G., Ferguson, C., Fevig, S., 2021. Reproducibility and sensitivity of 36 methods to quantify the SARS-CoV-2 genetic signal in raw wastewater: findings from an interlaboratory methods evaluation in the US. Environmental science: water research & technology 7, 504–520.

Pei, H., Xu, H., Wang, J., Jin, Y., Xiao, H., Ma, C., Sun, J., Li, H., 2017. 16S rRNA Gene Amplicon Sequencing Reveals Significant Changes in Microbial Compositions during Cyanobacteria-Laden Drinking Water Sludge Storage. Env. Sci Technol 51, 12774–12783. https://doi.org/10.1021/acs.est.7b03085

Pei, Y., Tao, C., Ling, Z., Yu, Z., Ji, J., Khan, A., Mamtimin, T., Liu, P., Li, X., 2020. Exploring novel Cr(VI) remediation genes for Cr(VI)-contaminated industrial wastewater treatment by comparative metatranscriptomics and metagenomics. Science of The Total Environment 742, 140435. https://doi.org/10.1016/j.scitotenv.2020.140435

Peng, Y., Leung, H.C.M., Yiu, S.M., Chin, F.Y.L., 2010. IDBA – A Practical Iterative de Bruijn Graph De Novo Assembler. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 6044 LNBI, 426–440. https://doi.org/10.1007/978-3-642-12683-3_28

Peng, Y., Leung, H.C.M., Yiu, S.M., Chin, F.Y.L., 2012. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics 28, 1420–1428. https://doi.org/10.1093/bioinformatics/bts174

Pereira, R.P.A., Peplies, J., Höfle, M.G., Brettar, I., 2017. Bacterial community dynamics in a cooling tower with emphasis on pathogenic bacteria and Legionella species using universal and genus-specific deep sequencing. Water Res 122, 363–376. https://doi.org/10.1016/j.watres.2017.06.011

Pereira, R.P.A., Peplies, J., Mushi, D., Brettar, I., Höfle, M.G., 2018. Pseudomonas-Specific NGS assay provides insight into abundance and dynamics of pseudomonas species including P. aeruginosa in a cooling tower. Front Microbiol 9, 1–15. https://doi.org/10.3389/fmicb.2018.01958

Peura, S., Sinclair, L., Bertilsson, S., Eiler, A., 2015. Metagenomic insights into strategies of aerobic and anaerobic carbon and nitrogen transformation in boreal lakes. Sci Rep 5, 12102. https://doi.org/10.1038/srep12102

Pinto, A.J., Marcus, D.N., Ijaz, U.Z., Bautista-de Lose Santos, Q.M., Dick, G.J., Raskin, L., 2016. Metagenomic Evidence for the Presence of Comammox Nitrospira-Like Bacteria in a Drinking Water System. mSphere 1. https://doi.org/10.1128/mSphere.00054-15

Pinto, A.J., Schroeder, J., Lunn, M., Sloan, W., Raskin, L., 2014. Spatial-temporal survey and occupancy-abundance modeling to predict bacterial community dynamics in the drinking water microbiome. MBio 5, e01135-14. https://doi.org/10.1128/mBio.01135-14

Pollard, M.O., Gurdasani, D., Mentzer, A.J., Porter, T., Sandhu, M.S., 2018. Long reads: Their purpose and place. Hum. Mol. Genet. 27, R234–R241. https://doi.org/10.1093/hmg/ddy177

Pope, P.B., Patel, B.K.C., 2008. Metagenomic analysis of a freshwater toxic cyanobacteria bloom. FEMS Microbiol. Ecol. 64, 9–27. https://doi.org/10.1111/j.1574-6941.2008.00448.x

Potgieter, S., Pinto, A., Sigudu, M., du Preez, H., Ncube, E., Venter, S., 2018. Long-term spatial and temporal microbial community dynamics in a large-scale drinking water distribution system with multiple disinfectant regimes. Water Res 139, 406–419. https://doi.org/10.1016/j.watres.2018.03.077

Proctor, C.R., Edwards, M.A., Pruden, A., 2015. Microbial composition of purified waters and implications for regrowth control in municipal water systems. Env. Sci Water Res Technol 1, 882–892. https://doi.org/10.1039/C5EW00134J

Prosser, J.I., 2010. Replicate or lie. Environmental Microbiology 12, 1806–1810. https://doi.org/10.1111/j.1462-2920.2010.02201.x

Pruden, A., Arabi, M., Storteboom, H.N., 2012. Correlation between upstream human activities and riverine antibiotic resistance genes. Environmental Science and Technology 46, 11541–11549. https://doi.org/10.1021/es302657r

Pruden, A., Bott, C., Blair, M. F., Miller, J. H., & Vaidya, R. (2020). *Characterization of Organic Carbon and Microbial Communities for the Optimization of Biologically-Active Carbon (BAC) Filtration for Potable Reuse.* Project 4872. Denver CO: The Water Research Foundation.

Pruden, A., Vikesland, P.J., Davis, B.C., de Roda Husman, A.M., 2021. Seizing the moment: now is the time for integrated global surveillance of antimicrobial resistance in wastewater environments. Current Opinion in Microbiology 64, 91–99. https://doi.org/10.1016/j.mib.2021.09.013

Prussin, A.J., Marr, L.C., Bibby, K.J., 2014. Challenges of studying viral aerosol metagenomics and communities in comparison with bacterial and fungal aerosols. FEMS Microbiol Lett 357, 1–9. https://doi.org/10.1111/1574-6968.12487

Pu, Y., Ngan, W.Y., Yao, Y., Habimana, O., 2019. Could benthic biofilm analyses be used as a reliable proxy for freshwater environmental health? ☆. https://doi.org/10.1016/j.envpol.2019.05.111

Putri, R.E., Kim, L.H., Farhat, N., Felemban, M., Saikaly, P.E., Vrouwenvelder, J.S., 2021. Evaluation of DNA extraction yield from a chlorinated drinking water distribution system. PLOS ONE 16, e0253799–e0253799. https://doi.org/10.1371/JOURNAL.PONE.0253799

Qin, H., Cui, L., Cao, X., Lv, Q., Chen, T., 2019. Evaluation of the Human Interference on the Microbial Diversity of Poyang Lake Using High-Throughput Sequencing Analyses. Int J Env. Res Public Health 16. https://doi.org/10.3390/ijerph16214218

Qin, K., Struewing, I., Domingo, J.W.S., Lytle, D., Lu, J., 2017. Opportunistic Pathogens and Microbial Communities and Their Associations with Sediment Physical Parameters in Drinking Water Storage Tank Sediments. Pathogens. https://doi.org/10.3390/pathogens6040054

Quail, M.A., Smith, M., Coupland, P., Otto, T.D., Harris, S.R., Connor, T.R., Bertoni, A., Swerdlow, H.P., Gu, Y., 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics 13, 341. https://doi.org/10.1186/1471-2164-13-341

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F.O., 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 41, D590-6. https://doi.org/10.1093/nar/gks1219

Quick, J., Cumley, N., Wearn, C.M., Niebel, M., Constantinidou, C., Thomas, C.M., Pallen, M.J., Moiemen, N.S., Bamford, A., Oppenheim, B., Loman, N.J., 2014. Seeking the source of Pseudomonas aeruginosa infections in a recently opened hospital: an observational study using whole-genome sequencing. BMJ Open 4, e006278. https://doi.org/10.1136/bmjopen-2014-006278

Quince, C., Delmont, T.O., Raguideau, S., Alneberg, J., Darling, A.E., Collins, G., Eren, A.M., 2017a. DESMAN: a new tool for de novo extraction of strains from metagenomes. Genome Biol 18, 181.

Quince, C., Walker, A.W., Simpson, J.T., Loman, N. J., Segata, N., 2017b. Shotgun metagenomics, from sampling to analysis. Nat Biotechnol 35, 833–844. https://doi.org/10.1038/nbt.3935

Radomski, N., Lucas, F.S., Moilleron, R., Cambau, E., Haenn, S. and Moulin, L. 2010. Development of a real-time qPCR method for detection and enumeration of *Mycobacterium* spp. in surface water. Appl Environ Microb 76(21), 7348-7351.

Rahman, S.F., Kantor, R.S., Huddy, R., Thomas, B.C., van Zyl, A.W., Harrison, S.T.L., Banfield, J.F., 2017. Genome-resolved metagenomics of a bioremediation system for degradation of thiocyanate in mine water containing suspended solid tailings. MicrobiologyOpen 6, e00446. https://doi.org/10.1002/mbo3.446

Ramos-Mandujano, G., Salunke, R., Mfarrej, S., Rachmadi, A.T., Hala, S., Xu, J., Alofi, F.S., Khogeer, A., Hashem, A.M., Almontashiri, N.A.M., Alsomali, A., Shinde, D.B., Hamdan, S., Hong, P.-Y., Pain, A., Li, M., 2021. A Robust, Safe, and Scalable Magnetic Nanoparticle Workflow for RNA Extraction of Pathogens from Clinical and Wastewater Samples. Global Challenges 5, 2000068. https://doi.org/10.1002/gch2.202000068

Randle-Boggis, R.J., Helgason, T., Sapp, M., Ashton, P.D., 2016. Evaluating techniques for metagenome annotation using simulated sequence data. FEMS Microbiol Ecol 92. https://doi.org/10.1093/femsec/fiw095

Raphael, B.H., Baker, D.J., Nazarian, E., Lapierre, P., Bopp, D., Kozak-Muiznieks, N.A., Morrison, S.S., Lucas, C.E., Mercante, J.W., Musser, K.A., Winchell, J.M., 2016. Genomic Resolution of Outbreak-Associated *Legionella pneumophila* Serogroup 1 Isolates from New York State. Appl. Environ. Microbiol. 82, 3582–3590. https://doi.org/10.1128/AEM.00362-16

Raskin, L.; Dowdell, K.; Haig, S.; Dai, D.; Edwards, M. A.; Pruden, A. 2022. *Methods for Detecting and Differentiating Opportunistic Premise Plumbing Pathogens to Determine Efficacy of Control and Treatment Technologies.* Project 4721. Denver, CO: The Water Research Foundation.

Ratnapradipa, D., Cardinal, C., Ratnapradipa, K., Scarborough, A. and Xie, Y. 2018. Implications of Hurricane Harvey on Environmental Public Health in Harris County, Texas. Journal of Environmental Health 81, 24-32.

Raynor, P.C., Adesina, A., Aboubakr, H.A., Yang, M., Torremorell, M., and Gooyal S.M. 2021. Comparison of samplers collecting airborne influenza viruses: 1. Primarily impingers and cyclones. PLoS ONE 16 (1): e0244977. DOI: 10.1371/journal.pone.0244977

Rehman, Z.U., Ali, M., Iftikhar, H., Leiknes, T., 2019. Genome-resolved metagenomic analysis reveals roles of microbial community members in full-scale seawater reverse osmosis plant. Water Res 149, 263–271. https://doi.org/10.1016/j.watres.2018.11.012

Reid, T., Chaganti, S.R., Droppo, I.G., Weisener, C.G., 2018. Novel insights into freshwater hydrocarbon-rich sediments using metatranscriptomics: Opening the black box. Water Res 136, 1–11. https://doi.org/10.1016/j.watres.2018.02.039

Reis, A.L.M., Deveson, I.W., Wong, T., Madala, B.S., Barker, C., Blackburn, J., Marcellin, E., Mercer, T.R., 2020. A universal and independent synthetic DNA ladder for the quantitative measurement of genomic features. Nat Commun 11, 3609. https://doi.org/10.1038/s41467-020-17445-5

Reis, M.P., Dias, M.F., Costa, P.S., Ávila, M.P., Leite, L.R., de Araújo, F.M.G., Salim, A.C.M., Bucciarelli-Rodriguez, M., Oliveira, G., Chartone-Souza, E., Nascimento, A.M.A., 2016. Metagenomic signatures of a tropical mining-impacted stream reveal complex microbial and metabolic networks. Chemosphere 161, 266–273. https://doi.org/10.1016/j.chemosphere.2016.06.097

Reiss, R.A., Guerra, P., Makhnin, O., 2016. Metagenome phylogenetic profiling of microbial community evolution in a tetrachloroethene-contaminated aquifer responding to enhanced reductive dechlorination protocols. Stand. Genomic Sci. 11. https://doi.org/10.1186/s40793-016-0209-z

Renshaw, M.A., Olds, B.P., Jerde, C.L., McVeigh, M.M., Lodge, D.M., 2015. The room temperature preservation of filtered environmental DNA samples and assimilation into a phenol–chloroform–isoamyl alcohol DNA extraction. Mol Ecol Resour 15, 168–176. https://doi.org/10.1111/1755-0998.12281

Reza, M.S., Mizusawa, N., Kumano, A., Oikawa, C., Ouchi, D., Kobiyama, A., Yamada, Y., Ikeda, Y., Ikeda, D., Ikeo, K., Sato, S., Ogata, T., Kudo, T., Jimbo, M., Yasumoto, K., Yoshitake, K., Watabe, S., 2018. Metagenomic analysis using 16S ribosomal RNA genes of a bacterial community in an urban stream, the Tama River, Tokyo. Fish. Sci. 84, 563–577. https://doi.org/10.1007/s12562-018-1193-6

Rice, E.W., Baird, R.B., Eaton, A.D. and Clesceri, L.S. (2012) Standard methods for the examination of water and wastewater, American Public Health Association Washington, DC.

Rimoldi, S.G., Stefani, F., Gigantiello, A., Polesello, S., Comandatore, F., Mileto, D., Maresca, M., Longobardi, C., Mancon, A., Romeri, F., Pagani, C., Moja, L., Gismondo, M.R., Salerno, F., 2020. Presence and vitality of SARS-CoV-2 virus in wastewaters and rivers. medRxiv 2020.05.01.20086009. https://doi.org/10.1101/2020.05.01.20086009

Roberto, A.A., Van Gray, J.B., Engohang-Ndong, J., Leff, L.G., 2019. Distribution and co-occurrence of antibiotic and metal resistance genes in biofilms of an anthropogenically impacted stream. https://doi.org/10.1016/j.scitotenv.2019.06.053

Rodriguez-Brito, B., Li, L., Wegley, L., Furlan, M., Angly, F., Breitbart, M., Buchanan, J., Desnues, C., Dinsdale, E., Edwards, R., Felts, B., Haynes, M., Liu, H., Lipson, D., Mahaffy, J., Martin-Cuadrado, A.B., Mira, A., Nulton, J., Pasić, L., Rayhawk, S., Rodriguez-Mueller, J., Rodriguez-Valera, F., Salamon, P., Srinagesh, S., Thingstad, T.F., Tran, T., Thurber, R.V., Willner, D., Youle, M., Rohwer, F., 2010. Viral and microbial community dynamics in four aquatic environments. ISME J 4, 739–751. https://doi.org/10.1038/ismej.2010.1

Rodriguez-R, L.M., Konstantinidis, K.T., 2014. Estimating coverage in metagenomic data sets and why it matters. ISME J 8, 2349–2351. https://doi.org/10.1038/ismej.2014.76

Roeselers, G., Coolen, J., van der Wielen, P.W.J.J., Jaspers, M.C., Atsma, A., de Graaf, B., Schuren, F., 2015. Microbial biogeography of drinking water: Patterns in phylogenetic diversity across space and time. Env. Microbiol 17, 2505–2514. https://doi.org/10.1111/1462-2920.12739

Rognes, T., Flouri, T., Nichols, B., Quince, C., Mahé, F., 2016. VSEARCH: A versatile open source tool for metagenomics. PeerJ 2016, 1–22. https://doi.org/10.7717/peerj.2584

Rohwer, F., Edwards, R., 2002. The Phage Proteomic Tree: a Genome-Based Taxonomy for Phage. J Bacteriol 184, 4529–4535. https://doi.org/10.1128/jb.184.16.4529-4535.2002

Rosario, K., Nilsson, C., Lim, Y.W., Ruan, Y., Breitbart, M., 2009. Metagenomic analysis of viruses in reclaimed water. Env. Microbiol 11, 2806–2820. https://doi.org/10.1111/j.1462-2920.2009.01964.x

Rose, R., Constantinides, B., Tapinos, A., Robertson, D.L., Prosperi, M., 2016. Challenges in the analysis of viral metagenomes. Virus Evol 2. https://doi.org/10.1093/ve/vew022

Rosso, G.E., Muday, J.A., Curran, J.F., 2018. Tools for Metagenomic Analysis at Wastewater

Treatment Plants: Application to a Foaming Episode. Proc. Water Environ. Fed. 2018, 1398–1417. https://doi.org/10.2175/193864718825137980

Rowe, W., Verner-Jeffreys, D.W., Baker-Austin, C., Ryan, J.J., Maskell, D.J., Pearce, G.P., 2016. Comparative metagenomics reveals a diverse range of antimicrobial resistance genes in effluents entering a river catchment. Water Sci. Technol. 73, 1541–1549. https://doi.org/10.2166/wst.2015.634

Roy, M.A., Arnaud, J.M., Jasmin, P.M., Hamner, S., Hasan, N.A., Colwell, R.R., Ford, T.E., 2018. A Metagenomic Approach to Evaluating Surface Water Quality in Haiti. Int J Env. Res Public Health 15. https://doi.org/10.3390/ijerph15102211

RStudio Team (2019). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/.

Ruggieri, J., Kemp, R., Forman, S., Van Eden, M.E., 2016. Techniques for Nucleic Acid Purification from Plant, Animal, and Microbial Samples 41–52. https://doi.org/10.1007/978-1-4939-3185-9_4

Ruiz-Moreno, H.A., López-Tamayo, A.M., Caro-Quintero, A., Husserl, J., Barrios, A.F.G., 2019. Metagenome level metabolic network reconstruction analysis reveals the microbiome in the Bogotá River is functionally close to the microbiome in produced water. Ecol. Model. 399, 1–12. https://doi.org/10.1016/j.ecolmodel.2019.02.001

Runcharoen, C., Moradigaravand, D., Blane, B., Paksanont, S., Thammachote, J., Anun, S., Parkhill, J., Chantratita, N., Peacock, S.J., 2017. Whole genome sequencing reveals high-resolution epidemiological links between clinical and environmental Klebsiella pneumoniae. Genome Med 9, 6. https://doi.org/10.1186/s13073-017-0397-1

Saingam, P., Li, B., Yan, T., 2018. Use of amplicon sequencing to improve sensitivity in PCR-based detection of microbial pathogen in environmental samples. J Microbiol Methods 149, 73–79. https://doi.org/10.1016/j.mimet.2018.05.005

Sakcham, B., Kumar, A., Cao, B., 2019. Extracellular DNA in Monochloraminated Drinking Water and Its Influence on DNA-Based Profiling of a Microbial Community. Environmental Science and Technology Letters 6, 306–312. https://doi.org/10.1021/ACS.ESTLETT.9B00185/SUPPL_FILE/EZ9B00185_SI_001.PDF

Saleem, F., Kamran Azim, M., Mustafa, A., Kori, J.A., Hussain, M.S., 2019. Metagenomic profiling of fresh water lakes at different altitudes in Pakistan. Ecol. Inform. 51, 73–81. https://doi.org/10.1016/j.ecoinf.2019.02.013

Saleem, F., Mustafa, A., Kori, J.A., Hussain, M.S., Kamran Azim, M., 2018. Metagenomic Characterization of Bacterial Communities in Drinking Water Supply System of a Mega City. Microb. Ecol. 76, 899–910. https://doi.org/10.1007/s00248-018-1192-2

Salter, S.J., Cox, M.J., Turek, E.M., Calus, S.T., Cookson, W.O., Moffatt, M.F., Turner, P., Parkhill, J., Loman, N.J., Walker, A.W., 2014. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biology 12, 1–12. https://doi.org/10.1186/S12915-014-0087-Z/FIGURES/4

Sánchez-Reyez, A., Batista-García, R.A., Valdés-García, G., Ortiz, E., Perezgasga, L., Zárate-Romero, A., Pastor, N., Folch-Mallol, J.L., 2017. A family 13 thioesterase isolated from an activated sludge metagenome: Insights into aromatic compounds metabolism. Proteins Struct. Funct. Bioinforma. 85, 1222–1237. https://doi.org/10.1002/prot.25282

Sanz, J.L., Köchling, T., 2019. Next-generation sequencing and waste/wastewater treatment: a comprehensive overview. Rev. Environ. Sci. Biotechnol. 18, 635–680. https://doi.org/10.1007/s11157-019-09513-0

Sato, M.P., Ogura, Y., Nakamura, K., Nishida, R., Gotoh, Y., Hayashi, M., Hisatsune, J., Sugai, M., Takehiko, I., Hayashi, T., 2019a. Comparison of the sequencing bias of currently available library preparation kits for Illumina sequencing of bacterial genomes and metagenomes. DNA Research 26, 391–398. https://doi.org/10.1093/dnares/dsz017

Sato, Y., Mizuyama, M., Sato, M., Minamoto, T., Kimura, R., Toma, C., 2019b. Environmental DNA metabarcoding to detect pathogenic *Leptospira* and associated organisms in leptospirosis-endemic areas of Japan. Sci Rep 9, 6575. https://doi.org/10.1038/s41598-019-42978-1

Savin, M., Bierbaum, G., Hammerl, J.A., Heinemann, C., Parcina, M., Sib, E., Voigt, A., Kreyenschmidt, J., Björkroth, J., 2020. ESKAPE Bacteria and Extended-Spectrum-Lactamase-Producing Escherichia coli Isolated from Wastewater and Process Water from German Poultry Slaughterhouses Downloaded from. https://doi.org/10.1128/AEM

Saxena, G., Mitra, S., Marzinelli, E.M., Xie, C., Wei, T.J., Steinberg, P.D., Williams, R.B.H., Kjelleberg, S., Lauro, F.M., Swarup, S., 2018. Metagenomics Reveals the Influence of Land Use and Rain on the Benthic Microbial Communities in a Tropical Urban Waterway. mSystems 3. https://doi.org/10.1128/msystems.00136-17

Scaturro, M., Fontana, S., Dell'eva, I., Helfer, F., Marchio, M., Stefanetti, M.V., Cavallaro, M., Miglietta, M., Montagna, M.T., De Giglio, O., Cuna, T., Chetti, L., Sabattini, M.A.B., Carlotti, M., Viggiani, M., Stenico, A., Romanin, E., Bonanni, E., Ottaviano, C., Franzin, L., Avanzini, C., Demarie, V., Corbella, M., Cambieri, P., Marone, P., Rota, M.C., Bella, A., Ricci, M.L., 2016. A multicenter study of viable PCR using propidium monoazide to detect Legionella in water samples. Diagn Microbiol Infect Dis 85, 283–288. https://doi.org/10.1016/j.diagmicrobio.2016.04.009

Schloss, P.D., Handelsman, J., 2005. Metagenomics for studying unculturable microorganisms: cutting the Gordian knot. Genome Biol. 6, 229. https://doi.org/10.1186/gb-2005-6-8-229

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., Sahl, J.W., Stres, B., Thallinger, G.G., Horn, D.J.V., Weber, C.F., 2009. Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. Appl. Environ. Microbiol. 75, 7537–7541. https://doi.org/10.1128/AEM.01541-09

Schlüter, A., Krause, L., Szczepanowski, R., Goesmann, A., Pühler, A., 2008. Genetic diversity and composition of a plasmid metagenome from a wastewater treatment plant. J Biotechnol 136, 65–76. https://doi.org/10.1016/j.jbiotec.2008.03.017

Sczyrba, A., Hofmann, P., Belmann, P., Koslicki, D., Janssen, S., Dröge, J., Gregor, I., Majda, S., Fiedler, J., Dahms, E., Bremges, A., Fritz, A., Garrido-Oter, R., Jørgensen, T.S., Shapiro, N., Blood, P.D., Gurevich, A., Bai, Y., Turaev, D., DeMaere, M.Z., Chikhi, R., Nagarajan, N., Quince, C., Meyer, F., Balvočiūtė, M., Hansen, L.H., Sørensen, S.J., Chia, B.K.H., Denis, B., Froula, J.L., Wang, Z., Egan, R., Don Kang, D., Cook, J.J., Deltel, C., Beckstette, M., Lemaitre, C., Peterlongo, P., Rizk, G., Lavenier, D., Wu, Y.-W., Singer, S.W., Jain, C., Strous, M., Klingenberg, H., Meinicke, P., Barton, M.D., Lingner, T., Lin, H.-H., Liao, Y.-C., Silva, G.G.Z., Cuevas, D.A., Edwards, R.A., Saha, S., Piro, V.C., Renard, B.Y., Pop, M., Klenk, H.-P., Göker, M., Kyrpides, N.C., Woyke, T., Vorholt, J.A., Schulze-Lefert, P., Rubin, E.M., Darling, A.E., Rattei, T., McHardy, A.C., 2017. Critical Assessment of Metagenome Interpretation—a benchmark of metagenomics software. Nat Methods 14, 1063–1071. https://doi.org/10.1038/nmeth.4458

Shakya, M., Lo, C.-C., Chain, P.S.G., 2019. Advances and Challenges in Metatranscriptomic Analysis. Front Genet 10:904. https://doi.org/10.3389/fgene.2019.00904

Shapiro, E., Biezuner, T., Linnarsson, S., 2013. Single-cell sequencing-based technologies will revolutionize whole-organism science. Nat. Rev. Genet. 14, 618–630. https://doi.org/10.1038/nrg3542

Shaw, J.L.A., Monis, P., Weyrich, L.S., Sawade, E., Drikas, M., Cooper, A.J., 2015. Using Amplicon Sequencing To Characterize and Monitor Bacterial Diversity in Drinking Water Distribution Systems. Appl Env. Microbiol 81, 6463–6473. https://doi.org/10.1128/AEM.01297-15

Shendure, J., Ji, H., 2008. Next-generation DNA sequencing. Nat Biotechnol 26, 1135–1145. https://doi.org/10.1038/nbt1486

Shi, P., Jia, S., Zhang, X.-X., Zhang, T., Cheng, S., Li, A., 2013. Metagenomic insights into chlorination effects on microbial antibiotic resistance in drinking water. Water Res 47, 111–120. https://doi.org/10.1016/j.watres.2012.09.046

Shin, J.H., 2018. Nucleic Acid Extraction and Enrichment. Advanced Techniques in Diagnostic Microbiology 273–273. https://doi.org/10.1007/978-3-319-33900-9_13

Shokralla, S., Spall, J.L., Gibson, J.F., Hajibabaei, M., 2012. Next-generation sequencing technologies for environmental DNA research. Mol Ecol 21, 1794–1805. https://doi.org/10.1111/j.1365-294X.2012.05538.x

Shrestha, R.G., Tanaka, Y., Malla, B., Bhandari, D., Tandukar, S., Inoue, D., Sei, K., Sherchand, J.B., Haramoto, E., 2017. Next-generation sequencing identification of pathogenic bacterial genes and their relationship with fecal indicator bacteria in different water sources in the Kathmandu Valley, Nepal. Sci. Total Environ. 601–602, 278–284. https://doi.org/10.1016/j.scitotenv.2017.05.105

Shrestha, R.G., Tandukar, S., Bhandari, D., Sherchan, S.P., Tanaka, Y., Sherchand, J.B., Haramoto, E., 2019. Prevalence of Arcobacter and other pathogenic bacteria in river water in Nepal. Water 11, 13–16. https://doi.org/10.3390/w11071416

Sible, E., Cooper, A., Malki, K., Bruder, K., Watkins, S.C., Fofanov, Y., Putonti, C., 2015. Survey of viral populations within Lake Michigan nearshore waters at four Chicago area beaches. Data Brief 5, 9–12. https://doi.org/10.1016/j.dib.2015.08.001

Sidhu, C., Vikram, S., Pinnaka, A.K., 2017. Unraveling the Microbial Interactions and Metabolic Potentials in Pre- and Post-treated Sludge from a Wastewater Treatment Plant Using Metagenomic Studies. Front. Microbiol. 8, 1382. https://doi.org/10.3389/fmicb.2017.01382

Siguier, P., Perochon, J., Lestrade, L., Mahillon, J., Chandler, M., 2006. ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res. 34, D32-6. https://doi.org/10.1093/nar/gkj014

Silva, C.C., Hayden, H., Sawbridge, T., Mele, P., De Paula, S.O., Silva, L.C.F., Vidigal, P.M.P., Vicentini, R., Sousa, M.P., Torres, A.P.R., Santiago, V.M.J., Oliveira, V.M., 2013. Identification of Genes and Pathways Related to Phenol Degradation in Metagenomic Libraries from Petroleum Refinery Wastewater. PLoS ONE 8, e61811–e61811. https://doi.org/10.1371/journal.pone.0061811

Silva, C.C., Hayden, H., Sawbridge, T., Mele, P., Kruger, R.H., Rodrigues, M.V.N., Costa, G.G.L., Vidal, R.O., Sousa, M.P., Torres, A.P.R., Santiago, V.M.J., Oliveira, V.M., 2012. Phylogenetic and functional diversity of metagenomic libraries of phenol degrading sludge from petroleum refinery wastewater treatment system. AMB Express 2, 18. https://doi.org/10.1186/2191-0855-2-18

Simmonds, P., 2015. Methods for virus classification and the challenge of incorporating metagenomic sequence data. J Gen Virol 96, 1193–1206. https://doi.org/10.1099/vir.0.000016

Simmons, K., 2016. Operating Procedure Surface Water Sampling.

Simpson, A., Topol, A., White, B.J., Wolfe, M.K., Wigginton, K.R., Boehm, A.B., 2021. Effect of storage conditions on SARS-CoV-2 RNA quantification in wastewater solids. PeerJ 9, e11933. https://doi.org/10.7717/peerj.11933

Singh, A., Chauhan, N.S., Thulasiram, H.V., Taneja, V., Sharma, R., 2010. Identification of two flavin monooxygenases from an effluent treatment plant sludge metagenomic library. Bioresour. Technol. 101, 8481–8484. https://doi.org/10.1016/j.biortech.2010.06.025

Skvortsov, T., de Leeuwe, C., Quinn, J.P., McGrath, J.W., Allen, C.C.R., McElarney, Y., Watson, C., Arkhipova, K., Lavigne, R., Kulakov, L.A., 2016. Metagenomic Characterisation of the Viral Community of Lough Neagh, the Largest Freshwater Lake in Ireland. PLoS One 11, e0150361. https://doi.org/10.1371/journal.pone.0150361

Smith, J., Banik, S. and Haque, U. 2018. Catastrophic hurricanes and public health dangers: lesson learned. Journal of Public Health and Emergency 2(2), 1-3.

Smith, R.J., Jeffries, T.C., Roudnew, B., Fitch, A.J., Seymour, J.R., Delpin, M.W., Newton, K., Brown, M.H., Mitchell, J.G., 2012. Metagenomic comparison of microbial communities inhabiting confined and unconfined aquifer ecosystems. Env. Microbiol 14, 240–253. https://doi.org/10.1111/j.1462-2920.2011.02614.x

Somvanshi, S., Kharat, P., Saraf, T., Somwanshi, S., Shejul, S., Jadhav, K., 2020. Multifunctional nano-magnetic particles assisted viral RNA-extraction protocol for potential detection of COVID-19. Materials Research Innovations 1–1. https://doi.org/10.1080/14328917.2020.1769350

Song, Z., Chen, S., Zhao, F., Zhu, W., 2019. Whole metagenome of injected and produced fluids reveal the heterogenetic characteristics of the microbial community in a water-flooded oil reservoir. J. Pet. Sci. Eng. 176, 1198–1207. https://doi.org/10.1016/j.petrol.2019.02.008

Sonthiphand, P., Ruangroengkulrith, S., Mhuantong, W., Charoensawan, V., Chotpantarat, S., Boonkaewwan, S., 2019. Metagenomic insights into microbial diversity in a groundwater basin impacted by a variety of anthropogenic activities. Environ. Sci. Pollut. Res. 26, 26765–26781. https://doi.org/10.1007/s11356-019-05905-5

Spencer, S.J., Tamminen, M.V., Preheim, S.P., Guo, M.T., Briggs, A.W., Brito, I.L., A Weitz, D., Pitkänen, L.K., Vigneault, F., Virta, M.P., Alm, E.J., 2016. Massively parallel sequencing of single cells by epicPCR links functional genes with phylogenetic markers. ISME J 10, 427–436. https://doi.org/10.1038/ismej.2015.124

Staley, J., 1985. Measurement of In Situ Activities of Nonphotosynthetic Microorganisms in Aquatic and Terrestrial Habitats. Annu. Rev. Microbiol. 39, 321–346. https://doi.org/10.1146/annurev.micro.39.1.321

Stamps, B.W., Leddy, M.B., Plumlee, M.H., Hasan, N.A., Colwell, R.R., Spear, J.R., 2018. Characterization of the Microbiome at the World’s Largest Potable Water Reuse Facility. Front. Microbiol. 9. https://doi.org/10.3389/fmicb.2018.02435

Stamps, B.W., Spear, J.R., 2020. Identification of Metagenome-Assembled Genomes Containing Antimicrobial Resistance Genes, Isolated from an Advanced Water Treatment Facility. Microbiol Resour Announc 9. https://doi.org/10.1128/MRA.00003-20

Steffen, M.M., Belisle, B.S., Watson, S.B., Boyer, G.L., Bourbonniere, R.A., Wilhelm, S.W., 2015. Metatranscriptomic evidence for co-occurring top-down and bottom-up controls on toxic cyanobacterial communities. Appl Env. Microbiol 81, 3268–3276.

https://doi.org/10.1128/AEM.04101-14

Steffen, M.M., Li, Z., Effler, T.C., Hauser, L.J., Boyer, G.L., Wilhelm, S.W., 2012. Comparative metagenomics of toxic freshwater cyanobacteria bloom communities on two continents. PLoS One 7, e44002. https://doi.org/10.1371/journal.pone.0044002

Strubbia, S., Phan, M.V.T., Schaeffer, J., Koopmans, M., Cotten, M., Le Guyader, F.S., 2019a. Characterization of Norovirus and Other Human Enteric Viruses in Sewage and Stool Samples Through Next-Generation Sequencing. Food Env. Virol 11, 400–409. https://doi.org/10.1007/s12560-019-09402-3

Strubbia, S., Schaeffer, J., Oude Munnink, B.B., Besnard, A., Phan, M.V.T., Nieuwenhuijse, D.F., de Graaf, M., Schapendonk, C.M.E., Wacrenier, C., Cotten, M., Koopmans, M.P.G., Le Guyader, F.S., 2019b. Metavirome Sequencing to Evaluate Norovirus Diversity in Sewage and Related Bioaccumulated Oysters. Front Microbiol 10, 2394. https://doi.org/10.3389/fmicb.2019.02394

Suffredini, E., Iaconelli, M., Equestre, M., Valdazo-González, B., Ciccaglione, A.R., Marcantonio, C., Della Libera, S., Bignami, F., La Rosa, G., 2018. Genetic Diversity Among Genogroup II Noroviruses and Progressive Emergence of GII.17 in Wastewaters in Italy (2011-2016) Revealed by Next-Generation and Sanger Sequencing. Food Env. Virol 10, 141–150. https://doi.org/10.1007/s12560-017-9328-y

Sul, W.-J., Kim, I.-S., Ekpeghere, K.I., Song, B., Kim, B.-S., Kim, H.-G., Kim, J.-T., Koh, S.-C., 2016. Metagenomic insight of nitrogen metabolism in a tannery wastewater treatment plant bioaugmented with the microbial consortium BM-S-1. J. Environ. Sci. Health Part A 51, 1164–1172. https://doi.org/10.1080/10934529.2016.1206387

Sun, Y., Guan, Y., Zeng, D., He, K., Wu, G., 2018. Metagenomics-based interpretation of AHLs-mediated quorum sensing in Anammox biofilm reactors for low-strength wastewater treatment. Chem. Eng. J. 344, 42–52. https://doi.org/10.1016/j.cej.2018.03.047

Suttner, B., Johnston, E.R., Orellana, L.H., Rodriguez-R, L.M., Hatt, J.K., Carychao, D., Carter, M.Q., Cooley, M.B., Konstantinidis, K.T., 2020. Metagenomics as a public health risk assessment tool in a study of natural creek sediments influenced by agricultural and livestock runoff: Potential and limitations. Applied and Environmental Microbiology 86. https://doi.org/10.1128/AEM.02525-19

Suzek, B.E., Huang, H., McGarvey, P., Mazumder, R., Wu, C.H., 2007. UniRef: comprehensive and non-redundant UniProt reference clusters. Bioinformatics 23, 1282–1288. https://doi.org/10.1093/bioinformatics/btm098

Suzuki, M.T., Taylor, L.T., DeLong, E.F., 2000. Quantitative analysis of small-subunit rRNA genes in mixed microbial populations via 5’-nuclease assays. Applied and Environmental Microbiology 66, 4605–4614. https://doi.org/10.1128/AEM.66.11.4605-4614.2000

Szczepanowski, R., Bekel, T., Goesmann, A., Krause, L., Krömeke, H., Kaiser, O., Eichler, W., Pühler, A., Schlüter, A., 2008. Insight into the plasmid metagenome of wastewater treatment plant bacteria showing reduced susceptibility to antimicrobial drugs analysed by the 454-pyrosequencing technology. J. Biotechnol. 136, 54–64. https://doi.org/10.1016/j.jbiotec.2008.03.020

Tamaki, H., Zhang, R., Angly, F.E., Nakamura, S., Hong, P.-Y., Yasunaga, T., Kamagata, Y., Liu, W.-T., 2012. Metagenomic analysis of DNA viruses in a wastewater treatment plant in tropical climate. Env. Microbiol 14, 441–452. https://doi.org/10.1111/j.1462-2920.2011.02630.x

Tamminen, M., Spaak, J., Caduff, L., Schiff, H., Lang, R., Schmid, S., Montealegre, M.C., Julian, T.R., 2020. Digital multiplex ligation assay for highly multiplexed screening of β-lactamase-encoding genes in bacterial isolates. Commun Biol 3, 1–6. https://doi.org/10.1038/s42003-020-0980-7

Tan, S.C., Yiap, B.C., 2009. DNA, RNA, and protein extraction: the past and the present. Journal of biomedicine & biotechnology 2009. https://doi.org/10.1155/2009/574398

Tanaka, N., Takahara, A., Hagio, T., Nishiko, R., Kanayama, J., Gotoh, O., Mori, S., 2020. Sequencing artifacts derived from a library preparation method using enzymatic fragmentation. PLOS ONE 15, e0227427. https://doi.org/10.1371/journal.pone.0227427

Tang, J., Bu, Y., Zhang, X.-X., Huang, K., He, X., Ye, L., Shan, Z., Ren, H., 2016. Metagenomic analysis of bacterial community composition and antibiotic resistance genes in a wastewater treatment plant and its receiving surface water. Ecotoxicol. Environ. Saf. 132, 260–269. https://doi.org/10.1016/j.ecoenv.2016.06.016

Tansirichaiya, S., Rahman, M.A., Roberts, A.P., 2019. The Transposon Registry. Mob DNA 10, 40. https://doi.org/10.1186/s13100-019-0182-3

Tap, J., Cools-Portier, S., Pavan, S., Druesne, A., Ohman, L., Tornblom, H., Simren, M., Derrien, M., 2019. Effects of the long-term storage of human fecal microbiota samples collected in RNAlater. 9:601. https://doi.org/10.1038/s41598-018-36953-5

Tatusov, R.L., Galperin, M.Y., Natale, D.A., Koonin, E.V., 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. 28, 33–36.

Tavernier, S., Coenye, T., 2015. Quantification of Pseudomonas aeruginosa in multispecies biofilms using PMA-qPCR. PeerJ 3, e787. https://doi.org/10.7717/peerj.787

Taylor, M.J., Bentham, R.H., Ross, K.E., 2014. Limitations of Using Propidium Monoazide with qPCR to Discriminate between Live and Dead Legionella in Biofilm Samples. Microbiol Insights 7, 15–24. https://doi.org/10.4137/MBI.S17723

The dMIQE Group, Huggett, J.F., 2020. The digital MIQE guidelines update: minimum information for publication of quantitative digital PCR experiments for 2020. Clinical Chemistry 66, 1012–1029. https://doi.org/10.1093/clinchem/hvaa125

The UnitProt Consortium, T.U., 2019. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res 47, D506–D515. https://doi.org/10.1093/nar/gky1049

Tian, M., Zhao, F., Shen, X., Chu, K., Wang, J., Chen, S., Guo, Y., Liu, H., 2015. The first metagenome of activated sludge from full-scale anaerobic/anoxic/oxic (A2O) nitrogen and phosphorus removal reactor using Illumina sequencing. J. Environ. Sci. 35, 181–190. https://doi.org/10.1016/j.jes.2014.12.027

Tomazetto, G., Wibberg, D., Schlüter, A., Oliveira, V.M., 2015. New FeFe-hydrogenase genes identified in a metagenomic fosmid library from a municipal wastewater treatment plant as revealed by high-throughput sequencing. Res. Microbiol. 166, 9–19. https://doi.org/10.1016/j.resmic.2014.11.002

Torii, S., Furumai, H., Katayama, H., 2021. Applicability of polyethylene glycol precipitation followed by acid guanidinium thiocyanate-phenol-chloroform extraction for the detection of SARS-CoV-2 RNA from municipal wastewater. Science of The Total Environment 756, 143067–143067. https://doi.org/10.1016/J.SCITOTENV.2020.143067

Touchon, M., Hoede, C., Tenaillon, O., Barbe, V., Baeriswyl, S., Bidet, P., Bingen, E., Bonacorsi, S., Bouchier, C., Bouvet, O., Calteau, A., Chiapello, H., Clermont, O., Cruveiller, S., Danchin, A., Diard, M., Dossat, C., Karoui, M.E., Frapy, E., Garry, L., Ghigo, J.M., Gilles, A.M., Johnson, J., Bouguénec, C.L., Lescat, M., Mangenot, S., Martinez-Jéhanne, V., Matic, I., Nassif, X., Oztas, S., Petit, M.A., Pichon, C., Rouy, Z., Ruf, C.S., Schneider, D., Tourret, J., Vacherie, B., Vallenet, D., Médigue, C., Rocha, E.P.C., Denamur, E., 2009. Organised Genome Dynamics in the Escherichia coli Species Results in Highly Diverse Adaptive Paths. PLOS Genetics 5, e1000344. https://doi.org/10.1371/journal.pgen.1000344

Tripathi, A., Marotz, C., Gonzalez, A., Vázquez-Baeza, Y., Song, S.J., Bouslimani, A., McDonald, D., Zhu, Q., Sanders, J.G., Smarr, L., Dorrestein, P.C., Knight, R., 2018. Are microbiome studies ready for hypothesis-driven research? Current Opinion in Microbiology 44, 61–69. https://doi.org/10.1016/J.MIB.2018.07.002

Truong, D.T., Franzosa, E.A., Tickle, T.L., Scholz, M., Weingart, G., Pasolli, E., Tett, A., Huttenhower, C., Segata, N., 2015. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. Nature Methods 12, 902–903. https://doi.org/10.1038/nmeth.3589

Tsementzi, D., Poretsky, R., Rodriguez-R, L.M., Luo, C., Konstantinidis, K.T., 2014. Evaluation of metatranscriptomic protocols and application to the study of freshwater microbial communities. Environ. Microbiol. Rep. 6, 640–655. https://doi.org/10.1111/1758-2229.12180

Tseng, C.-H., Chiang, P.-W., Shiah, F.-K., Chen, Y.-L., Liou, J.-R., Hsu, T.-C., Maheswararajah, S., Saeed, I., Halgamuge, S., Tang, S.-L., 2013. Microbial and viral metagenomes of a subtropical freshwater reservoir subject to climatic disturbances. ISME J 7, 2374–2386. https://doi.org/10.1038/ismej.2013.118

U.S. Centers for Disease Control and Prevention, 2022. Guidance for Reducing Health Risks to

Workers Handling Human Waste or Sewage [WWW Document]. URL https://www.cdc.gov/healthywater/global/sanitation/workers_handlingwaste.html (accessed 8.10.22).

U.S. Environmental Protection Agency, 2005. Membrane Filtration Guidance Manual.

U.S. Environmental Protection Agency, 2020. Contract Laboratory Program Guidance for Field Samplers 112.

U.S. EPA 2005. Method 1623: Cryptosporidium and Giardia in Water by Filtration/IMS/FA. 815-R-05-002.

US Environmental Protection Agency, 2012. Method 1611: Enterococci in Water by TaqMan® Quantitative Polymerase Chain Reaction (qPCR) Assay.

Uyaguari-Diaz, M.I., Chan, M., Chaban, B.L., Croxen, M.A., Finke, J.F., Hill, J.E., Peabody, M.A., Van Rossum, T., Suttle, C.A., Brinkman, F.S.L., Isaac-Renton, J., Prystajecky, N.A., Tang, P., 2016. A comprehensive method for amplicon-based and metagenomic characterization of viruses, bacteria, and eukaryotes in freshwater samples. Microbiome 4, 20. https://doi.org/10.1186/s40168-016-0166-1

Vadde, K.K., Feng, Q., Wang, J., McCarthy, A.J., Sekar, R., 2019. Next-generation sequencing reveals fecal contamination and potentially pathogenic bacteria in a major inflow river of Taihu Lake. Environ. Pollut. 254, 113108. https://doi.org/10.1016/j.envpol.2019.113108

van Dijk, E.L., Jaszczyszyn, Y., Thermes, C., 2014. Library preparation methods for next-generation sequencing: tone down the bias. Exp Cell Res 322, 12–20. https://doi.org/10.1016/j.yexcr.2014.01.008

Van Rossum, T., Peabody, M.A., Uyaguari-Diaz, M.I., Cronin, K.I., Chan, M., Slobodan, J.R., Nesbitt, M.J., Suttle, C.A., Hsiao, W.W.L., Tang, P.K.C., Prystajecky, N.A., Brinkman, F.S.L., 2015. Year-Long Metagenomic Study of River Microbiomes Across Land Use and Water Quality. Front Microbiol 6, 1405. https://doi.org/10.3389/fmicb.2015.01405

Vandeputte, D., Kathagen, G., D'hoe, K., Vieira-Silva, S., Valles-Colomer, M., Sabino, J., Wang, J., Tito, R.Y., De Commer, L., Darzi, Y., Vermeire, S., Falony, G., Raes, J., 2017. Quantitative microbiome profiling links gut community variation to microbial load. Nature 551, 507–511. https://doi.org/10.1038/nature24460

VanMensel, D., Chaganti, S.R., Droppo, I.G., Weisener, C.G., 2020. Exploring bacterial pathogen community dynamics in freshwater beach sediments: A tale of two lakes. Env. Microbiol 22, 568–583. https://doi.org/10.1111/1462-2920.14860

Vanysacker, L., Declerck, S.A.J., Hellemans, B., De Meester, L., Vankelecom, I., Declerck, P., 2010. Bacterial community analysis of activated sludge: an evaluation of four commonly used DNA extraction methods. Applied microbiology and biotechnology 88, 299–307.

https://doi.org/10.1007/S00253-010-2770-5

Varrone, C., Van Nostrand, J.D., Liu, W., Zhou, B., Wang, Z., Liu, F., He, Z., Wu, L., Zhou, J., Wang, A., 2014. Metagenomic-based analysis of biofilm communities for electrohydrogenesis: From wastewater to hydrogen. Int. J. Hydrog. Energy 39, 4222–4233. https://doi.org/10.1016/j.ijhydene.2014.01.001

Vasileski, G., 2000. Guideline on Sampling, Handling, Transporting, and Analyzing Legal Wastewater Samples. Canadian Water and Wastewater Association.

Veilleux, H.D., Misutka, M.D., Glover, C.N., 2021. Environmental DNA and environmental RNA: Current and prospective applications for biological monitoring. Science of The Total Environment 782, 146891. https://doi.org/10.1016/J.SCITOTENV.2021.146891

Venkateswaran, K., Vaishampayan, P., Cisneros, J., Pierson, D.L., Rogers, S.O., Perry, J., 2014. International Space Station environmental microbiome - microbial inventories of ISS filter debris. Appl Microbiol Biotechnol 98, 6453–6466. https://doi.org/10.1007/s00253-014-5650-6

Verheyen, J., Kaiser, R., Bozic, M., Timmen-Wego, M., Maier, B.K., Kessler, H.H., 2012. Extraction of viral nucleic acids: comparison of five automated nucleic acid extraction platforms. Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology 54, 255–259. https://doi.org/10.1016/J.JCV.2012.03.008

Vesper, S., McKinstry, C., Hartmann, C., Neace, M., Yoder, S., Vesper, A., 2008. Quantifying fungal viability in air and water samples using quantitative PCR after treatment with propidium monoazide (PMA). J Microbiol Methods 72, 180–184. https://doi.org/10.1016/j.mimet.2007.11.017

Vikram, A., Lipus, D., Bibby, K., 2016. Metatranscriptome analysis of active microbial communities in produced water samples from the Marcellus Shale. Microb. Ecol. 72, 571–581. https://doi.org/10.1007/s00248-016-0811-z

Vital, M., Dignum, M., Magic-Knezev, A., Ross, P., Rietveld, L., Hammes, F., 2012. Flow cytometry and adenosine tri-phosphate analysis: Alternative possibilities to evaluate major bacteriological changes in drinking water treatment and distribution systems. Water Res. 46, 4665–4676. https://doi.org/10.1016/j.watres.2012.06.010

Vosloo, S., Sevillano, M., Pinto, A., 2019. Modified DNeasy PowerWater Kit® protocol for DNA extractions from drinking water samples. https://doi.org/10.17504/protocols.io.66khhcw

Waak, M.B., Hozalski, R.M., Hallé, C., Lapara, T.M., 2019. Comparison of the microbiomes of two drinking water distribution systems - With and without residual chloramine disinfection. Microbiome 7, 1–14. https://doi.org/10.1186/s40168-019-0707-5

Walden, C., Carbonero, F., Zhang, W., 2017. Assessing impacts of DNA extraction methods on next generation sequencing of water and wastewater samples. J. Microbiol. Methods 141, 10–

16. https://doi.org/10.1016/j.mimet.2017.07.007

Wallace, J.C., Port, J.A., Smith, M.N., Faustman, E.M., 2017. FARME DB: a functional antibiotic resistance element database. Database 2017, baw165. https://doi.org/10.1093/database/baw165

Wan, J., Jing, Y., Rao, Y., Zhang, S., Luo, G., 2018. Thermophilic alkaline fermentation followed by mesophilic anaerobic digestion for efficient hydrogen and methane production from waste-activated sludge: Dynamics of bacterial pathogens as revealed by the combination of metagenomic and quantitative PCR ana. Appl Env. Microbiol 84, 1–14. https://doi.org/10.1128/AEM.02632-17

Wang, H., Masters, S., Hong, Y., Stallings, J., Falkinham, J.O., Edwards, M.A., Pruden, A., 2012. Effect of disinfectant, water age, and pipe material on occurrence and persistence of legionella, mycobacteria, pseudomonas aeruginosa, and two amoebas. Environmental Science and Technology 46, 11566–11574. https://doi.org/10.1021/es303212a

Wang, H., Proctor, C.R., Edwards, M.A., Pryor, M., Santo Domingo, J.W., Ryu, H., Camper, A.K., Olson, A., Pruden, A., 2014. Microbial Community Response to Chlorine Conversion in a Chloraminated Drinking Water Distribution System. Env. Sci Technol 48, 10624–10633. https://doi.org/10.1021/es502646d

Wang, M., Xiong, W., Liu, P., Xie, X., Zeng, J., Sun, Y., Zeng, Z., 2018. Metagenomic Insights Into the Contribution of Phages to Antibiotic Resistance in Water Samples Related to Swine Feedlot Wastewater Treatment. Front Microbiol 9, 2474. https://doi.org/10.3389/fmicb.2018.02474

Wang, Z., Zhang, X.-X., Lu, X., Liu, B., Li, Y., Long, C., Li, A., 2014. Abundance and diversity of bacterial nitrifiers and denitrifiers and their functional genes in tannery wastewater treatment plants revealed by high-throughput sequencing. PLoS One 9, e113603. https://doi.org/10.1371/journal.pone.0113603

Wanger, A., Chavez, V., Huang, R., Wahed, A., Dasgupta, A., Actor, J.K., 2017. Microbiology and Molecular Diagnosis in Pathology: A Comprehensive Review for Board Preparation, Certification and Clinical Practice. Elsevier.

Warrenfeltz, S., Basenko, E.Y., Crouch, K., Harb, O.S., Kissinger, J.C., Roos, D.S., Shanmugasundram, A., Silva-Franco, F., 2018. EuPathDB: The Eukaryotic Pathogen Genomics Database Resource. Methods Mol Biol 1757, 69–113. https://doi.org/10.1007/978-1-4939-7737-6_5

Watkins, S.C., Kuehnle, N., Anthony Ruggeri, C., Malki, K., Bruder, K., Elayyan, J., Damisch, K., Vahora, N., O'Malley, P., Ruggles-Sage, B., Romer, Z., Putonti, C., 2016. Assessment of a metaviromic dataset generated from nearshore Lake Michigan. Mar. Freshw. Res. 67, 1700. https://doi.org/10.1071/mf15172

Wattam, A.R., Abraham, D., Dalay, O., Disz, T.L., Driscoll, T., Gabbard, J.L., Gillespie, J.J., Gough, R., Hix, D., Kenyon, R., Machi, D., Mao, C., Nordberg, E.K., Olson, R., Overbeek, R., Pusch, G.D., Shukla, M., Schulman, J., Stevens, R.L., Sullivan, D.E., Vonstein, V., Warren, A., Will, R., Wilson, M.J.C., Yoo, H.S., Zhang, C., Zhang, Y., Sobral, B.W., 2014. PATRIC, the bacterial bioinformatics database and analysis resource. Nucleic Acids Res 42, D581-591. https://doi.org/10.1093/nar/gkt1099

Wattam, A.R., Davis, J.J., Assaf, R., Boisvert, S., Brettin, T., Bun, C., Conrad, N., Dietrich, E.M., Disz, T., Gabbard, J.L., Gerdes, S., Henry, C.S., Kenyon, R.W., Machi, D., Mao, C., Nordberg, E.K., Olsen, G.J., Murphy-Olson, D.E., Olson, R., Overbeek, R., Parrello, B., Pusch, G.D., Shukla, M., Vonstein, V., Warren, A., Xia, F., Yoo, H., Stevens, R.L., 2017. Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center. Nucleic Acids Res 45, D535–D542. https://doi.org/10.1093/nar/gkw1017

Weinroth, M.D., Thomas, K.M., Doster, E., Vikram, A., Schmidt, J.W., Arthur, T.M., Wheeler, T.L., Parker, J.K., Hanes, A.S., Alekoza, N., Wolfe, C., Metcalf, J.L., Morley, P.S., Belk, K.E., 2022. Resistomes and microbiome of meat trimmings and colon content from culled cows raised in conventional and organic production systems. Animal Microbiome 4, 21. https://doi.org/10.1186/s42523-022-00166-z

Weiss, S., Xu, Z.Z., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., Lozupone, C., Zaneveld, J.R., Vázquez-Baeza, Y., Birmingham, A., Hyde, E.R., Knight, R., 2017. Normalization and microbial differential abundance strategies depend upon data characteristics. Microbiome 5, 27. https://doi.org/10.1186/s40168-017-0237-y

Wenger, A.M., Peluso, P., Rowell, W.J., Change, P., Hall, R.J., Concepcion, G.T., Ebler, J., Fungtammasan, A., Kolesnikov, A., Olson, N.D., Topher, A., Alonge, M., Mahmoud, M., Qian, Y., Chin, C.S., Phillippy, A.M., Schatz, M.C., Myers, G., DePristo, M.A., Ruan, J., Marschall, T., Sedlazeck, F.J., Zook, J.M., Li, H., Koren, S., Carroll, A., Rank, D.R., Hunkapiller, M.W., 2019. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. Nat Biotechnol 37, 1155–1162. https://doi.org/10.1038/s41587-019-0217-9

Wexler, M., Bond, P.L., Richardson, D.J., Johnston, A.W.B., 2005. A wide host-range metagenomic library from a waste water treatment plant yields a novel alcohol/aldehyde dehydrogenase. Environ. Microbiol. 7, 1917–1926. https://doi.org/10.1111/j.1462-2920.2005.00854.x

Whiley, H., Taylor, M., 2016. Legionella detection by culture and qPCR: Comparing apples and oranges. Critical Reviews in Microbiology. https://doi.org/10.3109/1040841X.2014.885930

Whitney, O.N., Kennedy, L.C., Fan, V., Hinkle, A., Kantor, R., Greenwald, H., Crits-Christoph, A., Al-Shayeb, B., Chaplin, M., Maurer, A.C., Tjian, R., Nelson, K.L., 2021. Sewage, Salt, Silica and SARS-CoV-2 (4S): An economical kit-free method for direct capture of SARS-CoV-2 RNA from wastewater. Environ. Sci. Technol. 2021, 55, 8, 4880–4888. https://doi.org/10.1021/acs.est.0c08129

Wickham H (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org .
Wilke, A., Glass, E., Bischof, J., Braithwaite, D., Souza, M., Gerlach, W., 2013. MG-RAST technical report and manual for version 3.3. 6--Rev 1. Lemont IL Argonne Natl. Lab.

Williams, F., Stetler, R., Safferman, R., 2001. USEPA manual of methods for virology. Environmental Protection Agency 600, 4–84.

Williams, K.E., Huyvaert, K.P., Piaggio, A.J., 2016. No filters, no fridges: a method for preservation of water samples for eDNA analysis. BMC Res Notes 9, 298. https://doi.org/10.1186/s13104-016-2104-5

Williams, S.C.P., 2013. The other microbiome. Proc. Natl. Acad. Sci. 110, 2682–2684. https://doi.org/10.1073/pnas.1300923110

Wilton, S. and Cousins, D. 1992. Detection and identification of multiple mycobacterial pathogens by DNA amplification in a single tube. Genome Research 1(4), 269-273.

Wong, M.T., Zhang, D., Li, J., Hui, R.K.H., Tun, H.M., Brar, M.S., Park, T.-J., Chen, Y., Leung, F.C., 2013. Towards a metagenomic understanding on enhanced biomethane production from waste activated sludge after pH 10 pretreatment. Biotechnol Biofuels 6, 38. https://doi.org/10.1186/1754-6834-6-38

Wong, T., Deveson, I.W., Hardwick, S.A., Mercer, T.R., 2017. ANAQUIN: a software toolkit for the analysis of spike-in controls for next generation sequencing. Bioinformatics 33, 1723–1724. https://doi.org/10.1093/bioinformatics/btx038

Wood, D.E., Lu, J., Langmead, B., 2019. Improved metagenomic analysis with Kraken 2. Genome Biology 20. https://doi.org/10.1186/s13059-019-1891-0

World Health Organization (WHO), 2015. Global action plan on antimicrobial resistance 1–28.

Wüthrich, D., Gautsch, S., Spieler-Denz, R., Dubuis, O., Gaia, V., Moran-Gilad, J., Hinic, V., Seth-Smith, H.M.B., Nickel, C.H., Tschudin-Sutter, S., Bassetti, S., Haenggi, M., Brodmann, P., Fuchs, S., Egli, A., 2019. Air-conditioner cooling towers as complex reservoirs and continuous source of Legionella pneumophila infection evidenced by a genomic analysis study in 2017, Switzerland. Eurosurveillance 24, 1–7. https://doi.org/10.2807/1560-7917.ES.2019.24.4.1800192

Xia, Y., Yang, C., Zhang, T., 2018. Microbial effects of part-stream low-frequency ultrasonic pretreatment on sludge anaerobic digestion as revealed by high-throughput sequencing-based metagenomics and metatranscriptomics. Biotechnol. Biofuels 11. https://doi.org/10.1186/s13068-018-1042-y

Xu, H., Pei, H., Jin, Y., Ma, C., Wang, Y., Sun, J., Li, H., 2018. High-throughput sequencing reveals microbial communities in drinking water treatment sludge from six geographically distributed plants, including potentially toxic cyanobacteria and pathogens. Sci Total Env. 634, 769–779.

https://doi.org/10.1016/j.scitotenv.2018.04.008

Yadav, S., Kapley, A., 2019. Exploration of activated sludge resistome using metagenomics. Sci. Total Environ. 692, 1155–1164. https://doi.org/10.1016/j.scitotenv.2019.07.267

Yadav, T.C., Pal, R.R., Shastri, S., Jadeja, N.B., Kapley, A., 2015. Comparative metagenomics demonstrating different degradative capacity of activated biomass treating hydrocarbon contaminated wastewater. Bioresour Technol 188, 24–32. https://doi.org/10.1016/j.biortech.2015.01.141

Yáñez, M.A., Nocker, A., Soria-Soria, E., Múrtula, R., Martínez, L., Catalán, V., 2011. Quantification of viable Legionella pneumophila cells using propidium monoazide combined with quantitative PCR. J Microbiol Methods 85, 124–130. https://doi.org/10.1016/j.mimet.2011.02.004

Yang, Y., Hou, Y., Ma, M., Zhan, A., 2020a. Potential pathogen communities in highly polluted river ecosystems: Geographical distribution and environmental influence. Ambio 49, 197–207. https://doi.org/10.1007/s13280-019-01184-z

Yang, Y., Li, B., Ju, F., Zhang, T., 2013. Exploring variation of antibiotic resistance genes in activated sludge over a four-year period through a metagenomic approach. Env. Sci Technol 47, 10197–10205. https://doi.org/10.1021/es4017365

Yang, Y., Li, B., Zou, S., Fang, H.H.P., Zhang, T., 2014a. Fate of antibiotic resistance genes in sewage treatment plant revealed by metagenomic approach. Water Res. 62, 97–106. https://doi.org/10.1016/j.watres.2014.05.019

Yang, Y., Pan, J., Zhou, Z., Wu, J., Liu, Y., Lin, J.-G., Hong, Y., Li, X., Li, M., Gu, J.-D., 2020b. Complex microbial nitrogen-cycling networks in three distinct anammox-inoculated wastewater treatment systems. Water Research 168, 115142. https://doi.org/10.1016/j.watres.2019.115142

Yang, Y., Yu, K., Xia, Y., Lau, F.T.K., Tang, D.T.W., Fung, W.C., Fang, H.H.P., Zhang, T., 2014b. Metagenomic analysis of sludge from full-scale anaerobic digesters operated in municipal wastewater treatment plants. Appl. Microbiol. Biotechnol. 98, 5709–5718. https://doi.org/10.1007/s00253-014-5648-0

Yard, E.E., Murphy, M.W., Schneeberger, C., Narayanan, J., Hoo, E., Freiman, A., Lewis, L.S. and Hill, V.R. 2014. Microbial and chemical contamination during and after flooding in the Ohio River—Kentucky, 2011. Journal of Environmental Science and Health, Part A 49(11), 1236-1243.

Ye, L., Zhang, T., Wang, T., Fang, Z., 2012. Microbial Structures, Functions, and Metabolic Pathways in Wastewater Treatment Bioreactors Revealed Using High-Throughput Sequencing. Environ. Sci. Technol. 46, 13244–13252. https://doi.org/10.1021/es303454k

Yergeau, E., Masson, L., Elias, M., Xiang, S., Madey, E., Huang, H., Brooks, B., Beaudette, L.A.,

2016. Comparison of Methods to Identify Pathogens and Associated Virulence Functional Genes in Biosolids from Two Different Wastewater Treatment Facilities in Canada. PLoS One 11, 1–20. https://doi.org/10.1371/journal.pone.0153554

Yilmaz, P., Kottmann, R., Field, D., Knight, R., Cole, J.R., Amaral-Zettler, L., Gilbert, J.A., Karsch-Mizrachi, I., Johnston, A., Cochrane, G., Vaughan, R., Hunter, C., Park, J., Morrison, N., Rocca-Serra, P., Sterk, P., Arumugam, M., Bailey, M., Baumgartner, L., Birren, B.W., Blaser, M.J., Bonazzi, V., Booth, T., Bork, P., Bushman, F.D., Luigi Buttigieg, P., G Chain, P.S., Charlson, E., Costello, E.K., Huot-Creasy, H., Dawyndt, P., DeSantis, T., Fierer, N., Fuhrman, J.A., Gallery, R.E., Gevers, D., Gibbs, R.A., San Gil, I., Gonzalez, A., Gordon, J.I., Guralnick, R., Hankeln, W., Highlander, S., Hugenholtz, P., Jansson, J., Kau, A.L., Kelley, S.T., Kennedy, J., Knights, D., Koren, O., Kuczynski, J., Kyrpides, N., Larsen, R., Lauber, C.L., Legg, T., Ley, R.E., Lozupone, C.A., Ludwig, W., Lyons, D., Maguire, E., Meth, B.A., Meyer, F., Muegge, B., Nakielny, S., Nelson, K.E., Nemergut, D., Neufeld, J.D., Newbold, L.K., Oliver, A.E., Pace, N.R., Palanisamy, G., Peplies, rg, Petrosino, J., Proctor, L., Pruesse, E., Quast, C., Raes, J., Ratnasingham, S., Ravel, J., Relman, D.A., Assunta-Sansone, S., Schloss, P.D., Schriml, L., Sinha, R., Smith, M.I., Sodergren, E., Stombaugh, J., Tiedje, J.M., Ward, D. V, Weinstock, G.M., Wendel, D., White, O., Whiteley, A., Wilke, A., Wortman, J.R., Yatsunenko, T., Oliver Gl, F., 2011. Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. Nature Biotechnology. https://doi.org/10.1038/nbt.1823

Yoshitomi, H., Sera, N., Gonzalez, G., Hanaoka, N., Fujimoto, T., 2017. First isolation of a new type of human adenovirus (genotype 79), species Human mastadenovirus B (B2) from sewage water in Japan. J Med Virol 89, 1192–1200. https://doi.org/10.1002/jmv.24749

Yu, H., Zhao, Q., Dong, Q., Jiang, J., Wang, K., Zhang, Y., 2018. Electronic and metagenomic insights into the performance of bioelectrochemical reactor simultaneously treating sewage sludge and Cr(VI)-laden wastewater. Chem. Eng. J. 341, 495–504. https://doi.org/10.1016/j.cej.2018.01.159

Yu, K., Zhang, T., 2012. Metagenomic and metatranscriptomic analysis of microbial community structure and gene expression of activated sludge. PLoS One 7, e38183. https://doi.org/10.1371/journal.pone.0038183

Yuan, Q.B., Huang, Y.M., Wu, W.B., Zuo, P., Hu, N., Zhou, Y.Z., Alvarez, P.J.J., 2019. Redistribution of intracellular and extracellular free & adsorbed antibiotic resistance genes through a wastewater treatment plant by an enhanced extracellular DNA extraction method with magnetic beads. Environment international 131. https://doi.org/10.1016/J.ENVINT.2019.104986

Yuan, S., Cohen, D.B., Ravel, J., Abdo, Z., Forney, L.J., 2012. Evaluation of Methods for the Extraction and Purification of DNA from the Human Microbiome. PLOS ONE 7, e33865. https://doi.org/10.1371/journal.pone.0033865

Zacheus, O.M., Lehtola, M.J., Korhonen, L.K., Martikainen, P.J., 2001. Soft deposits, the key site for microbial growth in drinking water distribution networks. Water Res 35, 1757–1765. https://doi.org/10.1016/s0043-1354(00)00431-0

Zamyadi, A., Romanis, C., Mills, T., Neilan, B., Choo, F., Coral, L.A., Gale, D., Newcombe, G., Crosbie, N., Stuetz, R., Henderson, R.K., 2019. Diagnosing water treatment critical control points for cyanobacterial removal: Exploring benefits of combined microscopy, next-generation sequencing, and cell integrity methods. Water Res 152, 96–105. https://doi.org/10.1016/j.watres.2019.01.002

Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., Aarestrup, F.M., Larsen, M.V., 2012. Identification of acquired antimicrobial resistance genes. Journal of Antimicrobial Chemotherapy 67, 2640–2644. https://doi.org/10.1093/jac/dks261

Zerbino, D.R., Birney, E., 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18, 821–829. https://doi.org/10.1101/gr.074492.107

Zhang, S., He, Z., Meng, F., 2019. Floc-size effects of the pathogenic bacteria in a membrane bioreactor plant. Env. Int 127, 645–652. https://doi.org/10.1016/j.envint.2019.04.002

Zhang, S., Wei, B., Yu, X., Liu, B., Wu, Z., Gu, L., 2010. Development of pretreatment protocol for DNA extraction from biofilm attached to biologic activated carbon (BAC) granules. Frontiers of Environmental Science & Engineering 4, 459–465. https://doi.org/10.1007/S11783-010-0249-3

Zhang, T., Miao, J., Han, N., Qiang, Y., Zhang, W., 2018. MPD: a pathogen genome and metagenome database. Database 2018. https://doi.org/10.1093/database/bay055

Zhang, T., Yang, Y., Pruden, A., 2015. Effect of temperature on removal of antibiotic resistance genes by anaerobic digestion of activated sludge revealed by metagenomic approach. Appl. Microbiol. Biotechnol. 99, 7771–7779. https://doi.org/10.1007/s00253-015-6688-9

Zhang, Y., Liu, W.T., 2019. The application of molecular tools to study the drinking water microbiome – Current understanding and future needs. https://doi.org/10.1080/10643389.2019.1571351

Zhang, Y., Snow, D.D., Parker, D., Zhou, Z., Li, X., 2013. Intracellular and extracellular antimicrobial resistance genes in the sludge of livestock waste management structures. Environ Sci Technol 47, 10206–10213. https://doi.org/10.1021/es401964s

Zhao, Z., Cui, H., Song, W., Ru, X., Zhou, W., Yu, X., 2020. A simple magnetic nanoparticles-based viral RNA extraction method for efficient detection of SARS-CoV-2. bioRxiv 2020.02.22.961268-2020.02.22.961268. https://doi.org/10.1101/2020.02.22.961268

Zheng, X., Deng, Y., Xu, X., Li, S., Zhang, Y., Ding, J., On, H.Y., Lai, J.C.C., In Yau, C., Chin, A.W.H., Poon, L.L.M., Tun, H.M., Zhang, T., 2022. Comparison of virus concentration methods and RNA extraction methods for SARS-CoV-2 wastewater surveillance. Sci Total Environ 824, 153687–153687. https://doi.org/10.1016/j.scitotenv.2022.153687

Zhu, N., Ghosh, S., Strom, L., Pruden, A., Edwards, M., 2020a. Effects of BAC-filtration, disinfection, and temperature on water quality in simulated reclaimed water distribution systems. Environmental Science: Water Research & Technology. https://doi.org/10.1039/d0ew00581a

Zhu, N., Mapili, K., Majeed, H., Pruden, A., Edwards, M., 2020b. Sediment and Biofilm Affect Disinfectant Decay Rates During Long-term Operation of Simulated Reclaimed Water Distribution Systems. Environmental Science: Water Research & Technology. https://doi.org/10.1039/c9ew00978g

THE
Water
Research
FOUNDATION

advancing the science of water®

1199 North Fairfax Street, Suite 900
Alexandria, VA 22314-1445

6666 West Quincy Avenue
Denver, CO 80235-3098

www.waterrf.org | info@waterrf.org