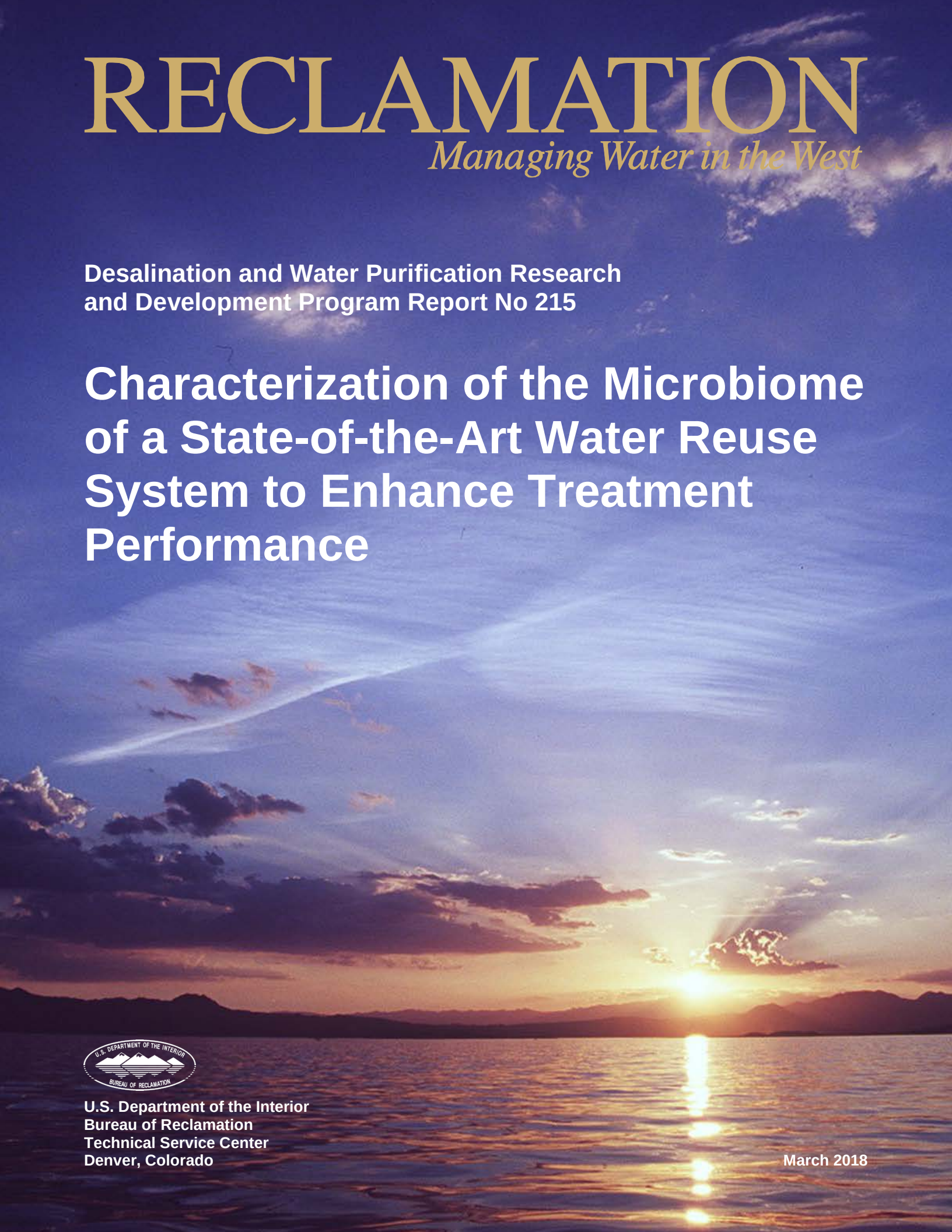
RECLAMATION

Managing Water in the West

Desalination and Water Purification Research and Development Program Report No 215

Characterization of the Microbiome of a State-of-the-Art Water Reuse System to Enhance Treatment Performance



U.S. Department of the Interior
Bureau of Reclamation
Technical Service Center
Denver, Colorado

March 2018

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
14-03-2018	Final	01-10-2016 – 31-10-2017

4. TITLE AND SUBTITLE
Characterization of the Microbiome of a State-of-the-Art Water Reuse System to Enhance Treatment Performance

5a. CONTRACT NUMBER
Agreement No. R16AC00127

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Blake W. Stamps, Postdoctoral Fellow, Colorado School of Mines
John R. Spear, Professor, Colorado School of Mines
Menu Leddy, Principal Scientist , Orange County Water District
Megan H. Plumlee, Director of Research, Orange County Water District
Rita Colwell, Professor, University of Maryland at College Park and CosmosID

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Colorado School of Mines
1500 Illinois Street
Golden, CO 80401

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Bureau of Reclamation
U.S. Department of the Interior
Denver Federal Center
PO Box 25007, Denver, CO 80225-0007

10. SPONSOR/MONITOR'S ACRONYM(S)
Reclamation

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
DWPR Report No. 215

12. DISTRIBUTION/AVAILABILITY STATEMENT
Available at https://www.usbr.gov/research/dwpr/DWPR_Reports.html

13. SUPPLEMENTARY NOTES
Online at https://www.usbr.gov/research/dwpr/DWPR_Reports.html

14. ABSTRACT

Potable water is a precious resource. Any technology that extends or allows the reuse of potable water supplies across the United States is in critical need. The Orange County Groundwater Replenishment System (GWRS) is an advanced water purification facility (AWPF) that uses a multi-barrier treatment process to allow for the indirect reuse of advanced-treated wastewater. This project sought to develop a better understanding of pathogen removal by studying the microbial community composition and assess the performance of various unit processes throughout the engineered system. The study aimed to advance and inform our understanding of microbial ecologies throughout the treatment works and to use this knowledge to optimize multi-barrier water treatment systems. The completed study has shown that the AWPF effectively removes the majority of the detected biomass through microfiltration and reverse osmosis. Following an advanced oxidation process (AOP) treatment, there was very little detectable biomass, despite the similarity of microbial communities identified in 16S rRNA gene sequencing. Our work illustrates the need to not only obtain microbial community composition but also to quantify the microbial load in treatment systems—advanced or otherwise. Metagenomic sequencing proved helpful in providing strain-level resolution of microorganisms in the AWPF. This approach accurately identifies potential pathogens, and genetic elements of interest, such as antibiotic resistance, that may persist through treatment. Coupling the CosmosID approach to identification with DNA and RNA will accurately identify microorganisms that are present, functional, and likely metabolically active.

15. SUBJECT TERMS
Advanced water treatment, microbiome, high-throughput sequencing, built-environment, water reuse

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	THIS PAGE			Yuliana Porras-Mendoza
U	U	U			19b. TELEPHONE NUMBER *(Include area code)* 303-445-2265

Desalination and Water Purification Research and Development Program Report No. 215

Characterization of the Microbiome of a State-of-the-Art Water Reuse System to Enhance Treatment Performance

Prepared for the Bureau of Reclamation Under Agreement No. R16AC00127

by

Blake W. Stamps, Colorado School of Mines

John R. Spear, Colorado School of Mines

Menu Leddy, Orange County Water District

Megan H. Plumlee, Orange County Water District

Rita R. Colwell, University of Maryland at College Park and CosmosID, Inc.

U.S. Department of the Interior
Bureau of Reclamation
Technical Service Center
Denver, Colorado

March 2018

Mission Statements

The U.S. Department of the Interior protects America's natural resources and heritage, honors our cultures and tribal communities, and supplies the energy to power our future.

The mission of the Bureau of Reclamation is to manage, develop, and protect water and related resources in an environmentally and economically sound manner in the interest of the American public.

Disclaimer

The views, analysis, recommendations, and conclusions in this report are those of the authors and do not represent official or unofficial policies or opinions of the United States Government, and the United States takes no position with regard to any findings, conclusions, or recommendations made. As such, mention of trade names or commercial products does not constitute their endorsement by the United States Government.

Acknowledgments

We thank the Desalination and Water Purification Research and Development Program and the Bureau of Reclamation, for sponsoring of this research.

Primary research was carried out by The Colorado School of Mines (Blake Stamps and John Spear) and Orange County Water District (Menu Leddy and Megan Plumlee). Sampling and site access were provided by OCWD. Total DNA and RNA sequencing, as well as analysis/identification of HTS sequence data was provided by Rita Colwell and CosmosID, Incorporated.

We thank Nur A. Hasan, CosmosID, for technical support throughout the study as well as assistance with figure generation comments during the revision process.

Acronyms and Abbreviations

ARG	antibiotic resistance gene
AOP	advanced oxidation process
AWPF	advanced water purification facility
BDL	below detectable limit
bp	basepair
CCP	critical control points
cDNA	complementary DNA
DDW	California Division of Drinking Water
DNA	deoxyribonucleic acid
EPS	extracellular polymeric substance
FPW	final product water
GWRS	Groundwater Replenishment System
HPOAS	high purity oxygen activated sludge
HTS	high throughput sequencing
ICP-OES	inductively coupled plasma optical emission spectrometer
LVC	large volume concentrator
MAG	metagenome assembled genome
MF	microfiltration
MFE	MF effluent water
MFF	MF feed water
NDMA	N-nitrosodimethylamine
NdN	nitrification-denitrification mode
OCSD	Orange County Sanitation District
OCWD	Orange County Water District
OTU	operational taxonomic unit
PBS	phosphate buffered saline
PCA	principal component analysis
PCR	polymerase chain reaction
PMMoV	pepper mild mottle virus
NdN	nitrification/de-nitrification
NTU	nephelometric turbidity unit
qPCR	quantitative polymerase chain reaction
RO	reverse osmosis
ROF	RO system feed
ROP	RO system permeate
Q1	secondary treated wastewater effluent from OCSD
rRNA	ribosomal RNA
RNA	ribonucleic acid
SEM	scanning electron microscope
sOTU	sub-operational taxonomic unit
SSU rRNA gene	small subunit ribosomal RNA gene
TF	trickling filter
UV	ultraviolet light
UV/AOP	UV advanced oxidation process
UVF	UV Feed/AOP system feed water
UVP	UV Product Water/AOP system product water

Measurements

°C	degree Celsius
L	liter
mgd	million gallons per day
mg/L	milligram per liter
mJ/cm2	millijoule per square centimeter
mL	milliliter
mm	millimeter
mmho/cm	millimho per centimeter
mS/cm	microSiemens per centimeter
nM	nanomolar
ng	nanogram
NTU	nephelometric turbidity unit
ppm	part per million
16S rRNA genes/mL	number of copies of 16S rRNA gene per mL
µg/L	microgram per liter
µL	microliter
µm	micron
µM	micromolar
v/v	volume-volume percentage

Contents

Figures

Tables

Executive Summary

Potable water is a precious resource. In many communities across the United States and the world, there is a critical need for additional, reliable water sources. Advanced treatment technologies coupled with water quality monitoring tools make possible the *reuse* of potable water supplies. To this end, the Orange County Water District (OCWD) Groundwater Replenishment System (GWRS) in Orange County, California is an advanced water purification facility (AWPF) that uses a multi-barrier treatment process to allow for the reuse of advanced-treated wastewater. This advanced treated water is then stored underground to augment groundwater as a potable water supply for the region. Planned water reuse is now being implemented more than in any time in our history to help meet the water needs of growing societies. Due to a number of scientific and technical advances, using highly treated, high quality reclaimed water is safe and acceptable. While non-planned reuse has been practiced since the beginning of time, new technologies, engineering, and knowledge have made it possible to produce highly purified water from wastewater for potable purposes.

The present study had defined goals which were successfully achieved. This project sought to apply high throughput sequencing (HTS) methods for the first time to a water reuse treatment facility, and to use HTS and other microbiology-based methods to characterize the microbial community composition, develop a better understanding of pathogen removal, and assess differences in treatment performance for the unit processes in the engineered treatment system. An overall goal was to advance, inform and optimize multi-barrier water treatment systems with regard to microbial ecology throughout the treatment works. Most pressing was identifying and quantifying the removal of biomass and identifying relevant pathogens present throughout the treatment system. Additionally, the study identified a viable filtration and concentration system for rapid sampling and identification of microorganisms from an ultra-low biomass environment, specifically, waters treated through reverse osmosis (RO) and an ultraviolet (UV) hydrogen peroxide (H_2O_2) advanced oxidation process (AOP) that contain little to no viable microorganisms. Method development was a critical part of this demonstration study on the use of existing and newly available microbial community analysis methods for potable reuse.

To characterize microbial community composition at various points in the AWPF system, the project team used a combination of methods, including HTS of total deoxyribonucleic acid (DNA) and total ribonucleic acid (RNA), also known as metagenomic or metatranscriptomic HTS, as well as small subunit ribosomal RNA gene (SSU rRNA gene) analysis to identify microorganisms and address water treatment effectiveness. Due to recent advances in DNA sequencing such as HTS, this was carried out with greater coverage (amount sequenced), and at a lower cost than previously possible. In addition, RNA HTS was carried out for the first time in an advanced water treatment system as a part of this study. The project team also conducted quantitative polymerase chain reaction (qPCR) to identify microbial community biomass at specific points across the AWPF

treatment train. The combination of HTS and qPCR methods were used to identify microbial communities and relative changes in the community structure at each step of the treatment process. With respect to method development, the sampling methods identified for water reuse systems can be used at OCWD and at similar facilities. Specifically, the project team identified a viable method using dialysis-style filters that allow for rapid sampling to determine the concentration of biomass contained in tens to hundreds of liters per sampling event. The combination of methods used in the presented study evaluate the microbial water quality of purified recycled water, further advancing the practice of potable reuse and providing fascinating insights not possible with previous methods.

The AWPF is the main component of the GWRS and the focus of this study. The AWPF treatment train is comprised of microfiltration (MF), RO, and UV disinfection with hydrogen peroxide (H_2O_2) addition. The UV/ H_2O_2 treatment step is referred to as the advanced oxidation process or AOP. This is followed by partial decarbonation and lime stabilization of the advanced treated water. The AWPF was sampled throughout the water treatment train at two distinct sampling events. OCWD meets all permit requirements for water quality and monitors the operational performance of the AWPF unit process using critical control points (CCP). CCP sampling locations cover key elements of the AWPF process and have identified the most important process monitoring and control points. This study was designed to accurately address the question of filtration effectiveness for microbial water quality and to determine if any potentially harmful microorganisms remain post RO treatment. A combined DNA and RNA sequencing approach was used to determine what microorganisms are present and functional throughout the treatment process.

Several key operational parameters were obtained from online readings every six hours throughout the AWPF alongside the presented microbial community analysis including: temperature, pH, turbidity, and total chlorine from September 1, 2016 to July 31, 2017 (over the study period and including the two microbial community sampling events). The average turbidity declined from 3.6 nephelometric turbidity units (NTU) at the secondary treated wastewater effluent (Q1 sampling site) which is the influent to the AWPF, to 1.6 NTU at MF feed (MFF) sample location. Average turbidity ranged between 0.08 and 0.04 NTU at the RO system feed (ROF), RO permeate (ROP) water, and final product water (FPW), respectively. Water temperature was stable throughout the system over the sampling period, averaging near 80 degrees Fahrenheit (°F) (≈ 26.7 degrees Celsius [°C]), with an average range between 79.5 °F at ROF to 80.8 °F at FPW. The pH was stable at 7.0 at MFF and ROF, declining to 5.51 - 5.64 at ROP and UVP respectively, as expected as a part of the acid addition and RO treatment process. An expected increase in pH was observed at FPW (average pH at 8.52) due to the addition of calcium hydroxide (hydrated lime) prior to the FPW exiting the AWPF.

Much like the water quality data obtained from the GWRS across the sampling period, SSU rRNA gene and metagenomics community analyses in 2016 and 2017 were highly concurrent, although some temporal variation was noted. Communities at Q1 and MFF were highly similar in taxonomic composition by SSU rRNA gene and total DNA sequence analyses despite the addition of sodium hypochlorite between these two sites. As expected, the greatest shift in microbial community distribution and abundance came at the microfiltration stage where particulate and biomass larger than 0.2 microns (µm) is removed—a pore size at or below the size of protozoa and most bacteria. Differences at Q1 and MFF between DNA and RNA samples by SSU rRNA gene analysis were mostly confined to the Clostridia and Betaproteobacteria. Betaproteobacteria are common in wastewater treatment systems and are responsible for the bulk of organic carbon removal during primary treatment (i.e., wastewater treatment process upstream of the AWPF); thus, their presence is expected. Water samples beyond MF effluent water (MFE) (i.e., RO permeate and finished water) produced insufficient quantities of DNA or RNA to obtain total DNA or RNA HTS data—indicative of the expected large-scale removal of biomass, particularly at RO where average RO membrane pore size is below that of even viral particles. As with the bacterial community analysis, the analysis detected a broader diversity and relative abundance of antibiotic resistance genes and viral sequences prior to microfiltration by using the CosmosID bioinformatic analysis and identification approach.

Data were further investigated with respect to the microbial community present throughout the AWPF using metagenomics and metatranscriptomics to identify any potential pathogens that might present a public health risk. Water samples retrieved from MFE had the lowest number of identifiable genes associated with virulence factors or known antibiotic resistance genes using HTS compared to Q1 MFF, MF Biofilm, or RO Biofilm. Microbial density, rather than differences in community distribution, is known to be a critical component of disease status in the human microbiome, the microbial associates of the human body. A similar correlation is likely in water purification and distribution systems. Another heavy reduction in the absolute abundance in the community estimated by qPCR was noted in RO feed (ROF) and product water (ROP), due to the physical filtration effect (pore size) of the RO system. Therefore, a combined multi-barrier filtration system (MF + RO) coupled to UV/AOP treatment is a clearly effective means of removing the microbial population present in secondary treated wastewater. The final step in the AWPF is the injection of treated water into the subsurface allowing for (indirect) potable reuse. The injection of high-purity finished water, and the associated microbial community, or lack thereof, into a subsurface aquifer is a potentially interesting area for future study related to such an effective treatment system. Cell numbers in the subsurface average near 10^3 to 10^4 cells per milliliter (mL), far lower than the concentration of cells in secondary wastewater effluent that is influent into the AWPF ($\approx 10^6$ cells per mL) yet much higher than the purified, finished water from the AWPF that is presumed to be below this concentration. Thus, the water injected into the subsurface may dilute what little

microbial community exists. What—if any—impact this might have on the subsurface microbial community should be a focus of future work, which should continue to consider the active microbial populations.

The completed study has shown that the AWPF effectively removes the majority of the detected biomass through MF and RO. Following UV/AOP treatment (the final water quality polishing step after RO), there is very little detectable biomass, despite the similarity of microbial communities identified in 16S rRNA gene sequencing at MFE, ROP, and after UV/AOP. This work illustrates the need to not only obtain microbial community composition but also to quantify the microbial load in treatment systems, advanced or otherwise. The project team suggests ongoing monitoring at the AWPF or other similar sites include qPCR estimation of the microbial load, in addition to community composition analyses for accurate assessment of microbial communities.

The presented work also shows the critical nature of large sample volumes, quantification, and importantly, laboratory controls in the analysis of ultra-low biomass environments, such as the AWPF. DNA was non-quantifiable, or present in very low abundance in samples after RO treatment. Furthermore, microorganisms (or their DNA) present in commonly used DNA extraction or PCR reagents in the laboratory setting (i.e., ultra-pure water) are likely to be very similar to those detected in the AWPF treatment train after RO. Future work should continue to sequence and quantify filtration and extraction controls to ensure detected microbial communities are actually those present in the water. Increasing filtration volumes to, or beyond, 100 liters (L) (as taken in the second sampling event of this study) will enhance the limit of detection and mitigate issues of low DNA concentration.

Total DNA and RNA HTS analyses proved helpful in providing strain-level identification of microorganisms in the AWPF. This approach accurately identifies potential pathogens and genetic elements of interest (such as antibiotic resistance) that may persist through treatment. Coupling the CosmosID bioinformatic analysis approach to identification techniques using both DNA and RNA will aid in accurately identifying microorganisms that are present, *functional*, and likely to be metabolically active.

1. Introduction

1.1. Problem

Due to cyclical weather cycles, drought (both short- and long-term), climate change and increasing human population, water stress has soared in the southern California region and projects that integrate water reuse (potable and non-potable) have become more desirable. This study's aim was to better understand the microbial diversity throughout an advanced treatment facility that produces purified, reclaimed water for potable reuse. This project sought to apply high throughput sequencing (HTS) methods for the first time to a water reuse treatment facility and to use HTS and other sequencing methods to: characterize the microbial community composition, develop a better understanding of pathogen removal, and assess differences in treatment performance for the unit processes in the engineered treatment system. To understand microbial community composition at various points in the system design, the project team used a combination of HTS, small subunit ribosomal RNA gene (SSU rRNA) sequencing, and quantitative polymerase chain reaction (qPCR) techniques to address overall removal efficiencies at each step of the treatment process and the relative change in the overall microbial community of the Orange County Water District (OCWD) Groundwater Replenishment System (GWRS) Advanced Water Purification Facility (AWPF).

1.2. Needs

Water is a critical component of the Earth's life support systems, yet many societies will face water scarcity due to acceleration in population growth, land use, and a corresponding acceleration in water use. With global population increase and climate change, many communities also face an increase in water quality degradation as demonstrated by increased eutrophication (Madigan et al., 2010). Both water scarcity and degradation of water quality can impact food production and economic prosperity, which can increase the pressure on available water resources. As a result, water reuse is attracting greater attention (EPA and CDM Smith, 2017). Reusing municipal wastewater instead of discharging to surface waters offers an opportunity to augment the water supply of the communities facing water shortages. Many of these communities serve large populations and can find themselves in immediate need when drought strikes as evidenced by cities in California over the past five years of drought.

Planned water reuse is now being implemented more than in any time in our history to help meet the needs of growing societies (Jimenez and Asano, 2008). Due to a number of scientific and technical advances, using highly treated, high

quality reclaimed water is accepted and considered an important part of a diverse water supply portfolio in water-scarce regions. While non-planned reuse has occurred since the beginning of time, new technologies, engineering, and knowledge have made it possible to produce highly purified water from wastewater for potable purposes.

1.3. Objectives

The project sought to:

1) Define the total microbial load of water at points along the AWPF treatment train;

2) Identify the resident microbial community in the AWPF using HTS coupled with metagenomics analysis as well as using 16S rRNA gene sequencing and qPCR;

3) Correlate the removal of microbial communities across the AWPF to existing water quality data demonstrating treatment;

4) Assess the removal efficiency of microorganisms at various points in treatment;

5) Develop sampling techniques for microbial analysis of high-purity water and potentially better indicators for systems similar to the GWRS AWPF.

1.4. Location

The project was carried out at the OCWD GWRS in central Orange County, California in Fountain Valley, California (Figure 1). The GWRS consists of three major components:

- AWPF, which consists of a multi-barrier treatment process
- Seawater intrusion barrier (injection wells)
- Percolation basins that facilitate groundwater recharge

The GWRS is a potable reuse project jointly operated by the OCWD and Orange County Sanitation District (OCSD). OCWD operates the AWPF, which is currently the world's largest water reclamation facility for potable reuse and is a recognized industry standard. OCSD supplies the secondary treated wastewater that would otherwise be discharged to the Pacific Ocean for disposal.

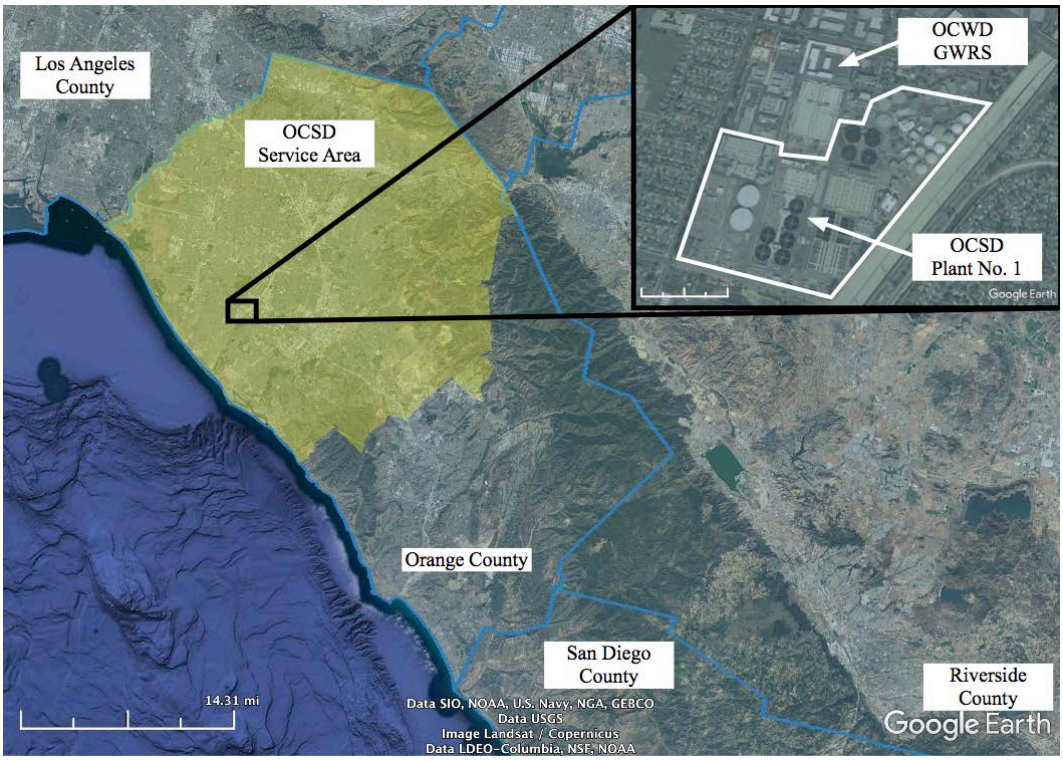


Figure 1.—Study area. The greater Los Angeles metropolis is shown, including county lines in blue, with the OCSD service area shaded in yellow. The Orange County Sanitation Reclamation Plant No. 1 and Orange County Water District GWRS (Including the AWPF) are inset. Images captured from Google Earth/Landsat. Approximate OCSD service area was obtained from the OCSD website.

The AWPF uses advanced water treatment to produce up to 100 million gallons per day (mgd) of potable-quality water from secondary treated wastewater. The AWPF treats approximately 130 mgd of OCSD's effluent to produce 100 mgd of product water. The treated wastewater that serves as the AWPF influent is comprised of an average blend of about 80% high purity oxygen activated sludge (HPOAS) operating in nitrification-denitrification mode (NdN) and 20% trickling filter (TF) sourced water, though the blend can be variable. The AWPF product water supplements existing water supplies by providing a reliable, high quality source of water to recharge the Orange County Groundwater Basin. OCWD stores the water in the groundwater basin, protecting that groundwater basin from degradation due to seawater intrusion from the Huntington Beach coast line. As a groundwater augmentation project, the GWRS is considered an "indirect" potable reuse project, as opposed to direct reuse in which water is supplied to a drinking water treatment plant ("raw water augmentation") or into the drinking water distribution system ("treated water augmentation") without groundwater or reservoir storage.

The AWPF is the main component of the GWRS System and the focus of this report. The AWPF treatment train (Figure 2) is comprised of:

- Microfiltration (MF)
- Reverse osmosis (RO)
- Ultraviolet (UV) disinfection and hydrogen peroxide (H_2O_2) addition as an advanced oxidation process (AOP)
- Decarbonation and lime stabilization

Pumping stations and pipelines convey the purified recycled water to the Talbert Seawater Intrusion Barrier and to the percolation basins (ponds) for groundwater recharge to augment potable supplies. GWRS meets all permit requirements for water quality and AWPF plant staff monitor the operational performance of the AWPF's unit processes using critical control points (CCP). The CCPs cover all key elements in the AWPF process at a sufficient frequency to anticipate potential problems and respond before problems become critical.

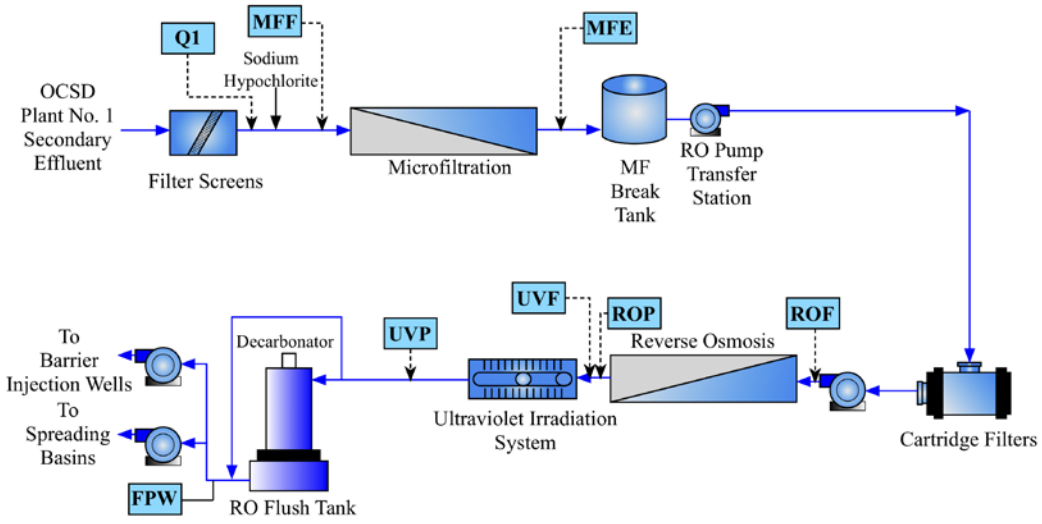


Figure 2.— AWPF process flow diagram. Sampling locations include Q1 (secondary treated wastewater, influent to AWPF), MF Feed (MFF) which is post-sodium hypochlorite addition, MF Effluent (MFE), RO Feed (ROF), RO Permeate (ROP), UV advanced oxidation process (UV/AOP) Feed (UVF), UV/AOP Product (UVP), and final product water (FPW).

Using CCPs allows for continuous online monitoring at multiple locations along each step of treatment through the AWPF. Monitoring demonstrates that the management system and treatment train are functioning according to design and operating within design specifications. This monitoring allows OCWD to track process performance while providing the ability to take corrective action when necessary.

This study collected samples from a subset of the AWPF CCPs, and alongside microbial quantification methods (qPCR), we used an HTS workflow (Figure 3) sequencing either deoxyribonucleic acid (DNA) or ribonucleic acid (RNA) to identify potentially pathogenic microorganisms and to characterize the total microbial community throughout the AWPF (Figure 2).

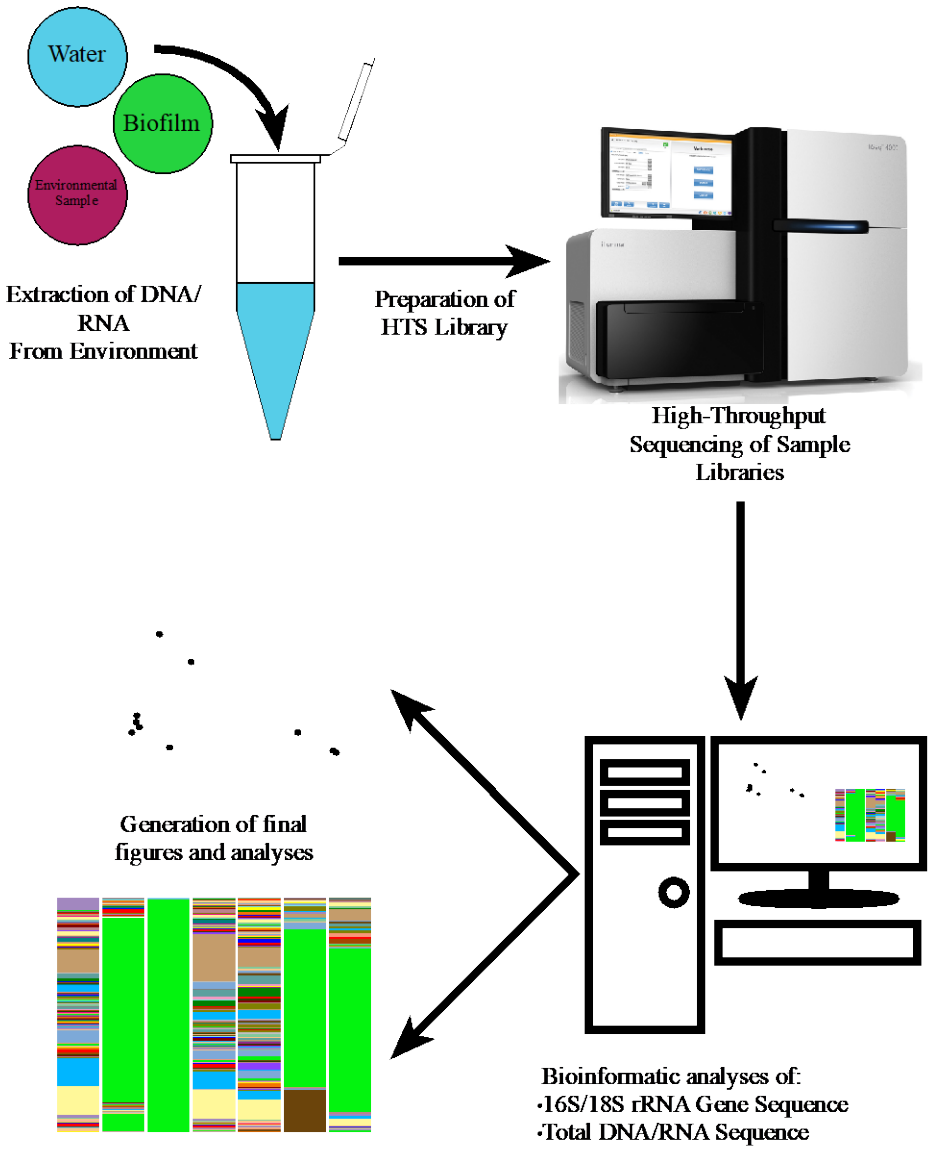


Figure 3.—Conceptual example HTS workflow. DNA or RNA is extracted from environmental samples including water or biofilm from an environment such as the AWPF. Extracted samples are prepared for sequencing, and then sequenced on a HTS instrument. Bioinformatic analyses are carried out either locally or using a tool such as CosmosID before final figures (e.g., the presented bar charts and principal component analyses) and analyses are produced.

1.5. Overall Method

The project sought to better understand the microbial community of the OCWD AWPF through several molecular microbial ecology approaches. Community structure and richness was determined through a combined 16S/18S rRNA gene sequencing approach using recently developed polymerase chain reaction (PCR) primers (Parada et al., 2015), that allows for examination of all three domains of microbial life: Bacteria, Archaea, and Eukarya. To resolve taxonomic information to strain level, the project team used an HTS method developed by CosmosID (Hasan et al., 2014; Lax et al., 2014; and Ponnusamy et al., 2016), which also allows for the identification of functional genes present in the sample (such as antibiotic resistance genes [ARG]) and identifies those genes in the sample that are associated with microbial pathogenicity.

To estimate community density which allows for the estimation of the treatment system to remove Bacteria and Archaea, we used previously designed 16S qPCR primers (Liu et al., 2012) were used. This study also developed sampling methods for water reuse systems for use at OCWD and similar facilities in the future. Specifically, the use of dialysis-style filters was developed that allows for rapid filtration (and sample concentration) of tens of liters to hundreds of liters of water. This method when integrated with qPCR allows for the rapid estimation of microbial biomass.

2. Technical Approach and Methods

2.1. Project Facility and Sampling

2.1.1. Project Facility

The AWPF treatment train uses a multi-barrier, multi-treatment approach including MF, RO, and UV/AOP. This is followed by partial decarbonation and lime stabilization (Figure 2) for pipeline corrosion control purposes due to the high purity, low alkalinity composition of the water. Each step in the treatment process has unique operational demands and potentially, unique microbial communities. A summary of the system, including sampled CCPs, is shown in Figure 2.

The source/influent water to the facility is secondary treated, undisinfected wastewater from OCSD ("Q1" sampling location, see Figure 2). The first step in the AWPF is addition of chlorine (as sodium hypochlorite) at the MF Feed (MFF) location; this water is then fed to the MF process (OCWD, 2018). Sodium hypochlorite is an effective chemical treatment for the inactivation of pathogens within drinking water, and thus some change in the microbial community is expected based on the addition of sodium hypochlorite. MF is a pretreatment step prior to the reverse osmosis (RO) process. Filtrate pumps continuously pull water through the MF membranes located in 36 MF cells using a piping manifold and discharge the filtrate, or MF effluent (MFE), to the MF Break Tank prior to entrance into the RO treatment process (OCWD, 2018). MF uses polypropylene hollow-fiber membranes with a nominal pore size of 0.2 micrometers (μm) to remove suspended and colloidal solids from Q1 water including bacteria and protozoa. Therefore, particulates larger than 0.2 μm remain on the outside of the membrane fiber, and particles smaller than 0.2 μm can lodge in the interstices within the membrane or pass through the membrane. The membrane is not a perfect barrier and a very small number of microorganisms can move further into the AWPF treatment train. The vast majority of Bacteria, Archaea, and Eukarya should, therefore, be removed at this MF step. There is a potential for additional microbial growth within the MF break tank, however few microorganisms are capable of passing through the MF system.

The RO process demineralizes water and removes chemical impurities consisting of both inorganics and organics, and with respect to microbial water quality, removes viruses and a wide range of other contaminants using spiral-wound, thin-film membranes. MF effluent is pumped from the MF Break Tank to the RO system by the RO Transfer Pump Station (Figure 2, not sampled). Pretreatment for the RO process which is after MF, includes adding sulfuric acid and anti-scalant (a threshold inhibitor) to lower the pH of the water and prevent RO

membrane scaling. Cartridge filtration is also performed after this chemical addition and prior to entry to the RO system via the RO Feed (ROF). There are 21 parallel RO treatment units, with 150 pressure vessels per unit at a ratio of 3.25:2:1 vessels per RO unit operated in three stages. Each unit has a rated capacity of 5 mgd, operating over a range of recovery with an average of ~85 %. The reject stream (concentrate) from the RO process constitutes the 15% loss that is sent back to OCSD as a waste product (carrying away the chemical and microbial constituents rejected by the RO process) and permitted discharged to the ocean. The RO permeate (ROP) is directed to the UV/H_2O_2 AOP via the UV Feed (UVF) for further treatment and as a final, water quality polishing step.

Hydrogen peroxide (H_2O_2) is added to the RO permeate, and this water is fed to the UV/AOP system (UV Feed [UVF] sampling location). UV light exposure is used for primary disinfection and for photolysis of UV light-sensitive contaminants such as N-nitrosodimethylamine (NDMA). Hydrogen peroxide exposed to UV light produces hydroxyl radicals that result in advanced oxidation to destroy UV-resistant contaminants such as 1,4-dioxane. A small number of unique chemical compounds such as NDMA, 1,4-dioxane, and other low molecular weight organics are able to pass through RO due to their size and chemical properties, and thus are removed or reduced via UV/AOP. The closed, in vessel type UV system uses low-pressure high output lamps. The UV/AOP system is arranged with thirteen identically sized trains. Each train contains six reactors and has a rated maximum capacity of 8.75 mgd, with a current daily capacity of 100 mgd. The UV/AOP system operates above the California Division of Drinking Water (DDW) minimum required disinfection dose of 101 millijoules per square centimeter (mJ/cm^2). The system was designed to achieve 1.2-log reduction of NDMA and provides a 6-log virus removal credit for DDW Title 22 requirements for injection of potable recycled water. Water after UV/AOP treatment at the UV/AOP product water (UVP) location was sampled, and should, by design, represent the lowest potential biomass environment throughout the treatment train.

Post-treatment at the AWPF consists of partial decarbonation and lime stabilization to raise the pH (from ≈5.5 to 8.5) post RO and UV/AOP treatment and add hardness and alkalinity to make the purified recycled water less corrosive and more stable for distribution systems and aquifer storage. This represents the final product water (FPW), as well as the final CCP at the AWPF.

2.1.2. Sampling Schedule and AWPF General Water Quality

Water was sampled twice at select sampling locations across the AWPF in October 2016 and July 2017, for microbial community analysis as well as general water quality. Sample dates were chosen based on project team availability, and in July 2017, to coincide with the sampling of biofilms from MF and RO membranes. Average water characteristics throughout the AWPF from October 2016 to July 2017 are summarized in Table 1, based on an average of data available from routine OCWD monitoring for this period.

OCSD supplies the secondary treated wastewater for the AWPF. The treated wastewater is comprised of an average blend of about 80% HPOAS operating in NdN mode and 20% TF sourced water, though the blend can be variable. The AWPF treats approximately 130 mgd of OCSD's effluent to produce 100 mgd of product water.

Table 1.—Summary of Average Water Quality Data from October 2016 to July 2017.

Parameter	Units	Q1	ROF	ROP	UVP	FPW
Ca	mg/L	82.21	81.42	0.05	0.05	13.63
Cl	mg/L	294.50	291.67	7.29	7.42	7.08
Electrical Conductivity (EC)	mS/cm	1622.63	1632.45	44.71	45.70	106.64
Fe	µg/L	545.83	120.58	1.93	1.06	0.70
Mg	mg/L	25.07	24.90	0.05	0.05	0.05
Mn	µg/L	53.92	48.08	0.18	0.1	0.4
NO2-N	mg/L	0.51	0.57	0.001	0.01	0.04
NO3-N	mg/L	10.61	10.01	1.27	1.32	1.19
pH	-	7.19	7.08	5.88	5.63	7.83
SIO2	mg/L	21.04	21.53	0.27	1.07	0.45
SO4	mg/L	186.83	198.17	0.21	0.19	0.41
Total Dissolved Solids (TDS)	mg/L	1006.92	1040.96	25.26	24.71	56.62
Total Alkalinity	mg/L	202.75	172.75	7.53	5.73	37.59

mg/L = milligrams per liter
µg/L = micrograms per liter
mS/cm = microSiemens per centimeter

2.1.3. Sampling

Water was sampled across multiple sampling locations (that also are monitored as CCPs) in the AWPF treatment train (Figure 2). All water samples were filtered using the Innova Prep Large Volume Concentrator (LVC) dialysis filter cartridge system (Innova Prep LLC, Drexel, Missouri). This filter-cartridge based system allowed for the filtration of 100 L of water in less than one hour per sample location. Briefly, water was fed through the LVC, and measured in 20 L intervals. Once 60 – 100 L of water were processed per CCP, biomass and particulate were eluted off the filter membrane using the supplied sterile phosphate buffered saline (PBS) elution buffer at a final volume of 50 - 100 mL and retained in sterile 250 mL bottles until further processing. Filtered water was retained in October 2016 for general water quality analysis (below). Q1, MFE, ROF, ROP, UVF, UVP, and FPW water samples were filtered in the initial sampling in October 2016, and Q1, MFF, MFE, ROP, and UVP during the final sampling in July 2017. Sampling points during July 2017 were determined after the identification of sampling locations that appeared to contribute most to the reduction of biomass (MF and RO), and the lack of detectable biology at FPW. Water volumes were 60 L at Q1, MFE, ROF, ROP, and FPW, and 80 L at UVF and UVP in October

2016, and 100 L at Q1, MFF, MFE, and UVP in July 2017. Additionally, biofilm samples were collected from the MF and RO membranes during July 2017 to compliment water samples taken from these locations. Because of the assumed low-biomass nature of the AWPF, filter and DNA extraction control (blank) samples were also taken to compare microbial communities.

2.2. Methodology

2.2.1. General Water Quality and Trace Metal Analysis

To characterize water samples taken in October 2016, we measured major anions (F^-, Cl^-, NO_2^-, NO_3^-, Br^-, PO_4^-, and SO_4^-) using a Dionex ICS-90 ion chromatography system running an AS14A (4 × 250 mm) column. Major cations (B, Ca, Cr, Fe, K, Mg, Na, P, S, Si, and Sr) were also measured using a Perkin-Elmer Optima 5300 DV Inductively Coupled Plasma Optical Emission Spectrometer (ICP-OES). Water was filtered using an 0.22 µm nitrocellulose filter. All ICP samples were acidified with trace-metal grade nitric acid as per standard procedure to ensure mobilization of all metal cations. No further preservation beyond filter sterilization was carried out for cation analysis. There was no chemical characterization of water sampled in July 2017. In addition, several key operational parameters were obtained from online readings every six hours through the AWPF including temperature, pH, turbidity, and total chlorine from September 1, 2016 to July 31, 2017 over the study period.

2.2.2. Scanning Electron Microscopy and Confirmation of Viable Microbial Populations

For scanning electron microscopy and growth, we took a single 0.22 µm nitrocellulose filter from each sample site. Filter papers were imaged on a Hitachi TM-1000 scanning electron microscope (SEM) under environmental conditions with no fixation. Individual filter papers were affixed to the microscope stage with carbon tape and allowed to dry under vacuum prior to imaging. Each filter was imaged under multiple magnification levels, and most of the surface of each sample was scanned to identify any possibly microbial morphology present.

In addition to microscopy, we used cultivation to recover any viable cells. While the vast majority of microbial life has yet to be cultivated (Staley and Konopka, 1985), cultivation can provide useful information as to the status and physiology of viable cultivable microbial populations. Briefly, biomass was re-suspended off of the sample filter by adding using 1 mL sterile phosphate buffered saline (pH 7.4) to the tube containing the filter. The phosphate buffered saline/filter mixture was vortexed vigorously for 30 seconds to remove biomass from the filter surface. Approximately 100 microliters (µL) of the re-suspended cell mass was plated onto both $^1/_{10}^{th}$ strength tryptic soy agar plates (Supplemented with 4.5 grams per Liter sodium chloride to maintain osmotic balance within the medium) and

R2A plates and allowed to grow at room temperature for one week. Any growth was recorded as a positive or negative measurement. Morphologically unique colonies were struck out on successive agar plates until isolated growth was achieved (3 successful transfers of a single colony morphotype), and preserved in 25 percent glycerol at -80 ˚C. After isolation of microbial cultures, no further characterization was carried out.

2.2.3. Small Subunit rRNA Gene Sequencing

Small subunit rRNA gene sequencing is a common microbiological method to identify whole-microbial communities from the environment. The gene is conserved throughout all known extant life, which allows for an ecological survey and identification of what microorganisms are present within an environment by sequencing the gene using HTS. Libraries of Bacterial, Archaeal, and Eukaryotic SSU rRNA gene fragments were amplified from each DNA extraction using PCR with primers (Integrated DNA Technologies Co., Coralville, Iowa) that spanned the ribosomal RNA gene V4 hypervariable region between position 515 and 926 (*E. coli* numbering) that produced a ~400 basepair (bp) fragment for Bacteria and Archaea, and a 600 bp fragment for the Eukaryotes. These primers evenly represent a broad distribution of all three domains of life (Parada et al., 2015).

The forward primer 515F-Y (**GTA AAA CGA CGG CCA G** CCG TGY CAG CMG CCG CGG TAA-3') contains the M13 forward primer (in bold) fused to the ssuRNA gene specific forward primer (underlined) while the reverse primer 926R (5'-CCG YCA ATT YMT TTR AGT TT-3') was unmodified from Parada et. al., 2015. 5 PRIME HOT master mix (5 Quanta BioSciences Inc., Gaithersburg, Maryland) was used for all reactions at a final volume of 50 µL. Reactions were purified using AmpureXP paramagnetic beads (Beckman Coulter Inc., Indianapolis, Indiana) at a final concentration of 0.8 x v/v. After purification, 4 µL of PCR product was used in a barcoding reaction to attach a unique 12 bp barcode to each library in duplicate 50 µL reactions. A mock community was also used as a positive control (Zymo Inc., Irvine, California). Duplicate reactions were pooled, purified using AmpureXP beads to a final volume of 40 µL, quantified using the QuBit HS DS DNA assay kit (Thermo Fisher Scientific Inc., Waltham, Massachusetts), and pooled in equimolar amounts before concentration using an Amicon 30 K centrifugation column (Merck Millipore, Burlington, Massachusetts) to a final volume of 80 µL. To mitigate the effects of reagent contamination (Salter et al., 2014), multiple extraction blanks and negative controls were sequenced. The pooled, prepared library was then submitted for sequencing on the Illumina MiSeq (Illumina Inc., San Diego, California) using V2 PE250 chemistry.

2.2.4. Quantitative PCR

Quantitative PCR is a useful tool to provide absolute values of a gene target of interest. In our case, qPCR was used to provide an absolute estimate of 16S rRNA genes within the AWPF, which is somewhat analogous to the number of microorganisms present within each CCP. To estimate the total bacterial/archaeal SSU rRNA gene count, we used a TaqMan based probe assay and primer set developed by Liu et. al (Liu et al., 2012). The assay was carried out in 25 µL reactions containing 1x final concentration of PerfeCTa qPCR ToughMix (Quanta BioSciences Inc.), 1.8 micromolars (µM) of each primer, and the probe at a final concentration of 225 nanomolars (nM). Each biological replicate was also assayed in technical triplicate. A seven-point standard curve was generated using serially diluted genomic *E. coli* DNA in triplicate. Results were reported as copies of 16S rRNA gene per mL water, or 16S/mL.

2.2.5. Metagenomic/Transcriptomic Sequencing and Bioinformatics Analyses

Total DNA or RNA sequencing from the environment (also known as metagenomic (DNA) or metatranscriptomic (RNA) sequencing) allows for a more highly detailed survey of the microorganisms present in any environmental sample. By sequencing total DNA from the environment, we expect to identify Bacteria, Archaea, or Eukarya to the species or strain level and to identify genes associated with virulence, or antibiotic resistance. Metatranscriptomic sequencing will provide this, and also allow identification of what genes are associated with microorganisms that are active, or currently metabolizing within an environment. CosmosID, Inc. prepared and sequenced the metagenomic and metatranscriptomic samples. Raw extracted DNA was provided by the Colorado School of Mines team to CosmosID after quantification using the Qubit HS or BR assay (Thermo Fisher Scientific Inc., Waltham, Massachusetts). Total RNA was converted to complementary DNA (cDNA) using the Protoscript II reverse transcriptase (New England Biolabs, Ipswich, Massachusetts). Second-strand cDNA synthesis was carried out using the NEBNext® Ultra II Non-Directional RNA Second Strand Synthesis module (New England Biolabs, Ipswich, Massachusetts), and final cDNA samples were quantified using the Qubit HS assay. Ribodepletion, a step that removes ribosomal RNA (rRNA) that is typically 80 to 90 percent of the RNA in a sample, was not attempted and total cDNA was submitted for sequencing at CosmosID. Because of the mixed nature of an environmental sample containing Bacteria, Archaea, and Eukarya, and the lack of a single efficient protocol to remove rRNA from each sample, it was determined the best approach was to not attempt ribodepletion. Each DNA sample was normalized in 3-18 µL of nuclease-free water for a final concentration of 0.5 ng µL-1 using the Biomek FX liquid handler (Beckman Coulter Life Sciences, Brea, California).

Libraries were constructed using the Nextera XT Library Prep Kit (Illumina, San Diego, California), followed by 13 cycles of PCR amplification using Nextera i7 and i5 index primers and 2X KAPA master mix in a modified Nextera XT protocol. The PCR products were purified using 1.0X speed beads and eluted in 15 µL of nuclease-free water and quantified by PicoGreen fluorometric assay (100X final dilution). The libraries were pooled by adding an equimolar ratio of each based on concentration determined by PicoGreen and loaded onto a high sensitivity (HS) chip run on the Caliper LabChipGX (Perkin Elmer, Waltham, Massachusetts) for size estimation. Pooled libraries were then sequenced using an Illumina HiSeq 3000 flowcell using PE150 chemistry. Unassembled metagenomic sequencing reads were directly analyzed by CosmosID bioinformatics software package (CosmosID Inc., Rockville, Maryland) described elsewhere (Ottensen et al., 2016; Ponnusamy et al., 2016; Hasan et al., 2014; and Lax et al., 2014) to identify at the species, subspecies, and strain levels and to quantify the organisms' relative abundance. Community diversity estimates and strain-level identities presented in this work are given as microbial species or strains that are above the confidence threshold of the CosmosID bioinformatic pipeline. That is, any sequence that CosmosID can successfully identify to the species or strain level with greater than 95 percent confidence is considered to be "above the threshold." Similarly, the community resistome and virulome, the collection of antibiotic resistance and virulence genes in the microbiome respectively, were also identified using the CosmosID curated antibiotic resistance and virulence gene database. Comparative metagenomic analyses (principal component analysis, heatmaps, and other statistical analyses) were done to determine changes and shifts in diversity correlated with different stages of treatment.

2.3. SSU rRNA Gene and Metagenomic Analyses

We de-multiplexed sequence reads using QIIME version 1.9.1 (Caporaso et al., 2010b). Sequence reads were first denoised and then clustered into sub-operational taxonomic units (sOTU) using UPARSE (Edgar, 2013). After clustering, sOTUs were assigned taxonomy using mothur (Schloss et al., 2009) against the SILVA database (r123, (Pruesse et al., 2007)). Each OTU was then aligned against the SILVA r123 database using pyNAST (Caporaso et al., 2010a) and filtered to remove uninformative bases. Then we created a tree using the maximum likelihood method and the Jukes Cantor evolutionary model within FastTree 2 (Price et al., 2010). For downstream analyses, we then produced a BIOM-formatted file (McDonald et al., 2012). To limit OTUs originating from contaminating microorganisms found in extraction and PCR reagents (Salter et al., 2014), we processed all extraction blanks and PCR controls and a computed a core microbiome. Any sOTU found in all controls was filtered from the overall dataset. Taxa heatmaps were generated using phyloseq (McMurdie and Holmes, 2013) and AmpVis.

Bacterial, Archaeal, and Eukaryotic taxa were identified from total RNA and DNA sequenced using HTS by a proprietary identification platform allowing for taxa identification to the strain level developed by CosmosID (Hasan et al., 2014; Lax et al., 2014; and Ponnusamy et al., 2016). Antibiotic resistance markers and virulence factors were similarly identified and cataloged using the CosmosID identification pipeline. To visualize how samples were related based on the relative abundance of identified Bacterial, viral, or ARG sequences within and between each sample, the project team used principal component analysis (PCA) generated by CosmosID. PCA is a way to expand and rotate data that allows for visualization of how samples are related by maximizing the variance (or distance within the PCA) between samples to determine what samples are more alike than others. In this case, each PCA was generated by calculating a distance matrix based on the relative abundance of detected Bacteria, viruses, or ARGs. This distance matrix was then used to produce the PCA for the Bacterial, viral, or ARG dataset. Samples that appear more closely together within a PCA are assumed to be more similar in sequence composition for each dataset. A more comprehensive discussion of PCA and its applications can be found in Jolliffe and Cadima (2016). To determine the diversity of bacteria, viruses, and ARGs, we calculated the Chao1 index (Chao, 1984) by using the frequency of detected organisms (Bacteria or viruses) or ARGs that were determined to be above the confidence threshold of detection. The confidence threshold is determined to be any sequence identified to the species or strain level that the CosmosID algorithm is $\geq$ 95 % confident in.

Additionally, we assembled metagenomes to provide further additional information and identification to characterize the system microbial ecology. Sequence reads were trimmed and adapters removed using the software tool PEAT (Li et al., 2015b) prior to assembly. To reduce the computational complexity of assembly, individual samples also underwent a k-mer reduction strategy using the software tool BBnorm (DOE, 2018) prior to concatenation of all samples, and co-assembly using the software tool MEGAHIT (Li et al., 2015a). Trimmed sequence reads from each sample were then mapped back to the co-assembly using the software tool Bowtie 2 (Langmead and Salzberg, 2012) to generate coverage profiles suitable for use with the software tool Anvi'o (Eren et al., 2015). A metagenome assembled genome (MAG) binning approach was used to recover near-complete microbial genomes from the sample set using the software tool CONCOCT (Alneberg et al., 2014), and finally refined and visualized within Anvi'o. Putative MAG identification was provided using the software tool CheckM (Parks et al., 2014). CheckM and Anvi'o give estimates of MAG "completeness" (how much of a putative microbial genome is represented by the MAG), and "contamination"(how much DNA sequence contained within the MAG is potentially from another microorganism), and not from the MAG itself. Genome GC content (the percentage of guanine and cytosine [two nucleotides]) and length (the total number of basepairs [bp]) within a MAG are also estimated by Anvi'o, and are useful as quality control estimates of a MAG.

3. Results and Discussion

3.1. Results

3.1.1. General Water Quality and Operational Characteristics

Several key operational parameters were obtained from online readings every six hours through the AWPF including: turbidity, temperature, pH, and total chlorine from September 1, 2016 to July 31, 2017. A summary of water quality data obtained from OCWD are presented in Table 1. Results included:

- **Turbidity.** Turbidity declined from averages of 3.6 NTU in Q1 (AWPF source water that is secondary treated wastewater effluent) to 1.6 NTU at MFF, to 0.08 - 0.04 NTU at ROF, ROP, and FPW. Turbidity is monitored at AWPF as a general water quality indicator and measure of solids and particles, demonstrating effective treatment through sequential unit processes.

- **Temperature.** Temperature was stable across the system during the sampling period measured, averaging near 80 F (≈ 26.7 C), ranging from a low of 79.5 F at ROF to a maximum average of 80.8 F at FPW.

- **pH**. pH was stable near 7.0 at MFF and ROF, declining to 5.51 to 5.64 at ROP and UVP, before increasing to an average of 8.52 at FPW due to addition of calcium hydroxide (lime) before the FPW exited the AWPF.

A summary of other water quality data obtained from OCWD are presented in Table 1 provided previously in Section 2. Concentrations of various constituents declined after RO treatment, a trend also reflected in trace metal analyses performed on water obtained in October 2016 (Figure 4). All major anions and cations declined to near detection limits, with a minor increase in nitrate from ROP to UVP (Figure 4). Similar to the decline in detectable trace metals and turbidity observed across the RO membrane, DNA and RNA quantities extracted declined across the treatment train (Table 2). These changes in water quality were expected for the advanced treatment process.

Table 2.—Summary of Total DNA or cDNA Quantities Extracted

Sample Site	October 2016[a]	July 2017[a]
Q1	1,604	4,524
Q1 (RNA)	-	33
MFF	-	4,599
MFF (RNA)	-	58
MFE	BDL[b]	53
ROF	BDL[b]	-
ROP	BDL[b]	22
UVF	BDL[b]	-
UVP	5.8	22
MF Biofilm	-	2,150
MF Biofilm (RNA)	-	3
RO Biofilm	-	4,066
RO Biofilm (RNA)	-	5
FPW	BDL[b]	-

[a]Quantities given in nanograms (ng).
[b]Below Detectable Limit (BDL) of Quantification Assay.
- indicates not measured (no sampling conducted)

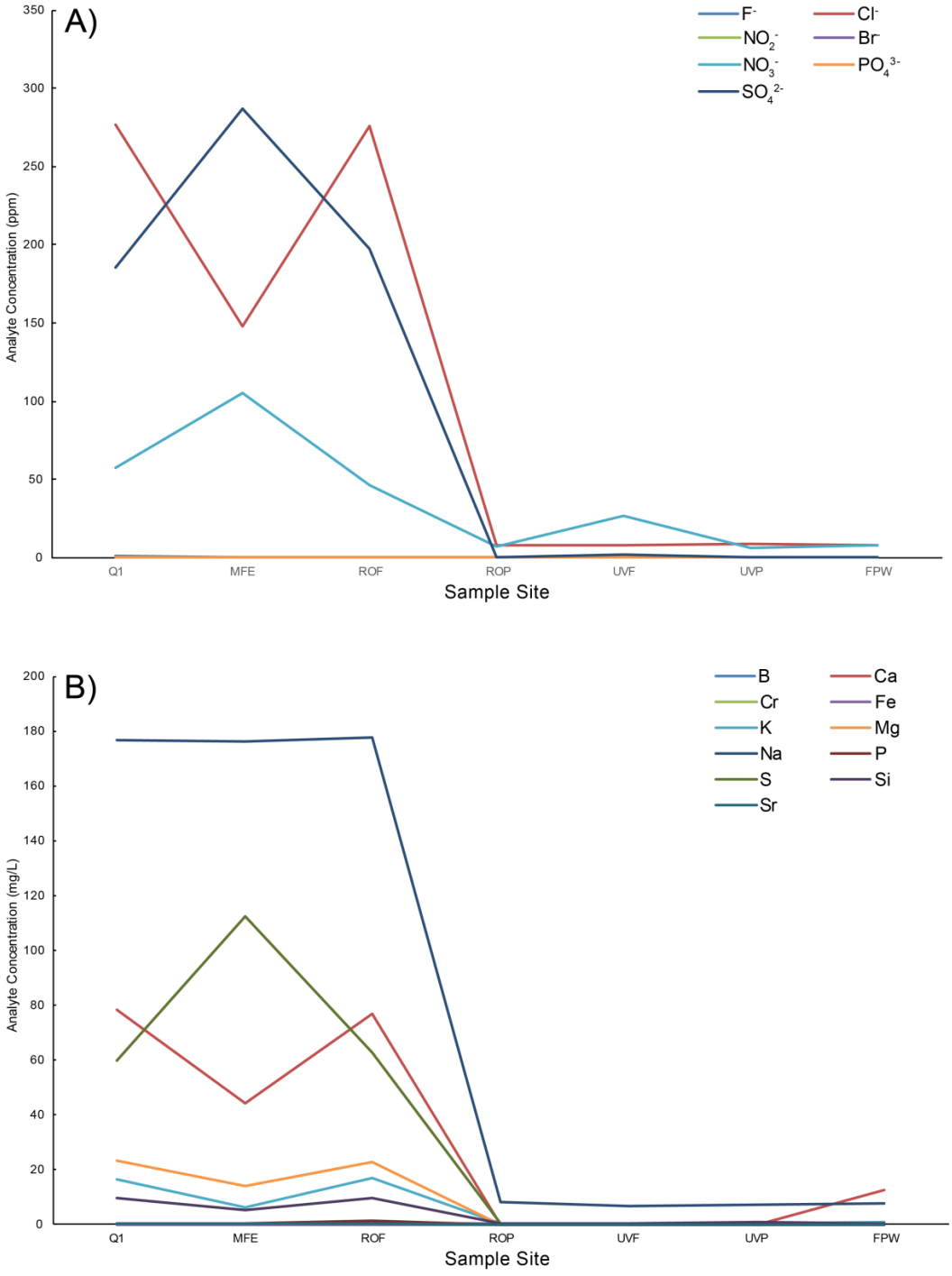


Figure 4.—Anion (A) and Cation (B) concentrations through the AWPF taken in October 2016. Anions are shown in parts per million (ppm), while cations are shown in mg/L.

3.1.2. Microscopy and Cultivation of October 2016 Samples

Overview images obtained from October 2016 samples are shown in Figure 5. Diatoms were visible at Q1, in addition to apparent microbial biofilm. Little to no particulate was present on filters obtained post MF. Distinct bacterial morphology was difficult to identify by SEM, although cells were heavily embedded in extracellular polymeric substance (EPS) in samples taken from Q1 at both sampling events (Figure 5). Cultivation supported the conclusion that visible microbial life was absent after microfiltration. A total of 36 distinct morphotypes were obtained, with only 6 identified from MFE. No viable colonies were obtained post MFE.

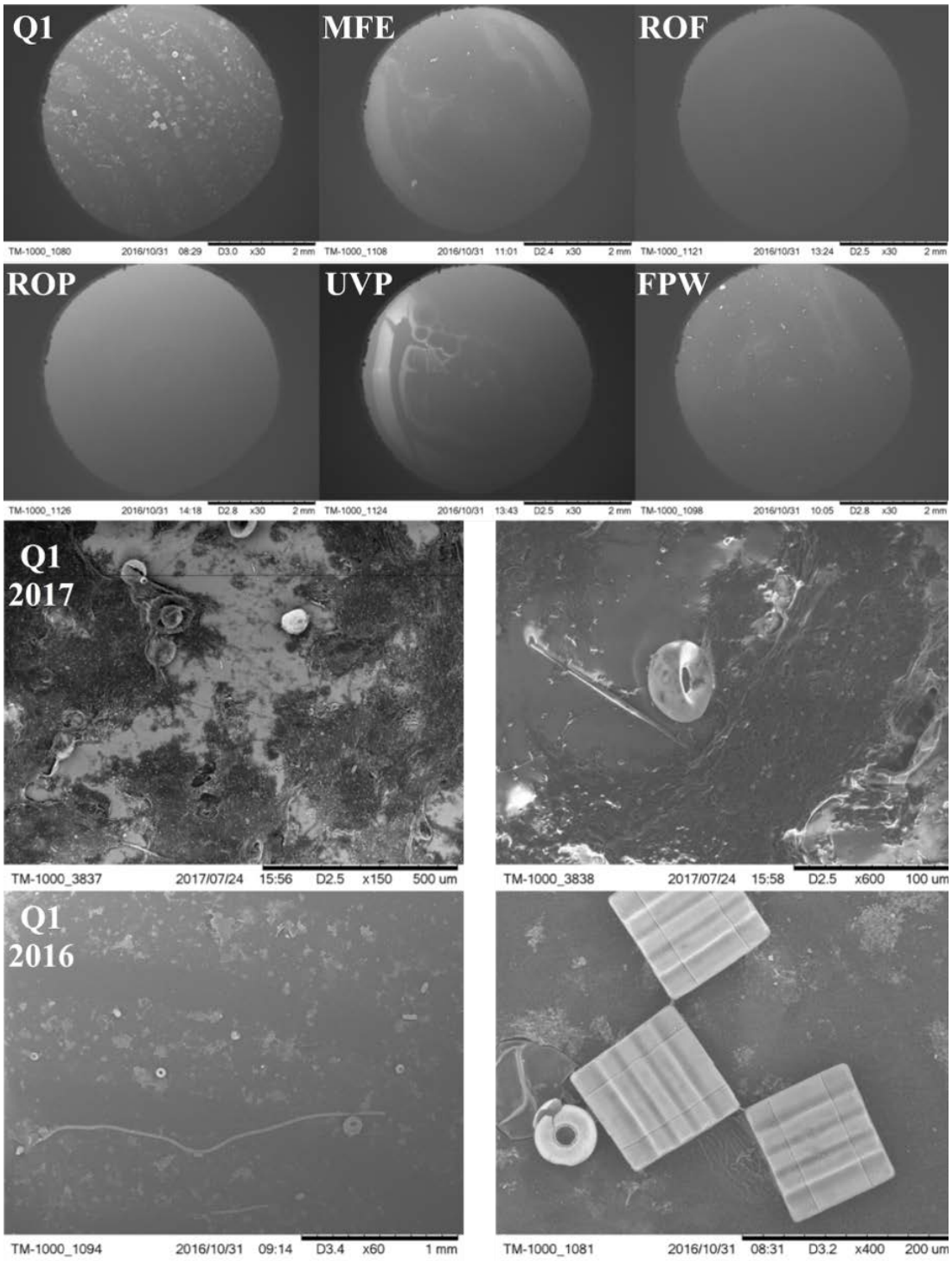


Figure 5.—Scanning electron microscopy (SEM) imaging of filters taken throughout the AWPF. The upper six images are from the initial October 2016 sampling event, while the lower, larger four images are a comparison of biofilms identified from both the 2016 (lower) and 2017 (upper) sampling at Q1.

3.1.3. Identification of AWPF Microbial Community

Communities at Q1 and MFF were highly similar. Differences in DNA, RNA and 16S rRNA gene microbial community composition at Q1 and MFF were mostly confined to the Clostridia and Betaproteobacteria, with sOTUs most closely related to *Romboutsia* and Clostridium *sensu strictu* group 1 and *Thauera* in greater abundance in RNA samples. Otherwise, sOTUs most closely related to *Flavobacterium, Zoogloea,* and unclassified Comamonadaceae were abundant at both locations. DNA samples from 2016 and 2017 were highly concurrent.

A large shift in the composition of the microbial community between Q1/MFF and MFE was apparent, with the complete removal of most abundant sOTUs found at Q1 and MFF (Figure 6). In their place, sOTUs within the Bacterial phylum Proteobacteria most closely related to *Nitrosomonas,* DSSF69 (Alphaproteobacteria), *Dongia,* and an sOTU within the Actinobacteria most closely related to the genus *Mycobacterium* were present in both DNA and RNA samples. The microbial diversity of samples from ROF, ROP, UVF, UVP, and FPW were similar to Q1 and MFF, although some sOTUs found in Q1 or MFF were not present in the samples taken after MF (Figure 6).

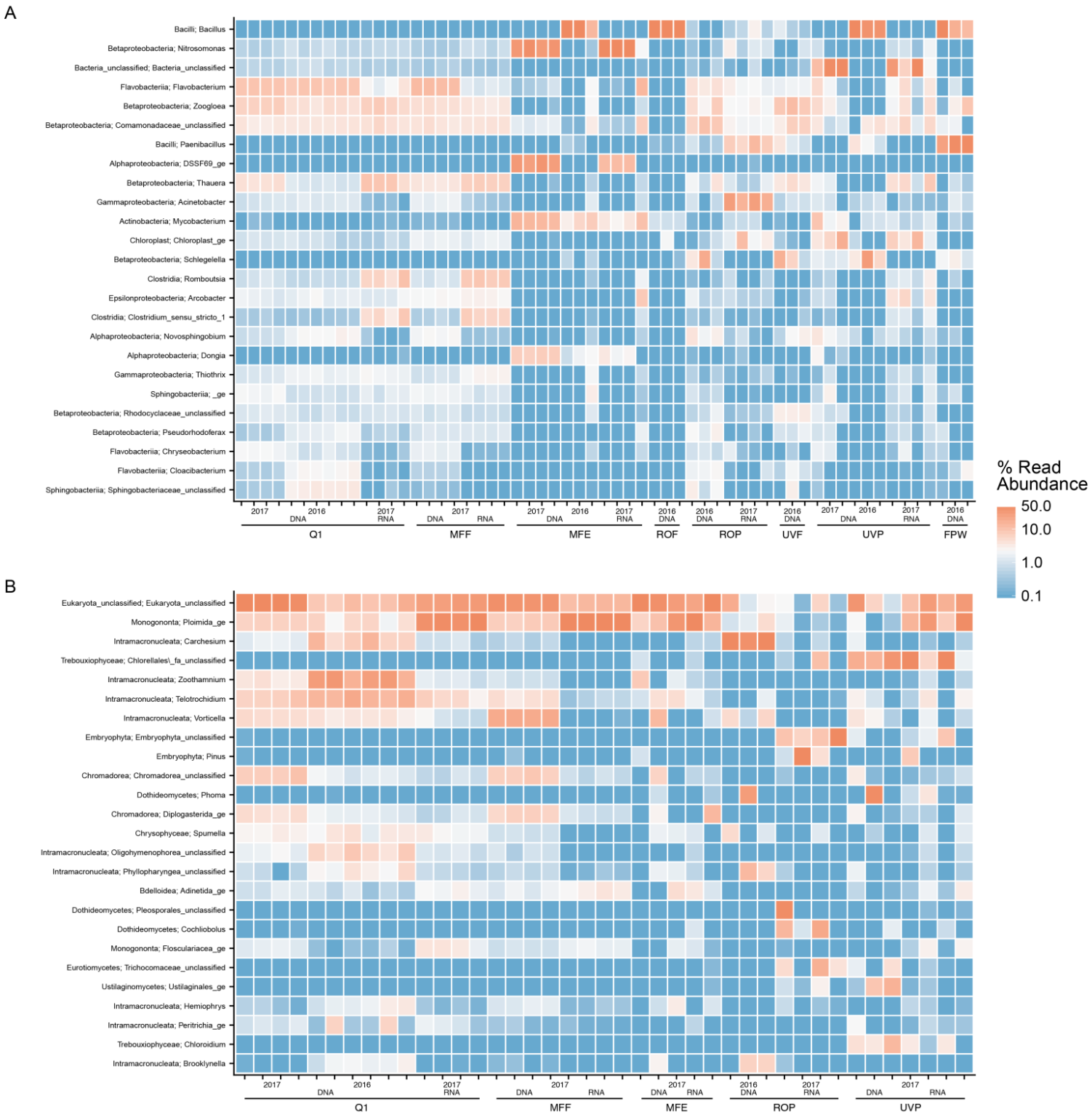


Figure 6.—Taxonomy summary of samples taken in 2016 and 2017 from 16S/18S rRNA gene sequencing. The top 25 Bacterial/Archaeal (A) or Eukaryotic (B) sOTUs are shown for each dataset.

3.1.4. HTS of Environmental DNA and RNA from AWPF

Despite optimized DNA extraction methods, water samples beyond Q1 in October 2016 produced insufficient quantities of DNA for metagenomic sequencing while still allowing for 16S/18S rRNA gene sequencing. Water samples beyond MFE in July 2017 produced insufficient quantities of DNA or RNA to obtain metagenomes or metatranscriptomes. However, we successfully sequenced these water samples using 16S/18S rRNA gene sequencing. The highest concentration of DNA was recovered from Q1 (2016 and 2017 sampling) and MFF water and RO biofilm samples (Table 2). Despite some samples containing less than the required amount of DNA for sequencing, we successfully obtained a total of 380 million paired sequence reads from Q1 (2016 and 2017), MFF, MFB, ROB, MFE, as well as RNA samples from Q1 (2017), MFF, MFB, and ROB. Overall, DNA and cDNA from RO-biofilm samples were more closely related to each and clustered together by PCA when compared to MF-biofilm and Q1-water (Figure 7). Based on the DNA/metagenomes and cDNA/metatranscriptomes, all three samples differed, with the greatest diversity observed between DNA and cDNA in Q1-water.

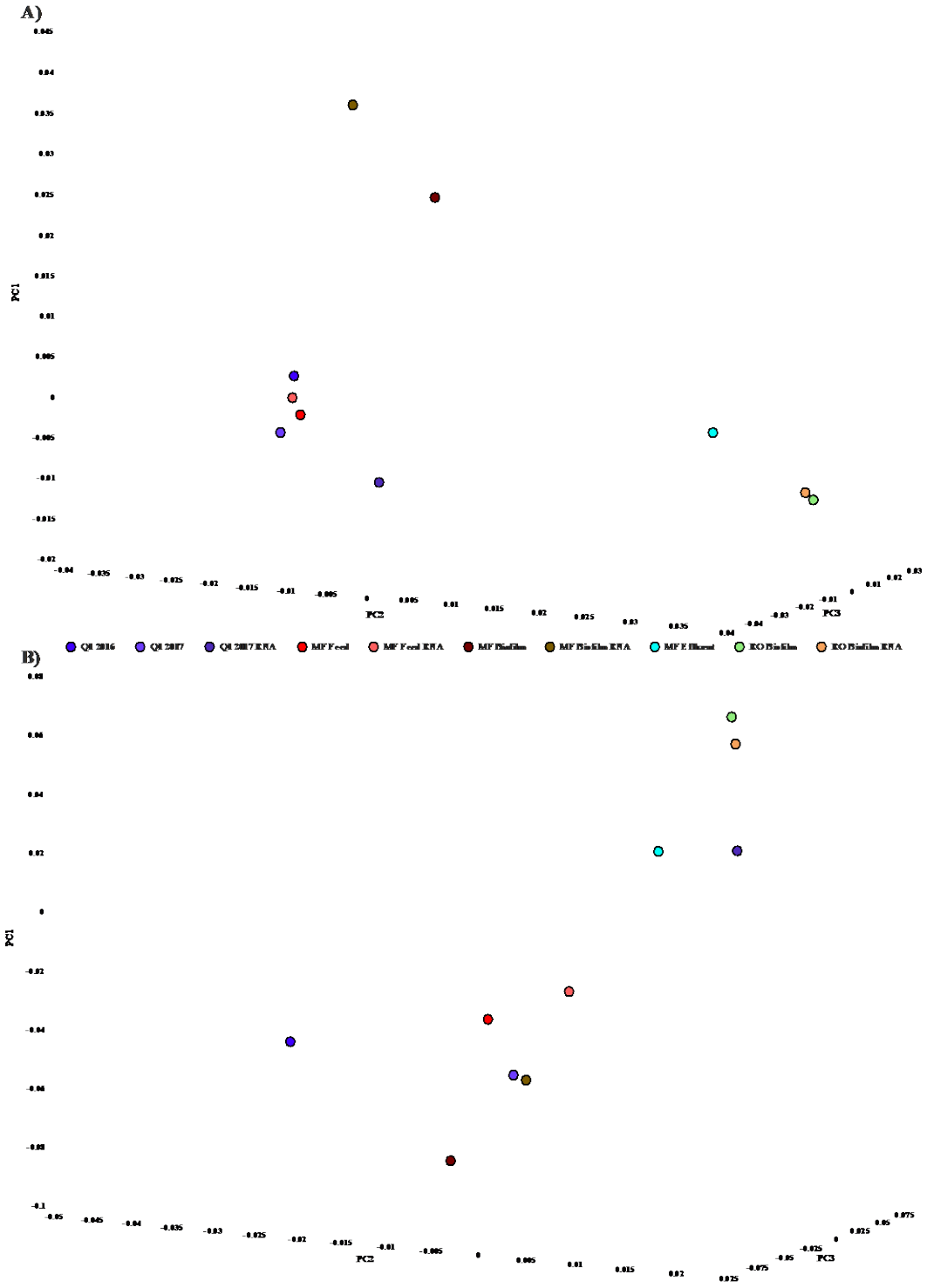


Figure 7.—Principal component analysis (PCA) of DNA and RNA samples taken in October 2016 and July 2017 for bacteria (A) and viruses (B) identified using the CosmosID pipeline. Principal component axis (PC1,2,3) values are the raw component scores of the PCA model. More critically, sample distance within the PCA is proportional to how similar a sample is to another.

More total bacterial species were identified at Q1 (2016 and 2017) than other sampled sites by the CosmosID analysis pipeline (Figure 8). A total of 2,180 bacterial species were identified at Q1 water in 2016, compared to 1,864 species in the 2017 Q1 water sample, correlating to a Chao1 diversity estimate of 790.40 and 637.78 respectively of Bacterial species above the CosmosID confidence threshold (Figure 8). Greater Chao1 diversity values are suggestive of greater species diversity, or more microorganisms present within a sample. Threshold values reported by the CosmosID database are defined by any sequence identified to a species or strain level with greater than 95 percent confidence. Biofilm samples from MF and RO were less diverse than Q1 water sample, with species diversity estimated to be 64% lower in the RO and 38% lower in the Q1. The lowest number of Bacterial species were identified in the MF Effluent water (MFE), which was only 11% of the estimated diversity in Q1 water from 2016 (Figure 8). Samples from Q1 and MFF (including RNA/metatranscriptomes) clustered together by PCA visualization, while post filtration samples (including ROB and MFE) clustered together (Figure 7).

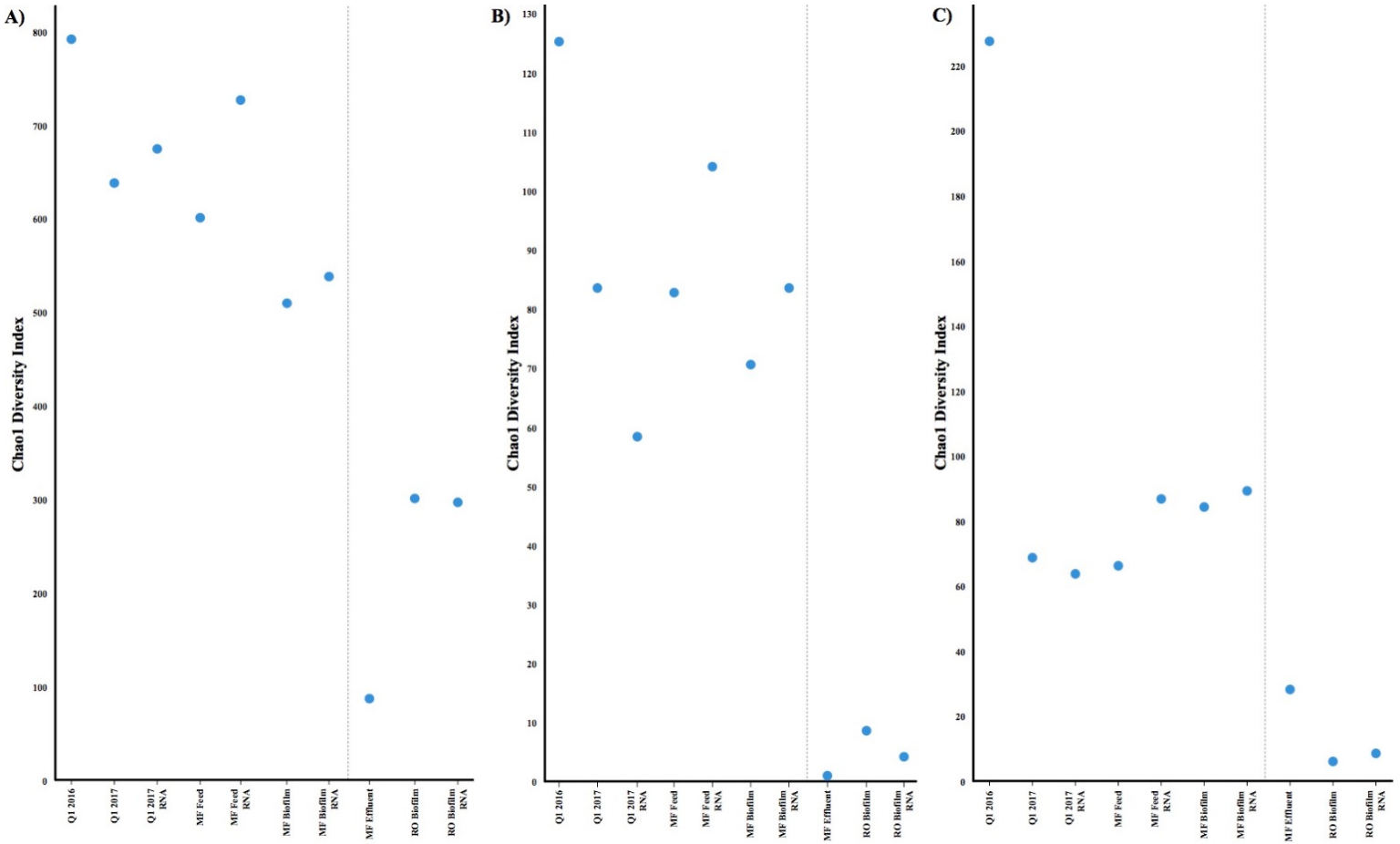


Figure 8.—Species diversity estimated by the Chao1 diversity index for the Bacteria (A), ARGs (B), and Viruses (C). Higher numbers indicate a greater number of species identified using the CosmosID pipeline. Values were calculated using filtered identities within CosmosID, representing > 95% confidence of the presence of a species.

A total of 117 MAGs were recovered from the DNA HTS co-assembly of varying genome completeness (Figure 9). Multiple MAGs were recovered from the most abundant taxonomic lineages identified in the 16S rRNA gene survey, including near complete genomes from the genera *Dongia, Zoogloea,* and *Thaurea,* which were also represented in the 16S rRNA gene dataset (Figure 6). No Eukaryotic MAGs were recovered. Both Q1 and MFF samples contained a greater diversity of recoverable MAGs than post microfiltration samples. Both MFB and ROB samples contained only a small number of MAGs in greater abundance, mostly from the *Mycobacteria.*

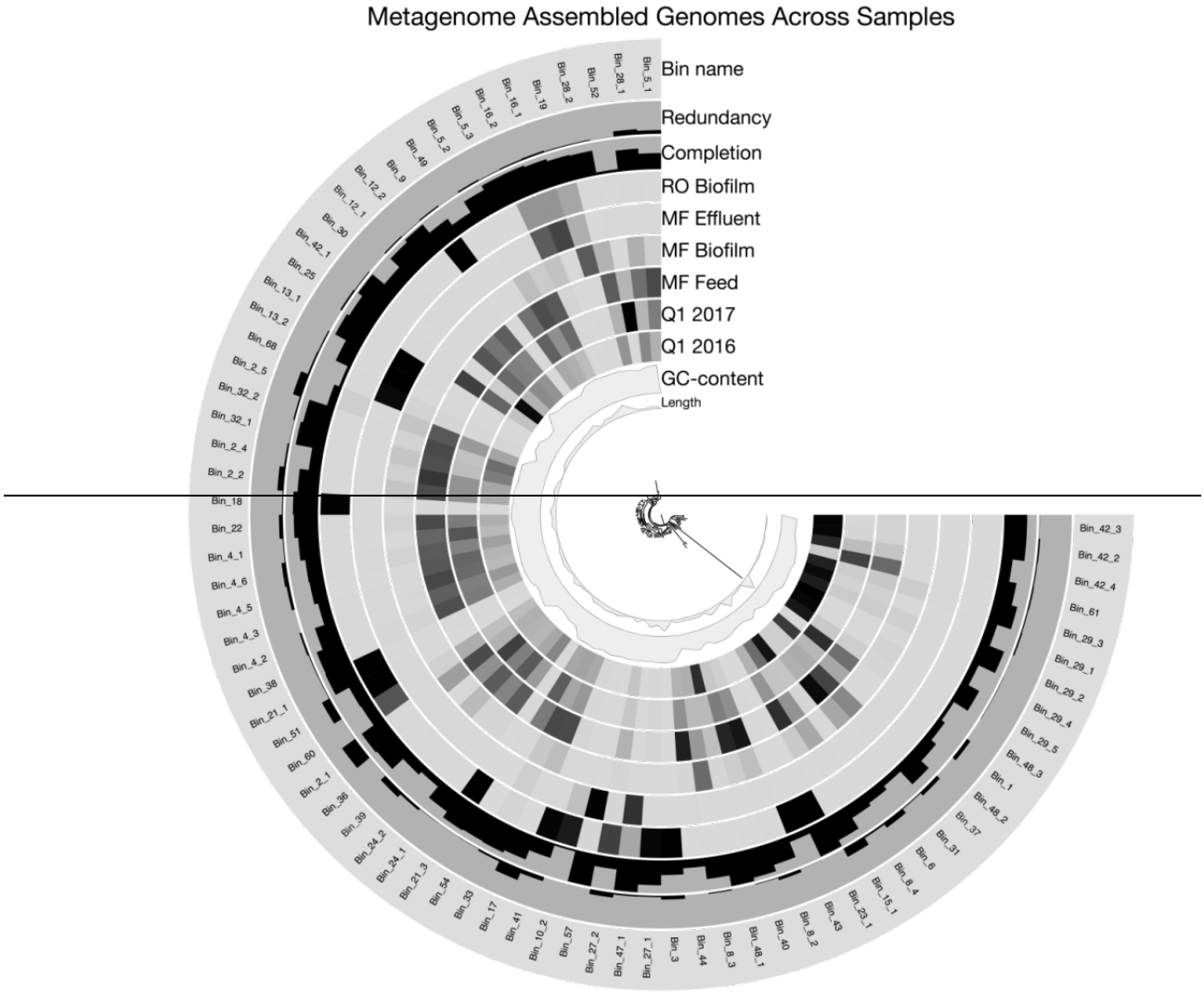


Figure 9.—Overview of metagenome assembled genomes (MAG) recovered from the AWPF. Completion and redundancy, or contamination, are represented by bars in the outer rings, and GC-Content and Genome Length are shown as a line plot in the innermost rings. The intensity of each rectangle reflects the relative abundance of each MAG within each sample.

3.1.5. Eukaryote Species Diversity Based on Metagenomic and Transcriptomic Sequencing

Both fungi and parasites were identified in metagenomic and metatranscriptomic sequence libraries. The greatest number of fungal species was identified in Q1 water (July 2017 sampling) from DNA/metagenomic and RNA/metatranscriptomes analysis (n=24) using the CosmosID identification platform. A large shift in fungal species was observed between Q1 water DNA and cDNA samples (Figure 10). The least number of fungal species were identified in the RO biofilm samples. Additionally, similar fungal species were identified in both the RO biofilm DNA (n=16) and cDNA (n=15) samples. In all samples collected during the 2017 sampling event, the following eight (8) fungal species were identified: *Clavaria fumosa*, *Enterocytozoon bieneusi*, *Epichloe sylvatica*, *Lentinus polychrous*, *Malassezia globosa*, *Malassezia restricta*, *Mitospordium daphnia*, and *Puccinia arachidis*. Even though these fungal species were identified in all samples, their relative abundance varied between samples (Figure 10).

Six parasite species were identified in all samples. However, their relative abundance varied between samples (Figure 11): *Acanthamoeba mauritaniensis*, *Acanthamoeba palestinensis*, *Hammondia hammondi* strain H, and *Paramecium biaurelia* strain V14. More parasite species were detected in the Q1 water (July 2017 sampling) DNA sample (n=38) compared to Q1 water (July 2017) RNA sample (n=31). Additionally, a large shift was observed between Q1 water with DNA and RNA, with removal of many organisms in the RNA sample, therefore, identifying fewer active parasites in the RNA sample (Table 3). The fewest parasite species were identified in the MF biofilm RNA sample, suggesting even fewer active parasites in the MF biofilm, possibly due to the activity of sodium hypochlorite on parasites within the biofilm.

Table 3.—Summary of identified taxa or ARGs at each sample location identified by the CosmosID Bioinformatic pipeline. Values in parentheses are total hits, otherwise values given are those above the 95 percent confidence threshold.

Sample Site	Bacteria	Fungi	Protists	Viruses	ARGs
Q1 2016	753 (2,180)	9 (9)	13 (13)	192 (192)	122 (203)
Q1 2017	601 (1,864)	7 (24)	14 (38)	69 (185)	81 (163)
Q1 2017 (RNA)	651 (1,638)	10 (24)	14 (31)	63 (200)	58 (134)
MFF	568 (1,791)	7 (26)	9 (29)	66 (172)	81 (163)
MFF (RNA)	695 (2,067)	8 (30)	13 (32)	87 (222)	100 (192)
MF Biofilm	489 (1,445)	8 (19)	9 (26)	84 (220)	68 (163)
MF Biofilm (RNA)	522 (1,317)	7 (19)	8 (18)	88 (225)	84 (163)
MFE	83 (286)	3 (13)	5 (12)	29 (109)	1 (7)
RO Biofilm	293 (821)	8 (16)	14 (22)	6 (43)	8 (33)
RO Biofilm (RNA)	291 (743)	7 (15)	9 (20)	8 (41)	4 (743)

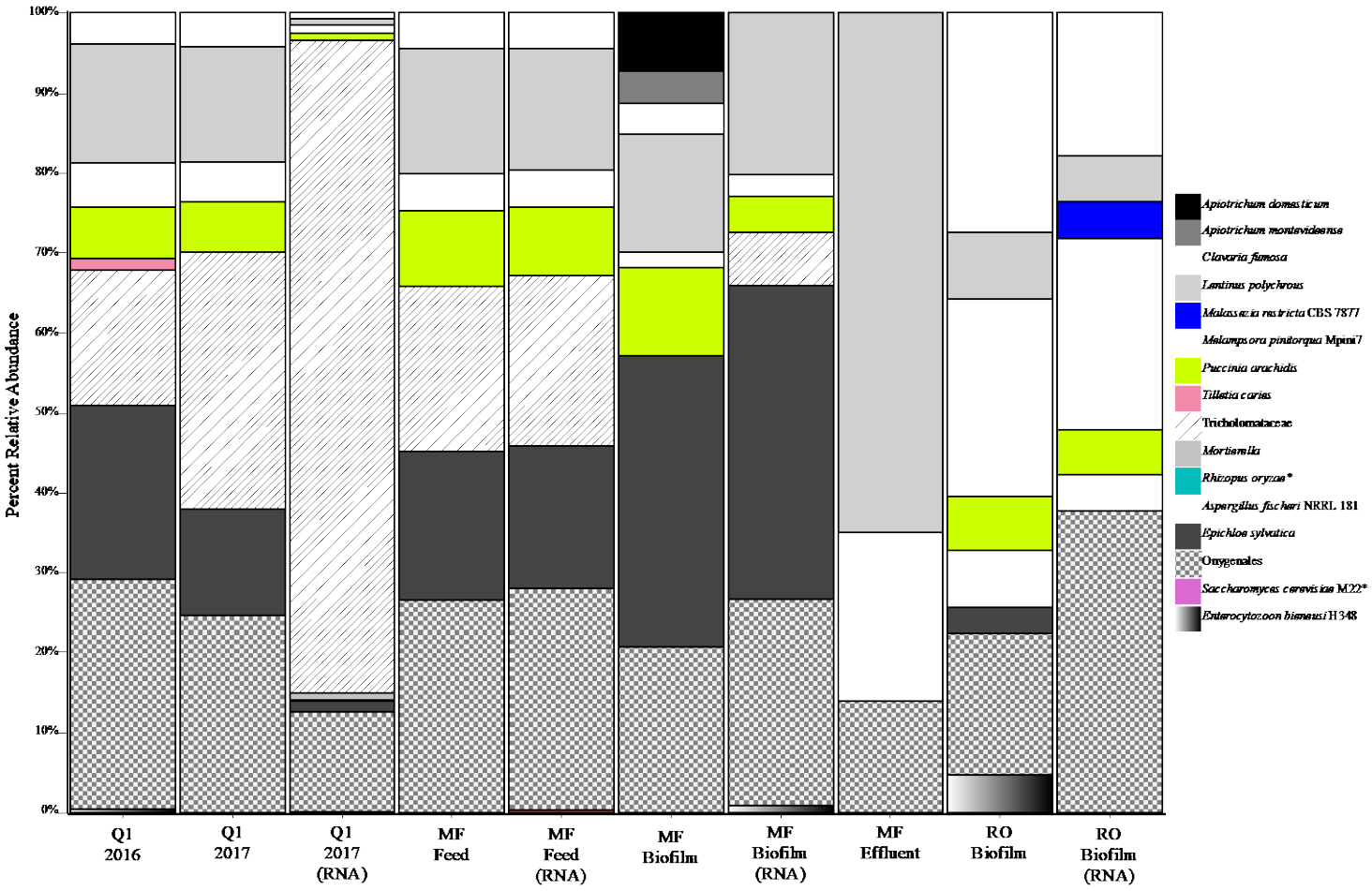


Figure 10.—Relative abundance bar chart of the diversity of fungal species identified across the AWPF. *Rhizopus oryzae* and *Saccharomyces cerevisiae* M22 are denoted with an asterisk to note that they are too low of a relative abundance to be visualized on the chart.

3.1.6. Identification of Viruses and ARGs in the AWPF

In addition to identifying Bacteria, Archaea, and Eukarya, the CosmosID pipeline can identify viral sequence and antibiotic resistance genes (ARG). As noted previously, HTS analyses of viruses, ARGs, and other species were not possible for samples beyond MFE due to the limited amount of DNA obtained. The greatest abundance of viral sequence was identified in Q1 water compared to any other sample in the treatment train, where a slightly greater diversity was identified at Q1 in 2016 at Q1 (192 total viral species identified) compared to 2017 at Q1 (185 total viral species identified, Table 3). The viral community identified in DNA and RNA samples were distinct, grouping similarly to the Bacterial community by (Figure 7). Greater numbers of viral sequence were identified in the Q1 2016 water sample than in the Q1 2017 water sample (Figure 8). The viral DNA sequences identified in both Q1 water samples were different from one another (Figure 12). However, the vast majority of the viral sequences identified were bacteriophages rather than human viruses (Figure 12).

A significantly lower number of viruses, including DNA and RNA viruses, were identified in the RO biofilm (Table 3), suggesting that most of the viruses were removed following MF treatment. Pepper mild mottle virus (PMMoV) is one of the most abundant RNA viruses found in a metagenomic survey of RNA viruses from human feces (Rosario et al., 2009) and thus it is unsurprising to find PMMoV in the treatment system. Recently, PMMoV has been used as an indicator of human fecal contamination in wastewater (Hughes et al., 2017).

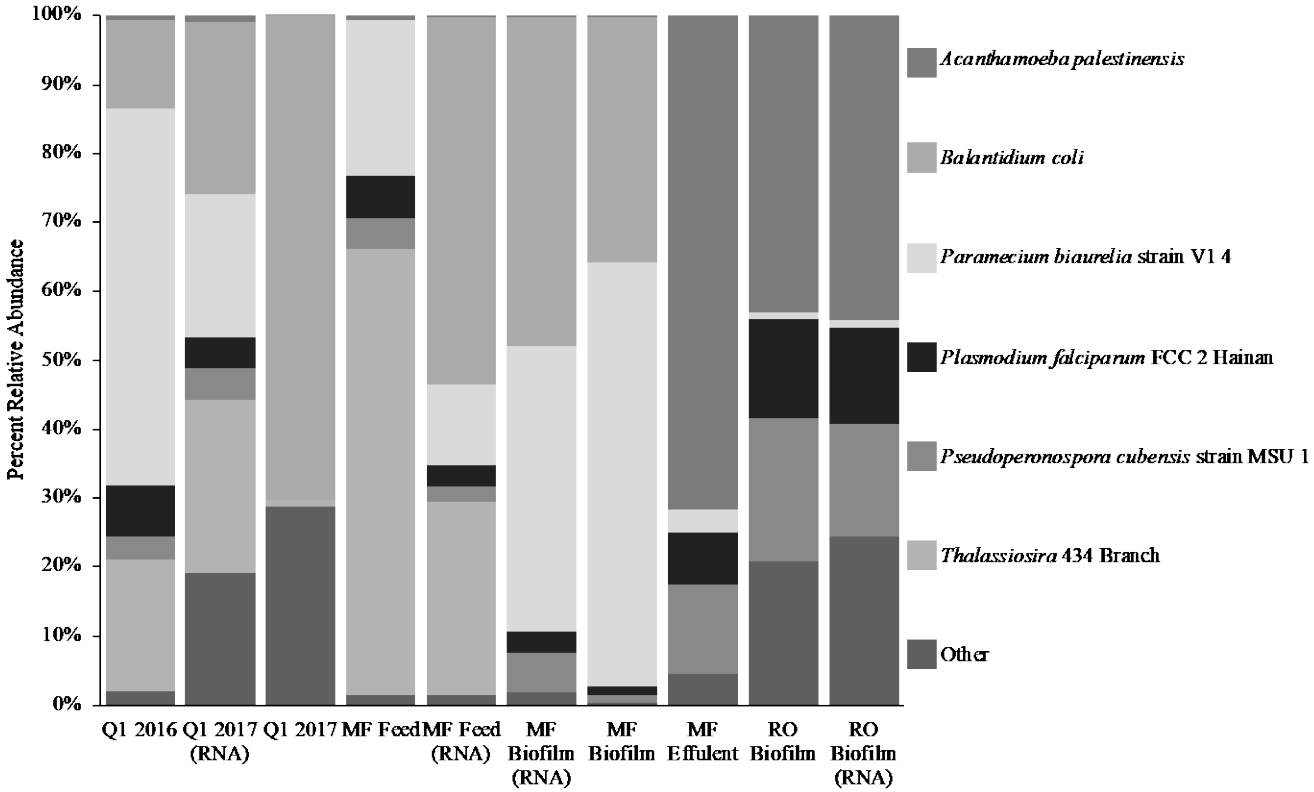


Figure 11.—Relative abundance bar chart of the diversity of the parasite species identified in samples collected across the AWPF. All detected parasite species with an average relative abundance of less than five percent are grouped into "Other."

The bar chart in Figure 12 provides an overview of viral species identified across the AWPF. Pepper mild mottle virus (PMMoV) was identified in both Q1 water and MF biofilm samples. Rhino and Adeno families of viruses (DNA viruses) were the predominant viruses identified in Q1 water and RO biofilm. PMMoV and the Rhino and Adeno family of viruses were not identified in the RO biofilm (Figure 12). While possibly human-associated viruses were found, predominately bacteriophage—and not pathogenic human viruses—were identified as the majority of detected viruses in all of the July 2017 samples.

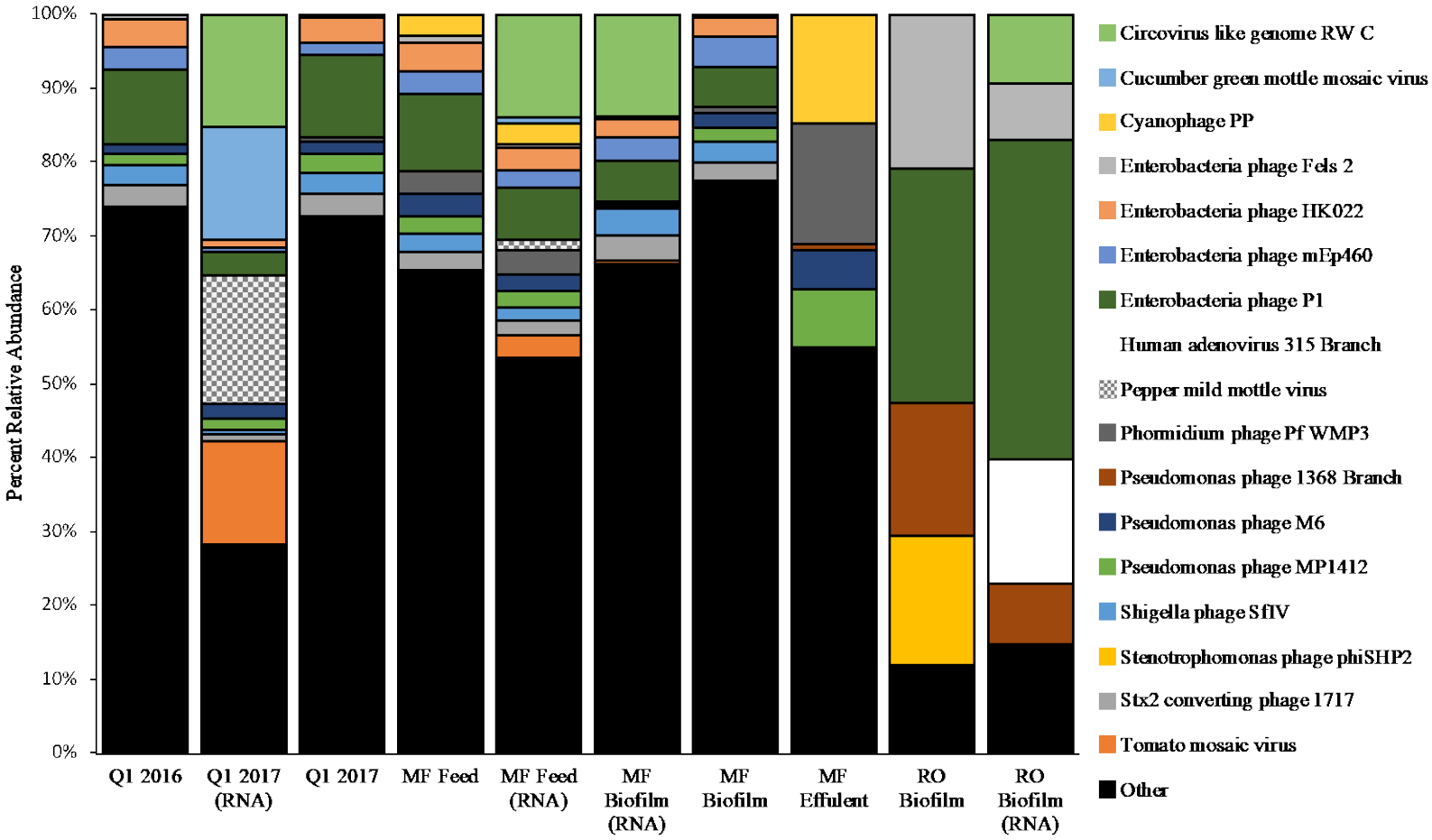


Figure 12.—Relative abundance (percent abundance within each sample, summed to 100 percent) bar chart of viral taxa identified across the AWPF. All viral taxa below an average of 1 percent relative abundance are grouped into "Other."

More ARGs were identified in the Q1 water 2016 sample than in Q1 water 2017 (Figure 8). The same numbers of ARGs were identified in the Q1 water (2016 and 2017) and in MF biofilm. However, fewer ARGs were identified in the RO biofilm. ARG samples clustered similarly to viral and bacterial communities, with Q1 and MFF appearing more similar than other sequenced samples visualized by PCA (Figure 13). Both the water samples clustered closer to each other than with either of the biofilms, suggesting that the ARG DNA sequences identified in the water samples are more similar to each other than the ARGs identified in the biofilms.

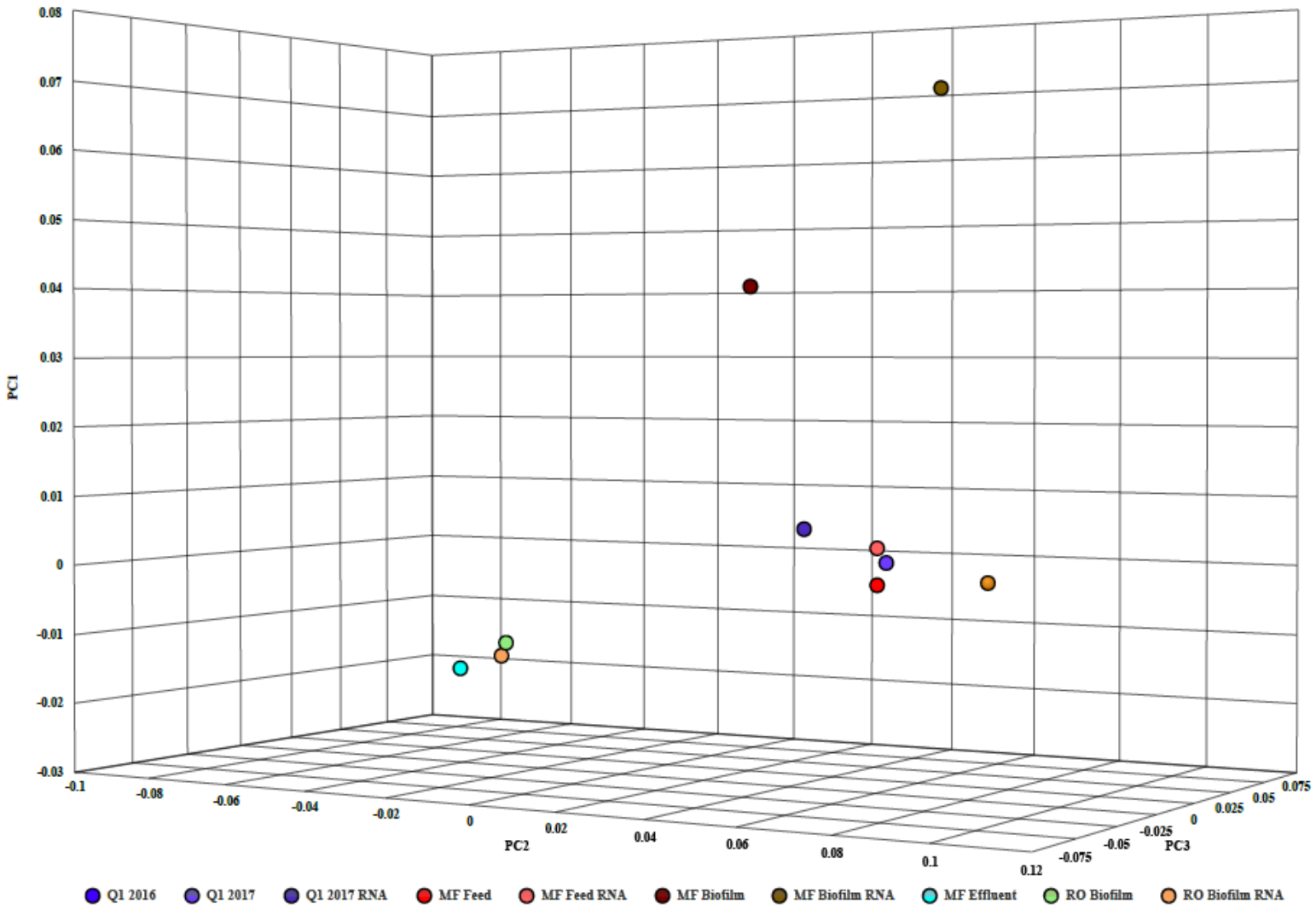


Figure 13. —Principal component analysis (PCA) of DNA and RNA samples taken in October 2016 and July 2017 for ARGs identified using the CosmosID pipeline. Principal component axis (PC1,2,3) values are the raw component scores of the PCA model. More critically, sample distance within the PCA is proportional to how similar a sample is to another.

3.1.7. Estimation of Microbial Cell Number Through qPCR

Quantitative PCR results are summarized in Figure 14. The number of 16S rRNA gene copies identified using this technique are analogous to the number of Bacterial and Archaeal cells present within each sample. Thus, a decline the copy number observed is roughly equivalent to a decline in cell number in each water sample. Gene copy number estimates remained high at Q1 and MFF, decreasing by three orders of magnitude from MFF to MFE. Controls processed in a second qPCR run suggest that any sample below 10^2 copies 16S rRNA genes/mL are below the limit of detection of the assay. Therefore, ROP and UVP samples from July 2017 and UVF, UVP, and FPW from October 2016 are below the limit of quantification. 16S rRNA gene copy number from MFE (October 2016) is near the value detected from filter/extraction blanks taken at the same time and should be interpreted with caution.

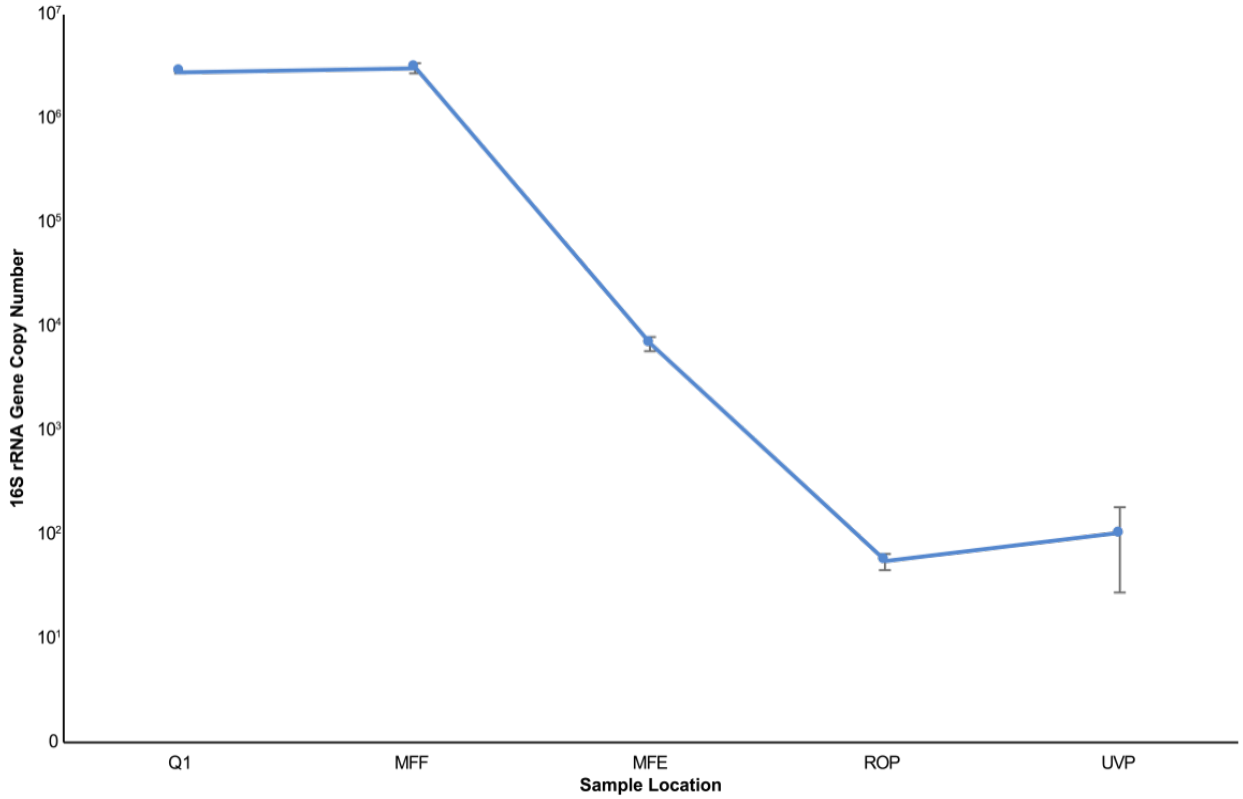


Figure 14. Bacterial/Archaeal Quantitative PCR of July 2017 water samples. Error bars represent average standard deviation of triplicate biological and triplicate technical replicates.

3.2. Discussion

Water is a critical resource in Earth's life support system, yet many societies across the Earth face scarcity due to accelerated population growth, land use, and increased eutrophication of critical water supplies (Madigan et al., 2010). Water reuse (direct or indirect) will only increase, and has attracted a great deal of public attention as we seek to reduce the pressure on ever more critical water resources (Miller, 2006 and EPA and CDM Smith, 2017). A comprehensive understanding of water quality is therefore paramount. Reusing municipal wastewater through advanced water treatment instead of disposal offers many municipalities throughout the United States a chance to augment their water supply. Notably, California has only recently emerged from an extended and severe drought—and more droughts may be on the horizon. As planned water reuse is now practiced more than in any other time in our history, a greater understanding of the operational characteristics of advanced water treatment systems designed to achieve potable reuse is critical to public health (Jimenez and Asano, 2008).

This study was designed to identify the resident microbial community through a combined HTS-based approach, which was successful. Microbial communities were correlated to online water quality and chemistry data, and the microbial load was estimated throughout the AWPF through a quantitative PCR based probe. Briefly, the study was able to achieve the below five objectives, restated from Section 1.3 with summary results:

1) Define the total microbial load of water at points along the AWPF treatment train. *The microbial load at multiple sampling locations was successfully defined using 16S qPCR, showing that much of the detectable biomass is removed by MF and RO.*

2) Identify the resident microbial community in the AWPF using HTS coupled with metagenomics analysis as well as using 16S rRNA gene sequencing and quantitative PCR (qPCR). *The microbial community was successfully identified using a combination of 16S/18S rRNA gene sequencing and HTS metagenomics.*

3) Correlate the removal of microbial communities across the AWPF to existing water quality data demonstrating treatment. *Microbial communities declined as did the concentrations of many chemical constituents such as trace metals and TOC. In this manner, chemical removal can be used as a surrogate or indicator of microbial biomass removal. No correlation to pH or temperature was observed.*

4) Assess the removal efficiency of microorganisms at various points in treatment. *The multi-barrier treatment system of the AWPF was highly successful in a 3-order of magnitude removal of the bulk of detectable biomass as identified using qPCR.*

5) Develop sampling techniques for microbial analysis of high-purity water and potentially better indicators for systems similar to the GWRS AWPF. *The project team successfully tested a method allowing for the filtration of hundreds of liters of water in a short period of time, allowing for rapid sampling across multiple sampling locations. Additional work across a longer timescale is required to identify promising new indicators.*

As expected, the greatest shift in microbial community distribution and abundance came at microfiltration. The MF treatment system is designed to remove microorganisms (including Bacteria and Protozoa) along with suspended and colloidal solids through a polypropylene hollow-fiber membrane with a nominal pore size of 0.2 µm. The biofilm collected from the MF membrane was less diverse than the influent water (Q1), likely due to a combination of adding sodium hypochlorite (which results in a MFF residual chloramine at approximately 3 to 5 mg/L as total chlorine), a known microbial disinfectant, and that fewer communities are adept at existing in the environment of a biofilm. Furthermore, most microbial cells are greater than the nominal pore size of the membrane system (0.2 µm) and thus were filtered out by the MF (Madigan et al., 2010). A small percentage of the population is nevertheless retained downstream in the MF effluent water (MFE), likely owing to a combination of a) incomplete (not 100%) removal given that industrial scale MF is not an absolute, perfect barrier, as evidenced by the need for periodic fiber pinning in the facility based on daily pressure testing, and b) the expected non-sterile conditions after MFE. Abundant operational taxonomic units (OTU) most closely related to *Flavobacterium,* as well as Betaproteobacteria most closely related to *Zooglea,* and unclassified Comamonadaceae were found in both Q1/MFF, and beyond RO and UVP, although they were lacking at MFE. Betaproteobacteria are common in wastewater treatment systems and are responsible for the bulk of organic carbon removal (Cydzik-Kwiatkowska and Zielińska, 2016) so their presence is expected. The microbial community at MFE was unique from the other sampled points across the AWPF, as it was comprised of:

- The uncultivated Alphaproteobacterial lineage DSSF69, which has been found in another drinking water treatment pilot-scale system (Williams et al., 2004);

- *Dongia,* which has been found in water treatment systems previously (Liu et al., 2010);

- *Mycobacterium,* which are common across both drinking water treatment and distribution systems (Le Dantec et al., 2002).

Further investigation of the microbial community present throughout the AWPF using total DNA and RNA HTS allowed for the identification of any potential pathogens that present a potential health risk if uncontrolled. Water at Q1 had the greatest diversity and number of bacterial species identified by the CosmosID bioinformatic pipeline, as well as the greatest biomass estimated by qPCR, both of which declined through the filtration process. No sequence was obtained beyond MFE due to the very low biomass, and therefore very low quantities of DNA. Compared to the hundreds or thousands of bacterial species identified, very few fungi or protest species (dozens) were identified at any point in treatment despite the large amount of DNA and RNA sequence obtained suggesting that they represent very little of the total microbial community. Two identified fungi, *Enterocytozoon* and Onygenales are potential human pathogens, infecting immunocompromised individuals (Sulaiman et al., 2014). Despite the identification of these two fungi, overall the fungi represented less than 1 percent of the sequence reads identified in influent (Q1) water, and were otherwise plant pathogens such as *Puccinia* and *Lentinus* (Fangkrathok et al., 2014 and Gill et al., 2015).

Protists were in similarly low abundance, representing less than 1 percent of the identified sequence data by CosmosID. *Paramecium aurelia*, a common environmental non-pathogenic protist (Siegel, 1958), was the most abundant overall. The pathogen *Plasmodium falciparum* (Encyclopedia of Parasitology, 2008) was detected, again representing only a fraction of a percent of the total sequence identified. These protists are commonly found in wastewater, and more work is required to confirm their activity beyond DNA or RNA sequencing.

The AWPF appears to be highly effective in the removal of ARGs from the environment. Compared to all other samples analyzed by HTS, MFE had the fewest identifiable known ARGs, and the fewest Bacterial species or detectable viral sequence. Transfer of antibiotic resistance from the environment to human pathogens is of concern, and any treatment method should strive to remove this risk as a precautionary approach (Boucher et al., 2009 and Pei et al., 2012). Previous research has shown the effectiveness of membrane based systems, like RO at the AWPF, in removing antibiotic resistance markers from the environment (Pruden et al., 2013). DNA quantities were too low to successfully sequence beyond MFE from water filtered during this study using metagenomic or metatranscriptomic HTS necessary to identify ARGs (Table 2). Biofilms at MF and RO however, were successful in HTS and ARGs were identified.

Removal of viral particulate is a key objective of advanced treatment for potable reuse, specifically through the RO filtration and UV/AOP processes (Figure 2) is another key point in the AWPF. Much of the detected viral sequences were bacteriophages and are harmless to humans (Figure 12). Again, Q1 had the greatest numbers of detected viral sequence compared to MFE and MF and RO biofilms. There was minor variation in the 2016 and 2017 samples. Despite filtering significantly less water, the 2016 Q1 sample had a greater number of

detected viral sequences. This could reflect temporal variations in the influent (Q1) microbial and viral load and warrants additional sampling and investigation to establish a baseline average viral load.

A key feature of the CosmosID pipeline is the ability to accurately predict many of the potentially pathogenic microorganisms, in addition to other markers such as viruses and ARGs. Approximately 33% of the bacteria in water at Q1 (2016 and 2017) were identified above the CosmosID confidence threshold of 95% at the species level. Similarly, approximately 36% of bacterial species in the RO biofilm were identified above the confidence threshold. Bacteria identified below the confidence threshold need further validation to determine if the predicted species are actually present in the sample.

If the microbial community identified by rRNA gene and metagenomic HTS are taken in isolation, it would appear as though the microbial community recovered after microfiltration and to some degree after RO, approximates the microbial community found at Q1 again in population distribution and relative abundance. However, a key finding is the vast reduction in total biomass after MF and RO. Abundance detected by rRNA gene sequencing is relative- that is to say that the percentages identified in figures relate to only the total number of organisms detected within that sample and does not indicate the total number of microbial cells (microbial density) within that sample. Microbial density, rather than differences in community relative abundance, has been found to be a critical component in disease status in the human microbiome (Vandeputte et al., 2017), and a similar correlation is likely in water purification and distribution systems. Therefore, microbial density is a critical component of ecology that cannot be ignored. Thus, similarities in microbial community profiles should not be misinterpreted as incomplete water treatment. Adding sodium hypochlorite at MFF appears to have no immediate impact on the density of the microbial population observed. This is more than likely due to the short exposure time of microorganisms to sodium hypochlorite at this sampling point and likely influences the viable microbial population further downstream at MF, and RO, given that it is used for membrane biofilm control in the AWPF. Chlorination is a well-proven method to disinfect drinking water (Aieta and Berg, 1986) although contact time is critical for disinfection.

MF was highly effective in removing a large percentage of the population, presenting a physical, rather than chemical barrier to any microorganism present. Most microorganisms are larger than the nominal pore size (0.22 µm) of the membrane system (Madigan et al., 2010), although this is not a perfect barrier and a small number of microorganisms may pass through the barrier and further into the AWPF. Further reductions in biomass occurred via RO, and again, a high percentage of the population was removed despite the similarity of the microorganisms detected by rRNA gene sequencing (Figure 6). Not only did the RO membrane further deplete the microbial population, but it served as an effective barrier to transmissible genomic elements, such as virulence factors, or

antibiotic resistance that may otherwise be transmitted to other organisms (Pruden et al., 2013). A combined multi-barrier filtration system coupled to UV/AOP treatment, therefore, represents an effective means of removing the microbial population present in secondary treated wastewater, in this way removing the wastewater "identity" or signature of the wastewater source.

There is no requirement or objective to entirely remove or inactivate (sterilize) the finished water. In fact, some current research in potable reuse, particularly for direct reuse (no environmental buffer) has suggested the potential importance of establishing healthy microbial consortia in finished water. As described previously, in the case of OCWD AWPF, the finished water is injected and percolated into the regional aquifer (groundwater basin) where it comingles with groundwater and is later withdrawn via production wells (after underground storage /travel times to wells on the order of months or years) by local cities and water agencies as a drinking water source. These agencies then provide additional treatment required for groundwater supplies, such as disinfection. A potentially interesting area for future study, related to the highly effective removal of the microbial community by the advanced treatment, is implications of the injection of high-purity, finished water, and the associated microbial community (or lack thereof) into a subsurface aquifer. Cell numbers in the subsurface average near 10^3 to 10^4 cells per mL (Colman et al., 2017), far lower than the concentration of cells in secondary treated wastewater that is the influent into the AWPF ($\approx 10^6$ cells per mL) yet likely much higher than the finished, highly purified water from the AWPF. Thus, the water injected into the subsurface may actually dilute what little microbial community exists. What—if any—impact this might have on the subsurface microbial community should be a focus of future work, which should continue to consider the active microbial populations.

4. Conclusions

4.1. Conclusions

This study successfully achieved the proposed goals of identifying the resident AWPF microbial communities and also provided a better understanding of the microbial load from inflow (secondary treated wastewater) to the final product (purified water). HTS based metagenomics have been used to study environmental systems, human health topics in medicine, and in other applications; in this study, to the authors' knowledge, HTS was applied to recycled water for the first time. Total microbial community abundance was quantified by using qPCR of the small subunit ribosomal RNA gene. This, coupled with rapid metagenomic classification provided by CosmosID, allows for a holistic view of the microbial communities within the advanced water purification facility. HTS using the CosmosID bioinformatic pipeline of DNA and RNA identified RNA viruses and accurately identified microorganisms that are present and likely metabolically active.

An additional contribution of the presented work is the development of improved sampling techniques for systems similar to the OCWD AWPF. The Innova Prep large volume concentration system used in this study was highly effective to concentrate large volumes of water (100 L) to a more manageable 50 to 100 mL for filtration and downstream processing. Filtration occurred in less than an hour at each site, rather than half a day. This process allowed for greater productivity and/or the ability to sample more locations in a single day. Due to extremely low biomass detected beyond reverse osmosis (RO) treatment, high volumes (> 100 L) and multiple lab-based controls of DNA extraction and PCR reagents as described in this study's methods are needed to identify microbial communities and DNA recovery rates accurately in AWPF treatment processes. Future work at any AWPF type treatment system that produces ultra-pure water should utilize similar dialysis-style filters to allow for rapid, high volume filtration.

4.2. Recommended Next Steps

This study showed that the AWPF is effective in removing the majority of the detected biomass through MF and RO. Following subsequent UV/AOP treatment, there is very little detectable biomass in the product water, despite the similarity of microbial communities after these steps identified in 16S rRNA gene sequencing, meaning that there are far fewer microbial cells present. This study illustrates the need to describe and identify microbial community composition and to quantify microbial load. We *strongly suggest* that any ongoing monitoring at the AWPF or other similar sites using community composition analyses include qPCR estimation of the microbial load for accurate assessment of microbial

communities. Such a low biomass environment as the ultra-pure water produced from the AWPF that is injected into the subsurface should be the focus of future work. Sampling of the aquifer that the GWRS AWPF injects its water into should be carried out, and the microbial community compared to that of surrounding aquifer systems to determine if such ultra-pure water produced by the AWPF impacts existing microbial communities in the subsurface.

The presented work also shows the critical nature of controls and quantification in the analysis of ultra-low biomass environments, such as the AWPF. DNA was non-quantifiable, or very low in samples after RO treatment. Furthermore, microorganisms (or their DNA) present in DNA extraction or PCR reagents (Salter et al., 2014) used in laboratories (i.e., ultra-pure waters) are very likely similar to those in the ultra-pure water produced by the AWPF treatment train. Any future work should continue to sequence and quantify filtration, extraction, and PCR blanks to ensure that the detected microbial communities are actually those found in the water and are not from lab reagents.

In this study, HTS could not be performed on AWPF waters after MF effluent due to insufficient DNA obtained, consistent with the higher amounts of DNA required for HTS compared to qPCR, or 16S rRNA gene sequencing based methods. Increasing filtration volumes beyond 100 L (as taken in the second sampling event) using the Innova Prep filters may enable HTS by increasing the limit of detection and mitigate issues of low DNA concentration in future work.

Metagenomic analysis via HTS is necessary for providing strain-level identification of microorganisms and viruses present in the AWPF and similar systems which is not possible with 16S rRNA sequencing methods via qPCR (which were also used in this study). Strain-level identification can identify potential pathogens, or genetic elements of interest such as antibiotic resistance that may persist through treatment. The identification of large numbers of viruses and ARGs in influent water warrants additional study. Additional sampling of Q1 over multiple months and years could establish a baseline viral and ARGs load, providing critical information for plant management and the water reuse industry, particularly as communities implement direct reuse (raw and treated water augmentation with advanced treated recycled water).

Further, with respect to assessing organism viability, coupling the CosmosID approach used in this study (HTS/metagenomics with bioinformatics) to identification of RNA, in addition to DNA, will help in the future to identify organisms that are not only present, but likely metabolically active.

References

Aieta, E. M. and Berg, J. D. (1986). A Review of Chlorine Dioxide in Drinking-Water Treatment. *J Am Water Works Assoc* 78, 62–72.

Alneberg, J., Bjarnason, B. S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U. Z., et al. (2014). Binning metagenomic contigs by coverage and composition. *Nat Meth* 11, 1144–1146. doi:10.1038/nmeth.3103.

Boucher, H. W., Talbot, G. H., Bradley, J. S., Edwards, J. E., Gilbert, D., Rice, L. B., et al. (2009). Bad Bugs, No Drugs: No ESKAPE! An Update from the Infectious Diseases Society of America. *Clin Infect Dis* 48, 1–12. doi:10.1086/595011.

Caporaso, J. G., Bittinger, K., Bushman, F. D., DeSantis, T. Z., Andersen, G. L., and Knight, R. (2010a). PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* 26, 266–267. doi:10.1093/bioinformatics/btp636.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010b). QIIME allows analysis of high-throughput community sequencing data. *Nat Meth* 7, 335–336. doi:10.1038/nmeth.f.303.

Chao, A. (1984). Nonparametric estimation of the number of classes in a population. *Scandanavian Journal of Statistics* 11, 265–270.

Colman, D. R., Poudel, S., Stamps, B. W., Boyd, E. S., and Spear, J. R. (2017). The deep, hot biosphere: Twenty-five years of retrospection. *Proc Natl Acad Sci U S A* 114, 6895–6903. doi:10.1073/pnas.1701266114.

Cydzik-Kwiatkowska, A. and Zielińska, M. (2016). Bacterial communities in full-scale wastewater treatment systems. *World J Microbiol Biotechnol* 32, 449. doi:10.1007/s11274-016-2012-9.

DOE. (2018). BBtools, https://jgi.doe.gov/data-and-tools/bbtools/. Accessed January, 2018.

Edgar, R. C. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Meth* 10, 996–998. doi:10.1038/nmeth.2604.

Encyclopedia of Parasitology. (2008). *Springer-Verlag Berlin Heidelberg*.

EPA and CDM Smith. (2017). 2017 Potable Reuse Compendium. Prepared by the Environmental Protection Agency (United States) and CDM Smith, Washington D.C.

Eren, A. M., Esen, Ö. C., Quince, C., Vineis, J. H., Morrison, H. G., Sogin, M. L., et al. (2015). Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* 3, e1319. doi:10.7717/peerj.1319.

Fangkrathok, N., Junlatat, J., Umehara, K., Noguchi, H., and Sripanidkulchai, B. (2014). Cytotoxic and immunomodulatory effects of polyhydroxyoctane isolated from Lentinus polychrous mycelia. *J Nat Med* 68, 302–309. doi:10.1007/s11418-013-0797-5.

Gill, U. S., Uppalapati, S. R., Nakashima, J., and Mysore, K. S. (2015). Characterization of Brachypodium distachyon as a nonhost model against switchgrass rust pathogen Puccinia emaculata. *BMC Plant Biol* 15, 113. doi:10.1186/s12870-015-0502-9.

Hasan, N. A., Young, B. A., Minard-Smith, A. T., Saeed, K., Li, H., Heizer, E. M., et al. (2014). Microbial Community Profiling of Human Saliva Using Shotgun Metagenomic Sequencing. *PLoS One* 9, e97699. doi:10.1371/journal.pone.0097699.

Hughes, B., Beale, D. J., Dennis, P. G., Cook, S., and Ahmed, W. (2017). Cross-Comparison of Human Wastewater-Associated Molecular Markers in Relation to Fecal Indicator Bacteria and Enteric Viruses in Recreational Beach Waters. *Appl Environ Microbiol* 83, e00028–17. doi:10.1128/AEM.00028-17.

Jimenez, B. and Asano, T. (2008). Water Reuse: An International Survey of current practice, issues and needs. *Water Intelligence Online* 7, 9781780401881–9781780401881. doi:10.2166/9781780401881.

Jolliffe, I. T. and Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philos Trans A Math Phys Eng Sci* 374. doi:10.1098/rsta.2015.0202.

Langmead, B. and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Meth* 9, 357–359. doi:10.1038/nmeth.1923.

Lax, S., Smith, D. P., Hampton-Marcell, J., Owens, S. M., Handley, K. M., and Scott, N. M., et al. (2014). Longitudinal analysis of microbial interaction between humans and the indoor environment. *Science* 345, 1048–1052. doi:10.1126/science.1254529.

Le Dantec, C., Duguet, J.-P., Montiel, A., Dumoutier, N., Dubrou, S., and Vincent, V. (2002). Occurrence of mycobacteria in water treatment lines and in water distribution systems. *Appl Environ Microbiol* 68, 5318–5325. doi:10.1128/AEM.68.11.5318-5325.2002.

Li, D., Liu, C. M., Luo, R., Sadakane, K., and Lam, T. W. (2015a). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics

assembly via succinct de Bruijn graph. *Bioinformatics* 31, 1674–1676. doi:10.1093/bioinformatics/btv033.

Li, Y.-L., Weng, J.-C., Hsiao, C.-C., Chou, M.-T., Tseng, C.-W., and Hung, J.-H. (2015b). PEAT: an intelligent and efficient paired-end sequencing adapter trimming algorithm. *BMC Bioinformatics* 16 Suppl 1, S2. doi:10.1186/1471-2105-16-S1-S2.

Liu, C. M., Aziz, M., Kachur, S., Hsueh, P.-R., Huang, Y.-T., Keim, P., et al. (2012). BactQuant: an enhanced broad-coverage bacterial quantitative real-time PCR assay. *BMC Microbiol* 12, 56. doi:10.1186/1471-2180-12-56.

Liu, Y., Jin, J. H., Liu, Y. H., Zhou, Y. G., and Liu, Z. P. (2010). Dongia mobilis gen. nov., sp. nov., a new member of the family Rhodospirillaceae isolated from a sequencing batch reactor for treatment of malachite green effluent. *Int J Syst Evol Microbiol* 60, 2780–2785. doi:10.1099/ijs.0.020347-0.

Madigan, M. T., Clark, D. P., Stahl, D., and Martinko, J. M. (2010). *Brock Biology of Microorganisms 13th edition*.

McDonald, D., Clemente, J. C., Kuczynski, J., Rideout, J. R., Stombaugh, J., Wendel, D., et al. (2012). The Biological Observation Matrix (BIOM) format or: how I learned to stop worrying and love the ome-ome. *GigaScience* 1, 7. doi:10.1186/2047-217X-1-7.

McMurdie, P. J. and Holmes, S. (2013). phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One* 8, e61217. doi:10.1371/journal.pone.0061217.

Miller, W.G. (2006). Integrated concepts in water reuse: managing global water needs. *Desalination* 187, 65–75. doi:10.1016/j.desal.2005.04.068.

OCWD. (2018). GWRS annual reports, https://www.ocwd.com/gwrs/annual-reports/. Accessed February, 2018.

Ottesen, A., Ramachandran, P., Reed, E., White, J. R., Hasan, N., Subramanian, P., et al. (2016). Enrichment dynamics of Listeria monocytogenes and the associated microbiome from naturally contaminated ice cream linked to a listeriosis outbreak. *BMC Microbiol.* 16, 275. doi:10.1186/s12866-016-0894-1.

Parada, A., Needham, D. M., and Fuhrman, J. A. (2015). Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time-series and global field samples. *Environ Microbiol* doi:10.1111/1462-2920.13023.

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from

isolates, single cells, and metagenomes. *Genome Res* 25, 1043–1055. doi:10.1101/gr.186072.114.

Pei, R., Storteboom, H., and Pruden, A. (2012). Antibiotic Resistance Genes (ARG) as Emerging Environmental Contaminants. in (Reston, VA: American Society of Civil Engineers), 1–10. doi:10.1061/40856(200)97.

Ponnusamy, D., Kozlova, E. V., Sha, J., Erova, T. E., Azar, S. R., Fitts, E. C., et al. (2016). Cross-talk among flesh-eating Aeromonas hydrophila strains in mixed infection leading to necrotizing fasciitis. *Proc Natl Acad Sci U S A* 113, 722–727. doi:10.1073/pnas.1523817113.

Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS One* 5, e9490. doi:10.1371/journal.pone.0009490.

Pruden, A., Larsson, D. G. J., Amézquita, A., Collignon, P., Brandt, K. K., Graham, D. W., et al. (2013). Management Options for Reducing the Release of Antibiotics and Antibiotic Resistance Genes to the Environment. *Environ Health Perspect* 121, 878–885. doi:10.1289/ehp.1206446.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J., et al. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* 35, 7188–7196. doi:10.1093/nar/gkm864.

Rosario, K., Symonds, E. M., Sinigalliano, C., Stewart, J., and Breitbart, M. (2009). Pepper Mild Mottle Virus as an Indicator of Fecal Pollution. *Appl Environ Microbiol* 75, 7261–7267. doi:10.1128/AEM.00410-09.

Salter, S., Cox, M. J., Turek, E. M., Calus, S. T., and Cookson, W. O. (2014). Reagent contamination can critically impact sequence-based microbiome analyses. *bioRxiv*.

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., et al. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 75, 7537–7541. doi:10.1128/AEM.01541-09.

Siegel, R. W. (1958). Hybrid Vigor, Heterosis, and Evolution in Paramecium aurelia. *Evolution* 12, 402. doi:10.2307/2405861.

Staley, J. T. and Konopka, A. (1985). Measurement of in Situ Activities of Nonphotosynthetic Microorganisms in Aquatic and Terrestrial Habitats. *Annu Rev Microbiol* 39, 321–346. doi:10.1146/annurev.mi.39.100185.001541.

Sulaiman, I. M., Jacobs, E., Simpson, S., and Kerdahi, K. (2014). Genetic Characterization of Fungi Isolated from the Environmental Swabs collected from a Compounding Center Known to Cause Multistate Meningitis Outbreak in United States Using ITS Sequencing. *Pathogens* 3, 732–742. doi:10.3390/pathogens3030732.

Vandeputte, D., Kathagen, G., D'hoe, K., Vieira-Silva, S., Valles-Colomer, M., Sabino, J., et al. (2017). Quantitative microbiome profiling links gut community variation to microbial load. *Nature* 352, 565. doi:10.1038/nature24460.

Williams, M. M., Domingo, J. W. S., Meckes, M. C., Kelty, C. A., and Rochon, H. S. (2004). Phylogenetic diversity of drinking water bacteria in a distribution system simulator. *J Appl Microbiol* 96, 954–964. doi:10.1111/j.1365-2672.2004.02229.x.

Metric Conversions

Unit	Metric equivalent
1 gallon	3.785 liters
1 gallon per minute	3.785 liters per minute
1 inch	2.54 centimeters
1 million gallons per day	3,785 cubic meters per day
1 pound per square inch	6.895 kilopascals
1 square foot	0.093 square meters
°F (temperature measurement)	(°F–32) × 0.556 = °C
1 °F (temperature change or difference)	0.556 °C