APPLICATION OF ADVANCED MOLECULAR TECHNIQUES IN APPLIED ENVIRONMENTAL MICROBIOLOGY

By

Brandon Charles Iker

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF SOIL, WATER, AND ENVIRONMENTAL SCIENCE

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College of

THE UNIVERSITY OF ARIZONA

2013

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by Brandon Charles Iker entitled "APPLICATION OF ADVANCED MOLECULAR TECHNIQUES IN APPLIED ENVIRONMENTAL MICROBIOLOGY" and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

_____Date:
07/24/13
Ian L. Pepper, Ph.D.

_____Date:
07/24/13
Charles P. Gerba, Ph.D.

_____Date:
07/24/13
Channah Rock, Ph.D.

_____Date:
07/24/13
Kelly Bright, Ph.D.

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

_____Date:
07/24/13
Dissertation Director: Ian L. Pepper, Ph.D.

STATEMENT BY AUTHOR

This statement has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University of Arizona Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College, when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: Brandon Charles Iker

ACKNOWLEDMENTS

Dr. Pepper, thank you for all the mentoring moments we have had together over the past few years. You have had a lot of patience with me and I thank you for all the challenges that you have given me. These experiences have made me even more determined and skilled then I could ever imagine.

Dr. Gerba, I owe you a lot for the opportunities that you have provided me. I could always count on you for a funny joke or hilarious Dr. Pepper story, something that I will take with me for the rest of my life. You have been truly inspiring with your vast knowledge of our pathogenic microbial world, reminding me that there still is a lot to learn. We students stand on the shoulders of giants. Thank you for all that you do.

Dr. Bright, your calm demeanor and supreme editorial skills have kept me a float for the past four years. To have the honor to have a chat with you, hear your opinion about one of my, at times, far out research ideas, or to obtain guidance on how to navigate the maze that is our version of academia, will always be among my most memorable experiences.

Dr. Rock, thank you so much for your guidance over the past few years. I am always so excited to talk or meet with you; you bring clarity, and always have excellent criticisms of my work, as well as a working knowledge of our field, something I value dearly. I hope to continue to have opportunities to work with you in the future!

DEDICATION

This dissertation is dedicated to those less fortunate, to those that went through today without a clean glass of water or safe healthy food. External factors like extreme weather, political, and religious unrest continue to drive our global food and water security into further crises. We must begin to apply even more advanced thinking to mitigate the effects of these ever bearing issues.

TABLE OF CONTENTS

TABLE OF CONTENTS - continued

TABLE OF CONTENTS - continued

TABLE OF CONTENTS - continued

ABSTRACT

Recent advancements in molecular biology such as next generation sequencing and more sensitive and rapid molecular detection methods like qPCR, have historically been developed for clinical applications in human genetics and for health care diagnostic purposes. The high demand for faster and more accurate molecular assays in the health care field has driven rapid development of inexpensive molecular techniques that when applied to the science of environmental microbiology, provides an unprecedented level of understanding of the microbial world around us. The goal of this dissertation is to begin to apply more advanced molecular technologies to problems in applied environmental microbiology. Appendix A is a brief literature review of next generation sequencing technologies for applications in environmental microbiology. Appendix B focuses on the development of a more robust virus nucleic extraction kit for the detection of viral genomes from environmental samples found to contain high concentrations of qPCR inhibitors, such as humic acids or heavy metals. Appendix C summarizes one of the largest virus surveys done in the US, using state of the art qPCR technologies in both wastewater influent and effluent from two wastewater treatment plants in the Southwest. Data suggests that traditional virus indicators may not be a viable tool to evaluate fecally impacted source water or virus removal during water treatment. The third study summarized in Appendix D, provides one of the first insights into the microbial ecology of biofilms utilized as biological treatment media using Roche 454 amplicon sequencing of the 16S rRNA gene.

INTRODUCTION

Problem Definition

I. Improving the extraction of viral nucleic acids in environmental virus concentrates

Appendix B reports on a new extraction procedure for environmental viruses. The detection of viruses in environmental matrices has relied on the use of cell culture or traditional gel based PCR for the presence or absence of viable or potentially viable viral genomes. However, there are major limitations for both of the techniques. Virus cell culture, while providing reasonable quantitative data is also limited in that few viruses are readily culturable, and factors such as assay cost and incubation times are greatly prohibitive. Polymerase chain reaction (PCR) provides some advantages over traditional cell culture technique in that it is typically more sensitive, and can be easily adapted to detect most any virus genome. However, PCR does not necessarily indicate viable viruses, and is found to be prone to inhibition that increases false negative results. PCR is also comparatively costly and time consuming compared to that of end point PCR like quantitative (qPCR). Developing techniques that mitigate false negatives, lower cost, improve detection limits, and speed up initial detection of viral genomes provides researchers with more useful tools when surveying the environment for potential viral contamination.

Environmental virus concentration and extraction methods commonly cite the use of a commercially available virus extraction kit to extract viral nucleic acids in favor of in-house extraction methods that lack reproducibility and quality assurance from lab to lab. Many extraction kits currently used for viral nucleic acid isolation were originally developed for use in a clinical setting according to manufacturers'

recommendations and provide excellent virus recovery for samples low in inhibitory compounds. An ideal environmental virus extraction kit should provide similar recoveries to established extraction methods and kits, while simultaneously providing effective inhibitor removal to provide the cleanest, highest quality nucleic acid extracts for subsequent molecular detection and quantification.

II. Simplifying the discussion on molecular detection of viruses

The accurate detection and quantification of viruses in the environment is inherently difficult for many reasons (Pang et al. 2012; Ikner et al. 2012):

1. Viruses in the environment are typically present in low numbers.
2. Concentrating virus numbers by sample processing also co-concentrates PCR inhibitors.
3. The recovery efficiency of specific viruses is highly variable due to differences in viral capsid physiology, structure and other environmental factors.
4. The development of efficient, secondary concentration methodologies is still lagging.
5. Poor optimization of molecular detection assays such as qPCR results in low sensitivity of detection of most viruses.

There are numerous methodologies and approaches to detect pathogenic enteric viruses (Ikner et al., 2012). One of the more popular and effective concentration protocols is the virus absorption elution method (VIRADEL), which includes: 1MDS with organic flocculation (Epa & Exposure, 2010; Ikner et al., 2011); NanoCeram with organic flocculation (Epa 2010; Ikner et al. 2011); HA negatively charged

membrane method with ultracentrifugation (Katayama et al., 2002); and Glass fiber, powder, or beads methodologies (Wyn-Jones et al. 2011). Other less common methods are tangential flow (Gibson and Schwab, 2011), ultrafiltration (Rhodes et al., 2011), and centrifugation (Prata et al., 2012). These methodologies are commonly cited in the literature, but are typically less practical in terms of cost and field application, and less efficient in terms of concentration and contamination removal.

In addition to the multiple techniques that make the comparison of different studies extremely difficult, other factors include, poor reporting of actual methodological details, particularly with respect to molecular assays and limit the discussion and understanding of virus monitoring from one study to another. One of the objectives of this review is to provide a better understanding of what can be called “tested water equivalency" (TWE). The tested water equivalency represents the amount of the original water sample that is actually assayed for analysis. This is determined by the careful tracking of water volume through the numerous steps that are necessary for molecular detection of pathogenic viruses. Recent improvements in nucleic acid extraction technologies, including more highly concentrated and inhibitor resistant reverse transcription master mixes, and more robust end point PCR reagents, has resulted in improved sensitivity and accuracy in today’s qPCR instrumentation. This in turn has allowed for molecular detection protocols to be more easily optimized, and the establishment of better standard operating procedures (SOPs). A summary for determining the TWE is outlined below in Table 1. This table also provides the user tool to input information about how the assay was conducted and then, by providing the estimated genome copy numbers observed in the qPCR tube,

one can easily calculate accurate estimations of how many viral genomes may be present in the original sample.

Table 1 Custom virus concentration calculator set to optimal reaction conditions

Methodology		Virus detection step	Sample volumes	Unit	Estimated genome copies per volume tested	Estimated % of virus loss during step	Water equivalency (L)	Water concentration efficiency	[a]Estimated genomic copies based on qPCR data with loss		[b]Estimated genomic copies based on qPCR data without loss	
									(L)	(mL)	(L)	(mL)
Virus concentration	Primary	Primary concentration volume	25	L	293	50%	25.00		12	0.01	4	0.00
		Primary eluate volume	0.03	L	195				6,500	6.50	3,611	3.61
	Secondary	Secondary concentration volume	0.03	L	195	20%	25.00	100%	6,500	6.50	3,611	3.61
		Secondary eluate volume	0.65	mL	163				250	0.25	167	0.17
Molecular detection	Extraction	Nucleic acid extraction volume	400	ul	100	25%	15.38	62%				
		Nucleic acid elution volume	50	ul	80							
	reverse transcription	RT-RNA template	15	ul	24	20%	4.62	30%				
		RT-reaction volume	20	ul	20							
	qPCR*	qPCR template volume	10	ul	10	0%	2.31	50%				
							Total=	9%				

***Input average genomic copy number from qPCR replicates into the green cell.**

[a,b]**Viruses cencntration of sample based on lowest possible detection limit a) calculates estimated percent loss in backcalculation b) assumes no loss of virus from step to step**

Table 1 outlines the optimized virus concentration conditions and provides automated calculation of virus quantification using Microsoft Excel

i. Molecular detection of viruses using qPCR

In order for qPCR assays to be truly quantitative, careful accounting of water equivalencies must be made in each of the molecular detection steps as well as the virus concentration steps, in order to provide an accurate estimate of how much water was evaluated. Typically the biggest limitations to efficient virus concentration lies in how much of the final virus concentrate can physically be utilized in nucleic acid extraction protocols (Abbaszadegan et al. 1999; Wong et al. 2012). This problem stems from two components, the first being that the virus concentration methodology chosen has not yielded a high enough concentration factor of either the primary or secondary concentration methods. The second involves the limitation of the amount of concentrated sample that can be put into an extraction kit (Table 1). Typically this limits the extractable volume up to but no more than 1 ml, and more commonly 400-200 μl for most commercially available kits. For example, if the concentrated virus sample is ≥ 1-10 ml, then the aliquot taken for extraction from the secondary concentrate drastically reduces the amount of water able to be extracted of the original water volume. This suggests that no matter the original sample volume of source water, if concentration methodologies cannot approach volumes close to 1-10 ml in total for final concentration volumes, then there is little hope to assay enough water to represent a risk relevant volume of water approaching what a human might be exposed to.

ii. Universal guideline on the minimal information for virus reporting by qPCR

Given the growing complexity of concentrating and quantifying viral genomes in the environment, future efforts need to be made to improve virus reporting amongst scientific communities, so that studies done by one group using one method and can be more easily compared by a different researcher using a different methodology. Already guidelines exist for the reporting of qPCR data (MIQE) (Bustin et al. 2009; Huggett et al. 2013), or the reporting of microbial community data (MIENS) (Yilmaz et al. 2009). Such minimal information for the reporting of environmental virus detection and quantification experiments requirements should undoubted include those guidelines already in use. To improve communication of virus reporting, guidelines should include a standard reporting format freely available on open access sites to be turned in as a supplemental criterion for manuscript submission. The reporting criteria for each sample should include environmental data on the source water i.e. location, time of year, water temp, pH and turbidity data if available. Methodological details should be listed in detail including the initial water volume sampled, eluate volume, buffer, type and pH. Other physical and chemical factors describing the reaction conditions of secondary and tertiary concentrations should also be listed in a similar manor. Additional molecular details must also be reported including volume of final concentrate used in extractions, how much template was use for cDNA preparation, as well as the ratio of template to reaction buffer, and finally how much DNA or cDNA template was used per qPCR assay.

iii. Surveying viruses in the environment

Reclaimed or recycled water systems derived from treated municipal wastewater can be used for various purposes, such as direct and indirect potable reuse, industrial use, agricultural irrigation, recreational use, and environmental enhancement. If done correctly, such reuse is a safe, sustainable, and feasible strategy to manage limited water resources. For potable water reuse, insufficient removal of viral pathogens and potential public health risks are of major concern because viruses show remarkable persistence during the wastewater treatment process, and are highly infectious to humans. The health risks associated with exposure to reclaimed water are minimized by identifying appropriate advanced treatment technology, and careful monitoring and management of water treatment trains (Toze 2006). Nevertheless, due to high concentrations of some pathogens in wastewater and the possibility of inadequate treatment, viral pathogens may be discharged as effluent from wastewater treatment plants (WWTPs) (Harwood et al. 2005). In addition, the concentration of viruses in treated wastewater may vary according to the type of the wastewater treatment process, season, geographical area, and hygiene conditions within the community.

III. Review of next generation sequencing technologies for applications in environmental microbiology

Next-generation sequencing (NGS) has taken DNA sequencing and detection to the next level of efficiency and speed. Coupled with the advent of environmental genomics and 16S rDNA molecular phylogenetics, next generation sequencing

technologies (NGST) have greatly improved our understanding of the microbial world. Historically, 16S rDNA molecular phylogenetic library construction pioneered many of the studies in the field of microbial ecology (Amann et al. 1992; Amaral-Zettler et al. 2009). More recently, the development of environmental shotgun sequencing, more commonly termed as "metagenomics" has taken the spotlight in sequencing applications (Handelsman et al. 1998). Metagenomics allows sequencing of portions of genetic material from individuals within a whole microbial community by randomly sequencing short (100-1000 bp) nucleic acid fragments found within community DNA.

Applications in environmental microbiology and public health have only recently been realized, due to high costs initially associated with NGST. However, in recent months these costs have dropped dramatically. Traditional techniques have relied heavily on culture or PCR based assays, both of which have been shown to have shortcomings. In addition, even assessment of basic water quality can take >24 hours to achieve using culture based assays. Also, in the case of PCR, assays are typically limited to just one organism or target piece of DNA/RNA, and provide no information about the organism's viability. In contrast, next generation sequencing allows for nearly autonomous, deep, rapid, and unbiased sequencing of microbial communities. Applications of NGS in environmental microbiology open up many opportunities for improvements in public health, and a greater understanding of the effects that human activities have on the environment.

IV. Low cost, low energy biological water treatment

Appendix D summarizes the study and application of a new point of use (POU) water treatment technology that can provide drinking water in water scarce regions with poor source water quality. With increasing concerns over access to safe drinking water, especially in arid undeveloped regions and in the aftermath of natural disasters, the development and greater use of portable, low cost, and efficient water treatment technologies has been, and continues to be, a global need. One approach for treating contaminated drinking water utilizes the development of a biological treatment layer, "schmutzdecke", or biofilm on porous media like that of sand or aragonite for use as a low cost, low energy point of use (POU) water filter. The physical exclusion of larger biological particles such as protozoa, fungal spores, and larger particles of organic matter occurs within a porous substrate is well understood, but the development of a biologically active treatment layer, schmutzdecke, or biofilm, seems to be essential for efficient bacteria removal of pathogens such as *Vibrio* and *Salmonella*. The mechanisms by which biofilms or biological treatment layers capture and remove pathogens to enhance drinking water quality remains unknown (Stauber et al. 2012). Therefore, by studying the ecology of developed biofilms used for water treatment (Rittmann 2006).

There have been few studies that have attempted to systematically characterize microbial communities within these biofilms. Some have utilized traditional techniques like culture-based assays (Hunter et al. 2012). Others have used molecular techniques like DGGE, clone libraries, phylochip and qPCR to quantify and characterize the microbial communities of biological treatments layers (Haig, et al., 2005; Wakelin et al., 2011). To

our knowledge, no studies have utilized deep sequencing analysis to evaluate the microbial communities of biofilms developed for POU water treatment. Developing insights into how biofilms develop and function to facilitate efficient biological water treatment may prove useful in gaining an understanding of how pathogen removal occurs via both filtration and potential inactivation. This understanding may allow use to improve or optimize the biological mechanisms involved in low cost low energy water treatment technologies that are essential to sustaining human health.

DISSERTATION FORMAT

The major focus of dissertation comprises four appendices. Appendix A comprises a short unpublished literature review of next generation sequencing technologies and their application in environmental microbiology. Appendix B contains a primary research article that is already published in the Journal of Virological Methods. It details the comparison of three virus extraction kits for the removal of PCR inhibitors and virus recovery in environmental samples considered to be high in PCR inhibitors. Appendix C summarizes a manuscript that will be submitted to the Journal of Environmental Science and Technology entitled "Relative abundance and treatment reduction of viruses during wastewater treatment processes – identification of potential viral indicators," which describes one of the largest virus surveys in wastewater systems in the US. Data is presented on 9 separate viruses and 2 protozoa of influent and effluent of two wastewater treatment plants in Tucson, Arizona. Appendix D a third and separate dissertation topic, summarizes pyrosequencing data on the microbial ecology of biofilms constructed on foam, for the biological treatment of water, as part of a novel point of use (POU) treatment system. This study will be published, in the journal *International Society of Microbial Ecology*.

PRESENT STUDY

This dissertation contains four appendices. A summary of each appendix follows below.

The manuscript in Appendix A provides a brief literature review of the next generation sequencing technology. While numerous literature reviews exist for this topic, few maintain a focus within the specific application of applied environmental microbiology. The cost and technical skills associated with next generation sequences rapidly decrease every few months as the technology becomes more available and well studied. Still to date, much of the limitation of next generation sequencing technology do not lie with the technical details or in its general application, but in the end users' ability to analyze data with a significant amount of computational power to be useful. Future improvements in sequencing technology will rely greatly on easily utilized and applied to problems within the field of environmental microbiology without the excessive dependence of advanced bioinformatics or high-powered computation needs.

The study conducted in Appendix B evaluated the extraction and purification of nucleic acids from a newly developed kit as a critical step in the molecular detection of enteric viruses from environmental or fecal samples. The performance of three commercially available kits, the MO BIO PowerVirus Environmental DNA/RNA Extraction kit, the Qiagen QIAamp Viral RNA Mini kit, and the Zymo ZR Virus DNA/RNA Extraction kit was evaluated. Viral particles of adenovirus 2 (AdV), murine norovirus (MNV), and poliovirus type 1 (PV1) spiked in molecular grade water, and three different types of sample matrices (i.e., biosolids, feces, and surface water concentrates)

were extracted with the kits, and the yields of the nucleic acids were determined by quantitative PCR (qPCR). The MO BIO kit performed the best with the biosolids, which were considered to contain the highest level of inhibitors. In addition this kit provided the most consistent detection of spiked virus from all of the samples. A qPCR inhibition test using an internal control plasmid DNA and a nucleic acid purity test using an absorbance at 230 nm for the nucleic acid extracts demonstrated that the MO BIO kit was able to remove qPCR inhibitors more effectively than the Qiagen and Zymo kits. Our results suggest that the MO BIO kit is appropriate for the extraction and purification of viral nucleic acids from environmental and clinical samples that contain high levels of inhibitors.

The study summarized in Appendix C investigates the relative abundance, occurrence, and reduction of nine different viruses at two wastewater treatment plants (WWTPs) in southern Arizona over a 12 month period, from August 2011 to July 2012. Influent and effluent samples from the two WWTPs were collected monthly. Viruses were concentrated using an electronegative filter method, and quantified using TaqMan-based qPCR assays for each of the nine virus types. A plant virus, pepper mild mottle virus, was the most prevalent virus in both influent and effluent wastewater (mean concentration of $3.1 \sim 3.3 \times 10^6$ copies/L and $6.3 \sim 6.4 \times 10^5$ copies/L in influent and effluent wastewater, respectively), showing a low reduction by the treatment processes (0.68~0.72 mean $\log_{10}$ reduction), and no significant seasonal change in concentration. Aichi virus, a human enteric virus, was also found in greater abundance, and was more resistant to wastewater treatment than other human enteric viruses. Our results suggest

that these viruses could be used as potential indicators of the wastewater reclamation system performance, with respect to virus occurrence and removal.

Appendix D summarizes a study on biofilms within a POU filter medium that were developed at three different locations in the US using different surface waters. Biofilm microbial communities that developed on biofoam were analyzed utilizing 454 pyrosequencing of the 16S rDNA genes, and showed a remarkable degree of shared community membership among the three locations. A large and diverse shared microboiome defined by the top 100 shared operational taxonomic units (OTUs) at 0.03 cut off (97% similarity) which represented 280,000 of the 306,000 sequences (>90%) was found. Of those 25% were classified within the genus *Pseudomonas*. Members of the microbial communities found within the shared microbiome of the biofoam were closely associated with organisms commonly found in activated sludge, drinking water biofilms, rhizosphere, phylloshere, and soil ecosystems. The biofoam provides a unique and effective porous matrix for biofilm formation, which appears to allow for the establishment of consistent microbial communities at different locations with varied water qualities.

REFERENCES

Abbaszadegan M, Stewart P. (1999). A Strategy for Detection of Viruses in Groundwater by PCR A Strategy for Detection of Viruses in Groundwater by PCR. Applied and environmental microbiology 65:444-449.

Amann RI, Stromley J, Devereux R, Key R, Stahl D. (1992). Molecular and microscopic identification of sulfate-reducing bacteria in multispecies biofilms. Applied and environmental microbiology 58:614–623.

Amaral-Zettler L, McCliment E, Ducklow HW, Huse SM. (2009). A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes. PloS one 4:e6372.

Bauer R, Dizer H, Graeber I, Rosenwinkel K-H, López-Pila JM. (2011). Removal of bacterial fecal indicators, coliphages and enteric adenoviruses from waters with high fecal pollution by slow sand filtration. Water research 45:439–452.

Bengtsson MM, Øvreås L. (2010). Planctomycetes dominate biofilms on surfaces of the kelp Laminaria hyperborea. BMC Microbiology 10:261.

Besemer K, Peter H, Logue JB, Langenheder S, Lindström ES, Tranvik LJ, et al. (2012). Unraveling assembly of stream biofilm communities. The ISME journal 6:1459–1468.

Branda SS, Vik S, Friedman L, Kolter R. (2005). Biofilms: the matrix revisited. Trends in microbiology 13:20–26.

Buckley DH, Huangyutitham V, Nelson T, Rumberger A, Thies JE. (2006). Diversity of Planctomycetes in Soil in Relation to Soil History and Environmental Heterogeneity. Applied and Environmental Microbiology 72:6429–6429.

Bustin S, Benes V, Garson J, Hellemans J, Huggett J, Kubista M, et al. (2009). The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clinical chemistry 55:611–622.

US EPA. (2010). Method 1615 measurement of enterovirus and norovirus occurrence in eater by eulture and RT-qPCR. Accessed January 19, 2012, from http://www.epa.gov/oamcinc1/1200011/method1615.pdf.

Flemming H-C, Wingender J. (2010). The biofilm matrix. Nature reviews. Microbiology 8:623–633.

Van Frankenhuyzen JK, Trevors JT, Lee H, Flemming C, Habash MB. (2011). Molecular pathogen detection in biosolids with a focus on quantitative PCR using propidium monoazide for viable cell enumeration. Journal of microbiological methods 87:263–272.

Gibson KE, Schwab KJ. (2011). Tangential-flow ultrafiltration with integrated inhibition detection for recovery of surrogates and human pathogens from large-volume source water and finished drinking water. Applied and environmental microbiology 77:385–391.

Haig S, Dorea C, Collins G, Davies R, Quince C. (2005). Validating Laboratory Slow Sand Filtration Studies Through Water Quality and Molecular analysis. userweb.eng.gla.ac.uk.

Hamady M. (2010). Fast UniFrac: Facilitating high-throughput phylogenetic analysis of microbial communities including analysis of pyrosequencing and Phylochip data. 4:17–27.

Harwood VJ, Levine AD, Scott TM, Chivukula V, Lukasik J, Farrah SR, et al. (2005). Validity of the Indicator Organism Paradigm for Pathogen Reduction in Reclaimed Water and Public Health Protection. Applied and environmental microbiology 7:3163-3170.

Henne K, Kahlisch L, Brettar I, Höfle MG. (2012). Analysis of structure and composition of bacterial core communities in mature drinking water biofilms and bulk water of a citywide network in Germany. Applied and environmental microbiology 78:3530–3538.

Hong P-Y, Hwang C, Ling F, Andersen GL, LeChevallier MW, Liu W-T. (2010). Pyrosequencing analysis of bacterial biofilm communities in water meters of a drinking water distribution system. Applied and environmental microbiology 76:5631–5635.

Huggett JF, Foy C, Benes V, Emslie K, Garson J, Haynes R, et al. (2013). Guidelines for minimum information for publication of quantitative digital PCR experiments. Clinical chemistry 59:6.

Hunter PJ, Calvo-Bado L, Parsons NR, Pettitt TR, Petch GM, Shaw E, et al. (2012). Variation in microbial communities colonizing horticultural slow sand filter beds: implications for filter function. Irrigation science. DOI 10.1007/s00271-012-0339-z.

Ikner L a, Gerba CP, Bright KR. (2012). Concentration and recovery of viruses from water: a comprehensive review. Food and environmental virology 4:41–67.

Ikner L, Soto-Beltran M, Bright K. (2011). New method using a positively charged microporous filter and ultrafiltration for concentration of viruses from tap water. Applied and environmental microbiology 79:215-223.

Kartal B, Kuenen JG, van Loosdrecht MCM. (2010). Engineering sewage treatment with anammox. Science (New York, N.Y.) 328:702–703.

Katayama H, Shimasaki A, Ohgaki S. (2002). Development of a virus Concentration Method and Its Application to Detection of Enterovirus and Norwalk Virus from Coastal Seawater. Applied and environmental microbiology 68:1033-1039.

Kim EH, Cho KH, Lee YM, Yim JH, Lee HK, Cho J-C, et al. (2010). Diversity of cold-active protease-producing bacteria from arctic terrestrial and marine environments revealed by enrichment culture. Journal of microbiology (Seoul, Korea) 48:426–432.

Koh KS, Matz C, Tan CH, Le HL, Rice S a, Marshall DJ, et al. (2012). Minimal increase in genetic diversity enhances predation resistance. Molecular ecology 21:1741–1753.

McCoy D, Cetin A, Hausinger R. (1992). Characterization of urease from Sporosarcina ureae. Archives of microbiology 157:411–416.

Molobela IP, Cloete TE, Beukes M. (2010). Protease and amylase enzymes for biofilm removal and degradation of extracellular polymeric substances (EPS) produced by Pseudomonas fluorescens bacteria. African journal of microbiology research 4:1515-1524.

Mulcahy H, Charron-Mazenod L, Lewenza S. (2008). Extracellular DNA chelates cations and induces antibiotic resistance in Pseudomonas aeruginosa biofilms. PLoS pathogens 4:e1000213.

Mußmann M, Ribot M, von Schiller D, Merbt SN, Augspurger C, Karwautz C, et al. (2013). Colonization of freshwater biofilms by nitrifying bacteria from activated sludge. FEMS microbiology ecology 85:104-115.

Pang XL, Lee BE, Pabbaraju K, Gabos S, Craik S, Payment P, et al. (2012). Pre-analytical and analytical procedures for the detection of enteric viruses and enterovirus in water samples. Journal of virological methods 184:77–83.

Pinto AJ, Xi C, Raskin L. (2012). Bacterial community structure in the drinking water microbiome is governed by filtration processes. Environmental science and technology 46:8851–8859.

Prata C, Ribeiro A, Cunha Â, Gomes NCM, Almeida A. (2012). Ultracentrifugation as a direct method to concentrate viruses in environmental waters: virus-like particle enumeration as a new approach to determine the efficiency of recovery. Journal of environmental monitoring 14:64–70.

Ramey BE, Koutsoudis M, von Bodman SB, Fuqua C. (2004). Biofilm formation in plant-microbe associations. Current opinion in microbiology 7:602–609.

Remold S, Brown C, Farris J, Hundley T, Perpich J, Purdy M. (2011). Differential habitat Use and niche partitioning by Pseudomonas species in human homes. Microbial ecology 62:505–517.

Rhodes ER, Hamilton DW, See MJ, Wymer L. (2011). Evaluation of hollow-fiber ultrafiltration primary concentration of pathogens and secondary concentration of viruses from water. Journal of virological methods 176:38–45.

Rittmann BE. (2006). Microbial ecology to manage processes in environmental biotechnology. Trends in biotechnology 24:261–266.

Schloss PD, Gevers D, Westcott SL. (2011). Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. PloS one 6:e27310.

Schloss PD, Handelsman J. (2003). Biotechnological prospects from metagenomics. Current Opinion in Biotechnology 14:303–310.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Applied and environmental microbiology 75:7537–7541.

Spiers AJ, Buckling A, Rainey PB. (2000). MINI REVIEW The causes of Pseudomonas diversity. Microbiology 146:2345–2350.

Stanier RY, Palleroni NJ, Doudoroff M. (1966). The aerobic pseudomonads: a taxonomic study. Journal of general microbiology 43:159–271.

Stauber CE, Kominek B, Liang KR, Osman MK, Sobsey MD. (2012). Evaluation of the impact of the plastic BioSand filter on health and drinking water quality in rural Tamale, Ghana. International journal of environmental research and public health 9:3806–3823.

Tamaki H, Wright CL, Li X, Lin Q, Hwang C, Wang S, et al. (2011). Analysis of 16S rRNA amplicon sequencing options on the Roche/454 next-generation titanium sequencing platform. PloS one 6:e25263.

Thomas JM, Ashbolt NJ. (2011). Do free-living amoebae in treated drinking water systems present an emerging health risk? Environmental science and technology 45:860–869.

Tobler DJ, Cuthbert MO, Greswell RB, Riley MS, Renshaw JC, Handley-Sidhu S, et al. (2011). Comparison of rates of ureolysis between Sporosarcina pasteurii and an

indigenous groundwater community under conditions required to precipitate large volumes of calcite. Geochimica et cosmochimica acta 75:3290–3301.

Toze S. (2006). Reuse of effluent water — benefits and risks. Agriculture water management 80:147–159.

Valster RM, Wullings B a, Bakker G, Smidt H, van der Kooij D. (2009). Free-living protozoa in two unchlorinated drinking water supplies, identified by phylogenic analysis of 18S rRNA gene sequences. Applied and environmental microbiology 75:4736–46.

Whitchurch CB, Tolker-Nielsen T, Ragas PC, Mattick JS. (2002). Extracellular DNA required for bacterial biofilm formation. Science (New York, N.Y.) 295:1487.

Wong K, Fong T-T, Bibby K, Molina M. (2012). Application of enteric viruses for fecal pollution source tracking in environmental waters. Environment international 45:151–164.

Wyn-Jones a P, Carducci A, Cook N, D'Agostino M, Divizia M, Fleischer J, et al. (2011). Surveillance of adenoviruses and noroviruses in European recreational waters. Water research 45:1025–1038.

Ye L, Zhang T. (2012). Bacterial communities in different sections of a municipal wastewater treatment plant revealed by 16S rDNA 454 pyrosequencing. Applied microbiology and biotechnology 97:2681-2690.

Yilmaz P, Kottmann R, Field D, Knight R, Cole J, Amaral-Zettler L, et al. (2010) The “Mininmum information about an environmental sequence” (MIENS) specification. Available from Nature precedings http://dx.doi.org/10.1038/npre.2010.5252.2.

Zhang T, Shao M-F, Ye L. (2012). 454 Pyrosequencing reveals bacterial diversity of activated sludge from 14 sewage treatment plants. The ISME journal 6:1137–1147.

APPENDIX A

ENVIRONMENTAL GENOMICS USING NEXT GENERATION SEQUENCING TECHNOLOGIES: APPLICATIONS AND LIMITATIONS IN ENVIRONMENTAL MICROBIOLOGY

Brandon C. Iker [a], Charles P. Gerba [a], and Ian Pepper [a].

[a] Department of Soil, Water and Environmental Science, The University of Arizona, Tucson, Arizona 85721, USA

Published: Not published

ABSTRACT

Waterborne and water-based pathogens pose a significant threat to human health. Improvements in microbial water quality monitoring of water infrastructure and in treatment technologies that may one day provide better insights into potential public health risks are necessary. Microbial water quality monitoring has begun to move towards more advanced technologies in recent years, with novel molecular tools that offer rapid, sensitive, and specific detection of various microbial pathogens that challenge current culture-based techniques. Technologies like quantitative real-time PCR (qPCR) and next generation sequencing such as pyrosequencing are presently emerging as rapid tools for pathogen detection and discovery in the environment. Future challenges of using such advanced molecular techniques for environmental microbiology lie with integrating these new molecular tools with user friendly bioinformatic platforms, in developing better standardized protocols, in reducing their costs and turnaround times, and in establishing the limitations of this technology for more pointed research objectives.

I. Introduction

Since Watson and Crick described the first molecule of DNA nearly 50 years ago, its simple yet complex nature has been used to attempt to describe and understand the biological world around us. The understanding of molecular biology has been applied to a range of topics including the most simple of viral plasmids to the most complex plant genomes. Newer areas are emerging in molecular biology that not only allow us to quickly sequence a single microbial genome within an environmental sample, but also begin to elucidate the inner workings of the most complex of microbial communities within oceans, soil, or even the human biome. We are now even able detect and quantify a single molecule of DNA contained within a pool of billions. Many of the advances in DNA sequencing have been fueled by a need for an unbiased, inexpensive, easy, and accurate tool for use in areas such as genomics, personalized medicine, microbial ecology, or bioengineering. Nucleic acid sequencing and amplification technologies like Sanger sequencing and PCR have advanced significantly over the past few decades. Now, what used to require days to go from sample to sequencing to data interpretation can now be accomplished in a few hours.

Historically, 16S rDNA molecular phylogenetic library construction pioneered many of the studies in the microbial ecology field (Amann et al. 1992; Amaral-Zettler et al. 2009). More recently, the development of environmental shotgun sequencing, more commonly termed as "metagenomics", has increased in significance for sequencing applications (Schloss & Handelsman 2003). Metagenomics allows the sequencing of

portions of genetic material from individuals within a whole microbial community by randomly sequencing short (100-1000 bp) nucleic acid fragments found within community DNA. However limitations to data interpretations are still abundant mostly due to the fact that ~90% of sequencing data in public databases lack meta-data along with persistent limitations in computational power. Continuous improvements in the volume of data produced by ever evolving sequencing technologies steadily outstrip the computational abilities available to analyze larger and larger data sets. In addition, unidentified biases and errors in data production, and a paucity of information on standard operating procedures leave many investigators unable to compare complicated data sets to public databases (Yilmaz et al. 20010)

Applications of next generation sequencing technologies (NGST) in environmental microbiology have only recently been realized. High sequencing costs have been among the limiting factors for widespread implementation of this technology; however, in recent years these costs have dropped dramatically. Traditional techniques in environmental microbiology have relied heavily on cultural or PCR based assays. Such techniques are still exceedingly useful components within the field, but limitations in cost, time, accuracy, etc. are driving the evolution of these techniques. In addition, even the assessment of basic water quality can take >24 hours using culture based assays. In the case of PCR, assays are typically limited to just one organism or target piece of DNA/RNA and provide no information about the organism's viability. In contrast, next generation sequencing allows for nearly autonomous, deep, rapid, and unbiased sequencing of microbial communities. Applications of NGST in environmental

microbiology open up many opportunities for improvements in public health and a greater understanding of the effects that human activities have on the environment.

Glossary of terms:

Next generation sequencing technology: The technology and equipment required to perform high throughput sequencing. Examples: 454-Pyrosequencing, Illumina, Ion Torrent

Metagenomics (the "shotgun-approach"): The application of modern genomic techniques to the study of communities of microorganisms *in situ* in their natural environment, eliminating the need for isolation and lab cultivation of individual species

Transcriptomics: Similar to metagenomics, but involves the analysis of all RNA molecules within a sample through the indirect sequencing of synthesized cDNA molecules via a process known as reverse transcription.

Sequencing "platform": the type of machine used in the sequencing analysis

RAM: Random access memory - the required computing power needed to undertake a large data set analysis.

OTU: Operational Taxonomic Unit - defined by the user as a theoretical cutoff to classify organisms at a molecular level.

Library: the isolation and preparation of one sample of environmental DNA/RNA. For example, one 16s or metagenomic DNA/RNA analysis from 100 ml of wastewater effluent.

Tagged sample run (pyro-tags, barcodes): Short sequences of DNA primers with known sequences that are adapted to each sequence from a single sample. Once the samples are "tagged," they are then combined with other samples and then separated bioinformatically following sequencing.

Genetic fingerprint: a molecular characterization of a microbial community based on a phylogenetic gene.

Bioinformatics: the application of statistics and computer science to the field of molecular biology.

Paired end reads: Short sequences that are sequenced at known distances from each other that act as a map for genome sequence analysis and longer rRNA coverage.

2. Next generation sequencing technologies (NGST): Review of Technology

Currently, billions of DNA fragments can be spread out over a plate no bigger than a credit card and sequenced simultaneously. To put this into perspective, all 3.4 billion base pairs of our human genome could be draft sequenced with one machine in a single afternoon. Similarly, when attempting to survey a microbial community, billions of 16S rRNA sequences could be sequenced just as rapidly. High throughput sequencing allows for a very deep, very rapid, genetic fingerprint of multiple microbial communities simultaneously and at a low cost per sample.

High throughput sequencing technologies, like 454-pyrosequencing, use a "sequence by synthesis" approach. Detailed discussions of these concepts and other NGST are described elsewhere (Ansorge 2009) and summary of the technologies can be found in Table 1. Briefly, DNA nucleotides are incorporated into the synthesis of a new strand of DNA, light is emitted that matches the color of the given base in a sequence. The light is derived from the release of pyrophosphate during DNA synthesis. The pyrophosphate is converted to ATP by the enzyme sulfurylase, which provides and energy source for luciferase that in turn oxidizes luciferin which produces a detectable light emission, indicating that a base has been incorporated (Pourmand et al. 2002; Ronaghi & Elahi 2002). The bases are then introduced one at a time, and the light emission from each nucleotide added is recorded. Pyrophosphate nucleic acid base incorporation detection has now been replaced by adding different colored fluorophores (light producing molecules) to each complement nucleotide base. For example, green = guanine, red = cytosine, blue = tyrosine, and yellow = adenine. The nucleotide base can

then be visualized on a nano scale by evaluating the different color light emission during base incorporation. Computer algorithms then can be used to identify the DNA bases based on the signal quality as each base is read. All sequencing platforms have basic sequencing analysis software that aid in the sequencing, cleaning, and basic manipulation of data. These data sets can then be exported to any number of external data analysis software (Table 2).

Table 1. Comparison of NGST for applications in applied environmental microbiology

Company	Throughput	Length	Total # of reads/run	Error rate per base	Cost	Cost/16S 2K reads	Applications	Reference
Sanger	6 Mb/day	800 nt	100	10^{-4}-10^{-5}	~$500/Mb	~$2,000/16S library	small sample size, genome gaps, long homopolymer regions	(Sanger & Coulson 1975)
454/Roche	750 Mb/day	400 nt	~10^{6}	10^{-3}-10^{-4}	~$20/Mb	~$250-500/16s library*	Complex genomes, SNPs, tagged sample runs, paired end reads.	(Margulies et al. 2005)
Illumina	5,000 Mb/day	100 nt	~35^{7}	10^{-2}-10^{-3}	~$0.50/Mb	~$200-500/16s library*	Complex genomes, SNPs, tagged sample runs, genome polishing, paired end reads	(Bentley 2006)
SOLiD	5,000 Mb/day	50 nt	~10^{9}	10^{-2}-10^{-3}	~$0.50/Mb	~$200-500/16s library*	Complex genomes, SNPs, genome polishing, paired end reads	(Hedges et al. 2011)
Helicos	5,000 Mb/day	32 nt	~10^{9}	10^{-2}	<$0.50/Mb	Not yet applicable	Non-amplified sequencing, quantification and direct RNA sequencing	(Pushkarev et al. 2009)

*costs are highly variable depending on type of library construct.

Platform = type of sequencing machine used

Throughput = number of sequences per day

nt = nucleotides

run = total sequencing capacity of one use of a platform

Quality = (Q) = -10 log p, where p is the probability of an incorrect base

Library = one sample preparation

SNPs = Single nucleotide polymorphisms

Paired end reads = used to complete genomes and increase 16S sequencing coverage

tagged sample runs = barcoding technique that allows for multiple samples per run

3. Effective data handling and analysis when dealing with large data sets

Bioinformatic support can prove challenging for microbiologists not well-versed in computer sciences (Kunin & Hugenholtz 2010; Kunin et al. 2010). The ability of NGST platforms to produce exceedingly large data sets leads to logistical problems with even basic data handling. Most sequencing analyses are RAM intensive, leaving investigators' desktop computers unable to approach the computational needs of even basic datasets. Some have worked around this problem by uploading their data sets over the web to online servers that perform the analyses for them. Other computer based software packages allow for more a customized analysis such as operational taxonomic units (OTUs) for 16S libraries. These are able to decrease the size of phylogenetic datasets by grouping sequences together into OTUs, therefore limiting the amount of data that needs to be analyzed and thus is approachable for commercially available computers.

Table 2. Summary of open source analysis software for NGST data

Software	GUI	Type of Data	Computational needs	Reference
Waters	Yes	LTS, alignment, classification, basic community analysis, data figures	MAC OS, Cloud RAM intensive	(Hartman et al. 2010)
Mothur	No	HTS/LTS, alignment, classification, basic community	MAC OS, Linux, Windows, Coud RAM intensive	(Schloss et al. 2009)

		analysis, data figures		
QIIME	No	HTS, alignment, classification, basic community analysis, data figures	Cloud, Linux, Windows RAM intensive	(Caporaso et al. 2010)
ARB/SILVA	Yes	LTS, tree construction, basic alignment and visualization	Linux, MAC OS RAM intensive	(Pruesse et al. 2007)
RDP-II	Web	Limited analysis of HTS, classification, alignment, basic community analysis and small tree building	Internet	(Cole 2003)
JGI (Genome Portal and IMG)	Web	metagenomes genomes	Internet, basic perl for data manipulation	(Markowitz et al. 2006)
RAST	Web	metagenomes	Internet, basic perl for data manipulation	(Aziz et al. 2008)
SEED	Web	genomes	Internet, basic perl for data manipulation	(Overbeek et al. 2004)
CAMERA	Web	metagenomes	Internet, basic perl for data manipulation	(Seshadri et al. 2007)

GUI = Graphical User Interface (Not command line driven i.e. more user friendly)

HTS = High throughput sequencing(datasets >3000 sequences)

LTS = Low throughput sequencing(datasets <3000 sequences)

RAM intensive = requires specialized computing equipment for larger data sets

Basic community analysis = species richness, evenness, relative abundance, shared OTUs, etc.

Depending on the type of analysis desired, different software packets offer varying strategies on how to handle the type of data (phylogenetic or genomic), the size of the data set, and the types of questions that can be addressed (Table 2). Overall, the computational abilities of investigators, with respect to the lack of computing power or agnostic manipulation of large and complex data sets, is a limiting factor in the widespread application of NGST in environmental microbiology.

4. Considerations for platform selection and experimental procedures.

Depending on the application, the sequence read length and platform used will affect the type analysis that can be performed. For applications such as metagenomics and genome sequencing, longer sequence read lengths give the best results, with a few caveats (Wommack et al. 2008; Wooley et al. 2010). A summary of the available technologies can be found in in Table 1. Longer sequence reads (i.e., >300 bps) allow for more overlap of sequence fragments into longer, more contiguous fragments and aid in more accurate identification since they provide more bases of a putative gene that may be present for comparison. For sequencing applications of phylogenetic

diversity using 16S rDNA sequences or other phylogenetic genes, short reads (e.g.,100-300 bps) of the V6 hyper-variable region of the 16S rRNA gene sequence have been shown to be effective in classifying bacteria and placing them phylogenetically to the genus level within an 80 % confidence interval. Some researchers feel that sequencing error rates in conjunction with shorter reads may lead to higher sequencing error rates, which can artificially inflate diversity estimates (Huse et al. 2010). Depending upon the question, the errors associated with a sequence may not be significantwhen attempting to fingerprint a microbial community, assuming that the OTUs are properly vetted during data analysis.

Unfortunately, many of the biases and assumptions that accompany PCR also apply to any high throughput sequencing attempts that involve a prior PCR amplification step. The PCR amplification of phylogenetic genes like 16S are exponentially amplified, leaving less abundant sequences in a sample under represented. Moreover, 16S rRNA gene copy number can lead to even greater bias in microbial community analysis (Crosby & Criddle 2003). Fortunately, PCR biases are not associated with DNA metagenomic applications because no amplification steps are required for direct sequencing of fragmented DNA from a sample. In contrast, cDNA library constructions require random primers and reverse transcriptase in order to be sequenced, and the errors and biases associated with these techniques are still not fully understood.

NGST currently offer little advancement in resolvinv the issue of live verses dead cells when using 16S rRNA libraries. Techniques using propidium monoazide (PMA) have been developed to eliminate the amplification/detection of nascent DNA and DNA that is inside of dead cells with compromised cell walls (Nocker et al. 2007;

Nocker & Camper 2009), but have only recently been tested with NGST (Yergeau et al. 2010). PMA works by entering the compromised cell and intercalates within the DNA backbone, rendering dead cells DNA un-amplifiable. Some investigators have suggested that targeting mRNA and rRNA may be more successful at identifying the viability of an organism in a sample because RNA theoretically persists for only a short amount of time after cell death (Dinsdale et al. 2008; Wooley et al. 2010). However, RNA analysis or transcriptomics may not account for cell dormancy and the biases in RNA extraction due to the secondary structures associated with RNA molecules.

5. Applications of NGST in environmental microbiology.

Environmental genomics using NGST can aid in the development of bioremediation approaches for sites difficult to clean up by potentially identifying genes that encode for enzymes that rapidly degrade the contaminant or provide novel enzymatic tools for applications in bioremediations projects (Stenuit et al. 2009; Schoenfeld et al. 2011). This may help eliminate costly exploratory stages in remediation approaches by assessing the genetic functional capacity of a given microbial community before spending millions of dollars on an unsuccessful remediation project.

Recently, NGST have been used to investigate the human microbiome (Nakamura et al. 2009) and as diagnostic tools in the hospital setting (Clarridge 2004). By analyzing DNA and/or RNA metagenomic libraries, investigators have been able to successfully identify organism(s) present that are associated with symptoms of an infected patient. Sampling efforts of mucosal swabs and fecal

samples yield hundreds of thousands of sequences; however, of those, few are identifiable as the pathogen of interest.

The success of pathogen detection using NGST in the environment still needs to be further investigated, especially in the case of environmental virus detection. Of the few studies attempting to directly detect viral human pathogens in wastewater or reclaimed water, few have been successful (Rosario et al. 2009; Bibby et al. 2010; Sanapareddy et al. 2009). However, it is important to note that the detection of many viral plant pathogens were readily detected in almost all of these studies, and may be of great relevance to the irrigation of arid crops using reclaimed water. A study conducted by Rosario et al. (2009) showed that RNA and DNA metagenomic sequence libraries were only able to classify ~40% of the reads. Of the classified sequences, approximately 60% were identified as pepper mild mottle virus (PMMV) (Zhang et al. 2006). Future efforts may need to be directed towards the detection of viral pathogens by improving qPCR, as well as the development of more statistically relevant multiple marker approaches.

Bacterial pathogen detection is now becoming more achievable because 16S rRNA sequencing depths are able to reach deep sequencing levels. Investigators are now able to obtain 10^9 sequences on some of the newer sequencing platforms, which can enhance exposure assessments. Bibby and colleagues (2010) found that human bacterial pathogens were detectible with greater sequencing depth (~30,000 sequences per sample) in class B biosolids. Of the samples tested in this study, all contained pathogens, but at very low abundances - ranging from 0.02% to 0.1% of the classified sequences. Of the identified pathogens, 61% were *Clostridium* and *Mycobacterium*, but similar distributions of the same pathogens were found in a native soil control

sample. Other studies of wastewater treatment plants have shown that even in environments thought to be rich in human pathogens, overwhelming genetic diversity has led to poor detection when using deep metagenomic coverage (Sanapareddy et al. 2009). Other large scale sequencing efforts to detect bacterial pathogens in human waste solids or biosolids have resulted in more progress towards improving virus detection.

The field of microbial source tracking may greatly benefit from deep sequencing approaches when trying to characterize the microbial community present in a given environment. By allowing researchers to identify the majority of microbes present, better correlations can perhaps be made when attempting to identify the molecular indicator sequences of the presence of fecal contamination. Multiple marker approaches in a field that relies heavily on statistical modeling is of the utmost importance in evaluating ideal pathogen indicator candidates in any number of matrices (Harwood et al. 2005). Future applications of NGST may allow for simultaneous selective detection, identification, and quantification of target sequences using novel sequence capture techniques in concert with high throughput sequencing technologies.

6. Concluding remarks and future directions

NGST may have applications in assessing environmental gene transfer (e.g., plasmids) and in discovering novel biocatalysts, drugs, chemicals, and other useful enzymes (Warnecke & Hess 2009). For example, the release of plasmid DNA encoding potentially problematic genetic sequences, like antibiotic resistance or

virulence factors into the environment and distribution systems is still poorly understood. We know that antibiotic resistance genes (ARG) and antibiotic resistant bacteria (ARB) are present in the environment and remain at high levels due to anthropogenic impacts, but selective pressures and mechanisms have not yet been well described (Rizzo et al. 2013; Fatta-Kassinos et al. 2011; Baquero et al. 2008; Rahube & Yost 2010; Schwartz et al. 2003). Information on the types of organisms that may be able to acquire ARG plasmids and other virulence factors may be more easily accessed by deep metagenomic surveillance of water distribution systems and their related biofilms.

NGST using transcriptomics may shed light on the microbial interactions in systems like drinking water biofilms. For example, Acanthamoebae can harbor *Legionella* intracellularly, protecting them fromdrinking water treatment and chlorination events. The reasons for this shared lifestyle are still not fully understood (Thomas et al. 2011). Potential studies of mRNA expression of both organisms, separately and combined, may allow further understanding of their interactions, regardless of whether they are parasitic, mutualistic, or beneficial based on changes in gene expression levels and functional estimations of each organism's transcriptome. Understanding interactions like these may help us understand bacterial re-growth and aid in developing protection and prevention methods for the management of water-based pathogens in water distribution systems.

References

Amann RI, Stromley J, Devereux R, Key R, Stahl D. (1992). Molecular and microscopic identification of sulfate-reducing bacteria in multispecies biofilms. Applied and environmental microbiology 58:614–623.

Amaral-Zettler L, McCliment E, Ducklow HW, Huse SM. (2009). A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes. PloS one 4:e6372.

Ansorge WJ. (2009). Next-generation DNA sequencing techniques. New biotechnology 25:195–203.

Aziz RK, Bartels D, Best A, DeJongh M, Disz T, Edwards R, et al. (2008). The RAST server: rapid annotations using subsystems technology. BMC genomics 9:75.

Baquero F, Martínez J-L, Cantón R. (2008). Antibiotics and antibiotic resistance in water environments. Current opinion in biotechnology 19:260–265.

Bentley DR. (2006). Whole-genome re-sequencing. Current opinion in genetics & development 16:545–552.

Bibby K, Viau E, Peccia J. (2010). Pyrosequencing of the 16S rRNA gene to reveal bacterial pathogen diversity in biosolids. Water research 44:4252–4260.

Caporaso JG, Lauber CL, Walters WA, Berg-lyons D, Lozupone CA, Turnbaugh PJ, et al. (2010). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. PNAS 108: 4516–4522.

Cole JR. (2003). The Ribosomal Database Project (RDP-II): previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy. Nucleic Acids Research 31:442–443.

Crosby LD, Criddle CS. (2003). Understanding bias in microbial community analysis techniques due to rrn operon copy number heterogeneity. BioTechniques 34:790–4, 796, 798.

Dinsdale E, Edwards R, Hall D, Angly F, Breitbart M, Brulc JM, et al. (2008). Functional metagenomic profiling of nine biomes. Nature 452:629–632.

Fatta-Kassinos D, Kalavrouziotis IK, Koukoulakis PH, Vasquez MI. (2011). The risks associated with wastewater reuse and xenobiotics in the agroecological environment. The Science of the total environment 409:3555–3563.

Hartman AL, Riddle S, McPhillips T, Ludäscher B, Eisen J. (2010). Introducing W.A.T.E.R.S.: a workflow for the alignment, taxonomy, and ecology of ribosomal sequences. BMC bioinformatics 11:317.

Harwood VJ, Levine AD, Scott TM, Chivukula V, Lukasik J, Farrah SR, et al. (2005). Validity of the Indicator Organism Paradigm for Pathogen Reduction in Reclaimed Water and Public Health Protection Validity of the Indicator Organism Paradigm for Pathogen Reduction in Reclaimed Water and Public Health Protection. Applied environmental microbiology 71:3163-3170.

Hedges DJ, Guettouche T, Yang S, Bademci G, Diaz A, Andersen A, et al. (2011). Comparison of three targeted enrichment strategies on the SOLiD sequencing platform. PloS one 6:e18595.

Huse SM, Welch DM, Morrison HG, Sogin ML. (2010). Ironing out the wrinkles in the rare biosphere through improved OTU clustering. Environmental microbiology 12:1889–1898.

Claridge JE. (2004). Impact of 16S rRNA Gene Sequence Analysis for Identification of Bacteria on Clinical microbiology and infectious diseases 17:840–862.

Kunin V, Engelbrektson A, Ochman H, Hugenholtz P. (2010). Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. Environmental microbiology 12:118–123.

Kunin V, Hugenholtz P. (2010). PyroTagger: A fast, accurate pipeline for analysis of rRNA amplicon pyrosequence data. The open journal 1:1–8.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben L, et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376–380.

Markowitz VM, Korzeniewski F, Palaniappan K, Szeto E, Werner G, Padki A, et al. (2006). The integrated microbial genomes (IMG) system. Nucleic acids research 34:344–348.

Nakamura S, Yang C-S, Sakon N, Ueda M, Tougan T, Yamashita A, et al. (2009). Direct metagenomic detection of viral pathogens in nasal and fecal specimens using an unbiased high-throughput sequencing approach. PloS one 4:e4219.

Nocker A, Camper AK. (2009). Novel approaches toward preferential detection of viable cells using nucleic acid amplification techniques. FEMS microbiology letters 291:137–142.

Nocker A, Sossa-Fernandez P, Burr MD, Camper AK. (2007). Use of propidium monoazide for live/dead distinction in microbial ecology. Applied and environmental microbiology 73:5111–5117.

Overbeek R, Disz T, Stevens R. (2004). The SEED: a peer-to-peer environment for genome annotation. Communication ACM 47:46–51.

Pourmand N, Elahi E, Davis RW, Ronaghi M. (2002). Multiplex Pyrosequencing. Nucleic acids research 30:e31.

Pruesse E, Quast C, Knittel K, Fuchs BM, 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 research 35:7188–7196.

Pushkarev D, Neff NF, Quake SR. (2009). Single-molecule sequencing of an individual human genome. Nature biotechnology 27:847–850.

Rahube T, Yost C. (2010). Antibiotic resistance plasmids in wastewater treatment plants and their possible dissemination into the environment. African journal of biotechnology 9:9183–9190.

Rizzo L, Manaia C, Merlin C, Schwartz T, Dagot C, Ploy MC, et al. (2013). Urban wastewater treatment plants as hotspots for antibiotic resistant bacteria and genes spread into the environment: A review. The science of the total environment 447:345–360.

Ronaghi M, Elahi E. (2002). Pyrosequencing for microbial typing. Journal of chromatography B 782:67–72.

Rosario K, Nilsson C, Lim YW, Ruan Y, Breitbart M. (2009). Metagenomic analysis of viruses in reclaimed water. Environmental microbiology 11:2806–2820.

Sanapareddy N, Hamp TJ, Gonzalez LC, Hilger H, Fodor A, Clinton SM. (2009). Molecular diversity of a North Carolina wastewater treatment plant as revealed by pyrosequencing. Applied and environmental microbiology 75:1688–1696.

Sanger F, Coulson AR. (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. Journal of molecular biology 94:441–448.

Schloss PD, Handelsman J. (2003). Biotechnological prospects from metagenomics. Current opinion in biotechnology 14:303–310.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. (2009). Introducing mothur: open-source, platform-independent, community-supported

software for describing and comparing microbial communities. Applied and environmental microbiology 75:7537–7541.

Schoenfeld T, Liles M, Wommack KE, Polson SW, Mead D. (2011). Molecular tools. 18:1–19.

Schwartz T, Kohnen W, Jansen B, Obst U. (2003). Detection of antibiotic-resistant bacteria and their resistance genes in wastewater, surface water, and drinking water biofilms. FEMS microbiology ecology 43:325–335.

Seshadri R, Kravitz S, Smarr L, Gilna P, Frazier M. (2007). CAMERA: a community resource for metagenomics. PLoS biology 5:e75.

Stenuit B, Eyers L, Schuler L, George I, Agathos S. (2009). Molecular Tools for Monitoring and Validating Bioremediation. In:Advances in Applied Bioremediation SE - 18, Singh, A, Kuhad, RC, & Ward, OP, (ed). Vol. 17, Springer Berlin Heidelberg, pp. 339–353; DOI 10.1007/978-3-540-89621-0_18.

Thomas JM, Ashbolt NJ. (2011). Do free-living amoebae in treated drinking water systems present an emerging health risk? Environmental science and technology 45:860–869.

Warnecke F, Hess M. (2009). A perspective: metatranscriptomics as a tool for the discovery of novel biocatalysts. Journal of biotechnology 142:91–95.

Wommack KE, Bhavsar J, Ravel J. (2008). Metagenomics: read length matters. Applied and environmental microbiology 74:1453–1463.

Wooley JC, Godzik A, Friedberg I. (2010). A primer on metagenomics. PLoS computational biology 6:e1000667.

Yergeau E, Hogues H, Whyte LG, Greer CW. (2010). The functional potential of high Arctic permafrost revealed by metagenomic sequencing, qPCR and microarray analyses. The ISME journal 4:1206–1214.

Yilmaz P. The “Mininmum information about an environmental sequence” (MIENS) specification. Available from Nature Precedings http://dx.doi.org/10.1038/npre.2010.5252.2.

Zhang T, Breitbart M, Lee WH, Run J-Q, Wei CL, Soh SWL, et al. (2006). RNA viral community in human feces: prevalence of plant pathogenic viruses. PLoS biology 4:e3.

APPENDIX B

RELATIVE ABUNDANCE AND TREATMENT REDUCTION OF VIRUSES DURING WASTEWATER TREATMENT PROCESSES – IDENTIFICATION OF POTENTIAL VIRAL INDICATORS

Masaaki Kitajima,*[,†] *Brandon C. Iker*,[†] *Ian L. Pepper*,[†] *Charles P. Gerba*[†]

[†]Department of Soil, Water and Environmental Science, The University of Arizona, Tucson, Arizona 85721, United States

KEYWORDS: Viruses; Quantitative PCR; Wastewater, Treatment; Indicators, Seasonal Occurrence

Submitted to: Environmental Science & Technology (ACS Publications)

ABSTRACT

Waterborne pathogenic viruses discharged from wastewater treatment plants (WWTP) pose potential public health risks. In the present study, we investigated the occurrence and relative abundance of nine different viruses in wastewater and their removal by two WWTPs in southern Arizona over a 12 month period from August 2011 to July 2012. Influent and effluent samples from the two WWTPs were collected monthly. Viruses were concentrated using an electronegative filter method and quantified using TaqMan-based qPCR assays for each of the nine virus types. The pepper mild mottle virus, a plant virus, was found to be the most prevalent virus in both the influent and effluent wastewater (mean concentration of $3.1 \sim 3.3 \times 10^6$ copies/L and $6.3 \sim 6.4 \times 10^5$ copies/L in influent and effluent samples, respectively), showing a low reduction by the treatment processes (0.68~0.72 mean $\log_{10}$ reduction), and no significant seasonal change in concentration. Aichivirus, a human enteric virus, was also found in greater abundance, and was more resistant to wastewater treatment than other human enteric viruses. Our results suggest that these viruses could be used as potential indicators of wastewater reclamation system performance, with respect to virus occurrence and removal.

1. INTRODUCTION

Increased water consumption associated with exploding human populations and limited precipitation within arid and semi-arid areas in the United States as well as in other parts of the world has perpetuated a growing shortage of water supply. To address this problem, reclaimed or recycled water derived from treated municipal wastewater is being used for various purposes, such as direct and indirect potable reuse, industrial use, agricultural irrigation, recreational use, and environmental enhancement, which if implemented correctly, provide a safe, sustainable, and feasible strategies to manage limited water resources (Levine 2004).

The potential public health risks associated with wastewater reuse are mainly derived from the insufficient removal of pathogenic viruses, which are commonly found in high concentrations in wastewater and are highly infectious to humans. Thus, the possibility of inadequate treatment of pathogenic viruses by wastewater treatment plants (WWTP) that use treated wastewater for reuse purposes requires additional scrutiny (Harwood 2005). In addition, the concentration of viruses in treated wastewater may vary according to the type of the wastewater treatment process, season, geographical area, and hygiene conditions within the community, which makes it difficult to generalize about the occurrence of pathogenic viruses in treated wastewater using traditional indicator or model organisms (Gerba 2013).

Currently, the microbiological safety of reclaimed water is indirectly assessed through the routine monitoring of bacterial indicators in the disinfected effluent water; however, human pathogenic viruses are more resistant than bacterial indicators such as total coliforms and fecal coliforms (e.g., *Escherichia coli*) during the wastewater

treatment process (Gerba 2013). Traditional bacterial indicators are therefore not always appropriate predictors of the occurrence and fate of viral pathogens during wastewater treatment (Baggi et al., 2001; Gerba 1979). Bacteriophages have also been proposed as indicators of viral contamination (AWPRC 1991), but their presence does not always correlate with the occurrence of human enteric viruses (Hot 2003). Accordingly, several types of viruses such as adenoviruses (AdVs), polyomaviruses, enteroviruses (EVs), and pepper mild mottle virus (PMMoV), have recently been suggested as potential indicators of the presence of viruses in water (Silva 2011; Albinana-Gimenez 2009; Hot 2003; Hamza 2011). Recent advancements in molecular techniques, especially quantitative PCR (qPCR), have enabled the detection and quantification of a wide range of pathogenic and indicator viruses, including emerging and non-culturable viruses, in water (Girones 2010).

In the present study, we investigated the relative abundance, occurrence, and reduction of nine different viruses at two WWTP in southern Arizona throughout a one-year period with the goal of identifying a conservative viral indicator of human fecal contamination for tracking the fate and transport of pathogenic viruses in the environment and wastewater reuse schemes. The criteria that we used to identify the optimal indicator viruses included the following: 1) no observable seasonal changes in abundance, 2) a low removal during wastewater treatments, 3) a high relative abundance in comparison to well-studied enteric viruses such as AdV and EVs, and 4) considered to be specific to human fecal contamination.

2. MATERIALS AND METHODS

2.1. Collection of wastewater samples. Between August 2011 and July 2012, influent and effluent wastewater grab samples were collected monthly from two WWTP (Plants A and B) located in southern Arizona. Plant A utilized a conventional activated sludge process and plant B utilized a biological trickling filter process or biotower. In addition, both plants used chlorination for disinfection. All samples were collected in sterile plastic bottles, stored on ice, and transported to the laboratory, where they were processed within 12 h of collection. To determine whether the microbiological water quality of effluent water met the criteria for recreational water (USEPA 1986), *E. coli* in 100 mL of the effluent water sample was assayed by the Colilert ® method (SM 9223B), and expressed as the most probable number (MPN)/100 mL (APHA 2005).

2.2. Concentration of viruses in wastewater samples. A total of 48 wastewater samples (12 influent and 12 effluent samples each from both plants) were collected and concentrated using an electronegative filter method as described previously (Kitajima et al. 2012). Briefly, 2.5 M $MgCl_2$ was added to the wastewater samples to obtain a final concentration of 25 mM. The samples (100 mL influent and 1,000 mL effluent) were subsequently passed through the electronegative filter (cat. no. HAWP-090-00; Millipore, Billerica , MA) attached to a glass filter holder (Advantec, Tokyo, Japan). Magnesium ions were removed by passing 200 mL of 0.5 mM H_2SO_4 (pH 3.0) through the filter, and the viruses eluted with 10 mL of 1.0 mM NaOH (pH 10.8). The eluate was recovered in a tube containing 50 µL of 100 mM H_2SO_4 (pH 1.0) and 100 µL of 100× Tris–EDTA buffer (pH 8.0) for neutralization, followed by further centrifugal concentration using a Centriprep YM-50 (Millipore) to obtain a final

volume of approximately 650 µL. A previous study showed that the recovery efficiencies of poliovirus type 1 spiked into influent and effluent wastewater using this method were 23 ± 19 % and 65 ± 28 %, respectively (Katayama et al. 2008). The water concentrates were stored at −80°C until further analysis.

2.3. Sample process control for quantification of viral genomes. Murine norovirus (MNV, S7-PP3 strain), kindly provided by Dr. Y. Tohya (Nihon University, Kanagawa, Japan) and propagated in RAW 264.7 (ATCC TIB-71) cells (American Type Culture Collection, Manassas, VA, USA), was used as a sample process control to determine the efficiency of extraction-RT-qPCR, as previously described (Hata et al., 2012). Briefly, 2.0 µL of MNV stock (4.0×10^4 copies/µL) was spiked into 200 µL of concentrated wastewater samples, and pure water (as a control). MNV-RNA was co-extracted with other indigenous viral nucleic acids from the water samples, and the MNV-RNA yield was subsequently determined by RT-qPCR (Kitajima et al. 2010) to calculate the extraction-RT-qPCR efficiency (%).

2.4. Extraction of viral nucleic acids and RT. Viral DNA and RNA was extracted from the concentrated wastewater sample spiked with the MNV process control (202 µL in total) using the ZR Viral DNA/RNA Kit (Zymo Research, Irvine, CA) to obtain a final volume of 100 µL, according to the manufacturer's protocol.

The RT reaction was performed using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Foster City, CA). Briefly, 10 µL of extracted RNA was added to 10 µL of RT mixture containing 2 µL of 10× reverse transcription buffer, 0.8 µL of 25× deoxynucleoside triphosphates (dNTPs), 2 µL of 10× random hexamers, 50 units of MultiScribe™ reverse transcriptase, and 20 units of RNase

inhibitor. The RT reaction mixture was incubated at 25°C for 10 min, followed by 37°C for 120 min, and finally 85°C for 5 min to inactivate the enzyme.

2.4. Quantification of viral genomes by qPCR. TaqMan-based qPCR assays for viruses were performed with a LightCycler® 480 Real-Time PCR Instrument II (Roche Diagnostics, Mannheim, Germany). Reaction mixtures (25 μL) consisted of 12.5 μL of LightCycler® 480 Probes Master (Roche Diagnostics), forward and reverse primers, probe(s), and 2.5 μL of (c) DNA template. The sequences of primers and probes are shown in Table S1 in the Supporting Information. The reaction mixtures were subjected to thermal cycling and fluorescence readings were collected and analyzed with LightCycler® 480 Software version 1.5 (Roche Diagnostics). The genome copy numbers of each virus were determined based on the standard curve prepared with 10-fold serial dilutions of plasmid DNA containing each virus gene to be amplified, at a concentration of 10^7 to 10^0 copies per reaction.

2.5. Statistical analyses. Student's *t*-tests were performed with Microsoft Excel for Mac 2011 (Microsoft Corp., Redmond, WA) to determine whether the $\log_{10}$ reductions at Plant A and B were statistically different. Differences were considered statistically significant if the resultant *P* value was 0.05 or lower.

3. RESULTS

3.1. Viral nucleic acid extraction-RT-qPCR efficiency. In order to monitor the efficiency of RNA extraction-RT-qPCR for quantitative detection of viruses, a known amount of MNV (8.0×10^4 copies) was spiked into the concentrated wastewater samples as a process control. The mean recovery efficiencies of MNV were determined to be more than 75% (Table 1).

3.2. Occurrence and abundance of viruses and E. coli in wastewater. The occurrence of a total of nine types of viruses including eight types of human enteric viruses [norovirus (NoV) genogroups I (GI), GII, and GIV, sapovirus (SaV), Aichivirus (AiV), adenovirus (AdV), enteroviruses (EV), group A rotavirus (ARV)] and pepper mild mottle virus (PMMoV), a plant virus, was determined in influent (untreated wastewater) and effluent (treated wastewater after disinfection) samples by (RT-)qPCR (Figure 1).

GI and GII NoVs were detected in all influent samples from both plants and were detected in nine (75%) effluent samples (Figure 1A and 1B). GIV NoV, which has rarely been identified from environmental samples, was detected in eight (67%) influent and three (25%) effluent samples (Figure 1C). SaV, a human calicivirus, was detected in all but one (collected in September 2011) of the influent samples from Plant A, and was detected in nine (75%) effluent samples for both plants (Figure 1D). AiV, a picornavirus, was detected in all influent and effluent samples from both plants and its concentration was fairly high and consistently stable in both influent and effluent wastewater throughout the year (influent: 8.1×10^4 and 1.5×10^6 copies/L in Plant A and B, respectively; effluent: 1.1×10^4 and 2.4×10^5 copies/L in Plant A and B, respectively) (Figure 1E). EV were detected in all influent and in 11 (92%) effluent

samples (Figure 1F). AdV was detected in all samples except two influent samples (collected in March and July 2012) from Plant B (Figure 1G). ARV was detected in eight (67%) influent and 10 (83%) effluent samples, with a clear seasonality (i.e., higher positive rate in spring to early summer seasons than the other seasons; Figure 1H). PMMoV had the highest mean concentration in the wastewater samples (influent: 3.1×10^6 and 3.3×10^6 copies/L in Plant A and B, respectively; effluent: 6.4×10^5 and 6.3×10^5 copies/L in Plant A and B, respectively) among the viruses tested and showed little seasonal variation (Figure 1I; also see Figure S1 in the Supporting Information).

E. coli was detected in nine (75%) effluent samples from both of the plants, with an annual maximum concentration of 1.4×10^2 and 2.3×10^2 MPN/100 ml in Plant A and B, respectively (Figure 1J). In addition, one (8.3%) out of 12 effluent samples for each plant exceeded the criteria for *E. coli* in recreational water recommended by the US Environmental Protection Agency (i.e., 126 MPN/100 ml) (USEPA 1986).

3.3. Reduction of viruses by wastewater treatment. The reduction of viruses was calculated from the samples that were qPCR-positive for both the influent and effluent(Figure 2). Among the nine virus types tested, GII NoV showed the highest reductions (2.04 ± 1.01 and 2.64 ± 0.61 $\log_{10}$ reductions for Plants A and B, respectively), followed by GI NoV (1.57 ± 1.13 and 2.37 ± 1.05 $\log_{10}$ reductions for Plant A and B, respectively), with high variability in the $\log_{10}$ reduction over the year; however, these differences between the reductions of GI and GII NoV were not statistically significant (*t*-test, $P > 0.05$). In contrast, the reduction of AiV was relatively low and almost constant throughout the year (0.86 ± 0.36 and 0.81 ± 0.08 $\log_{10}$ reductions for Plants A and B, respectively). The reduction of PMMoV ($0.68 \pm$

0.51 and 0.72 ± 0.70 $\log_{10}$ reductions for Plants A and B, respectively) was even less than that of AiV.

When the reductions of viruses at Plants A and B were compared, the $\log_{10}$ reduction of only SaV at Plant A was significantly higher than that of Plant B (*t*-test, $P = 0.042$). No statistically significant differences were observed between the $\log_{10}$ reductions between Plants A and B for the other viruses tested (*t*-test, $P > 0.05$).

4. DISCUSSION

The goals of this study were to assess the occurrence and relative abundance of viruses in wastewater and their removal by two types of wastewater treatment plants in an attempt to identify novel conservative viral indicators. We attempted to identify viruses that meet the essential criteria for viral indicators such as a high abundance, low removal, and little seasonal variation. It is hoped that these virus indicators may be used as model viruses for routine monitoring during advanced tertiary treatment processes such as soil aquifer treatment, reverse osmosis (RO), and other advanced oxidation processes (AOPs) prior to reuse applications.

One of the most significant findings of this study was that PMMoV and AiV may be useful viral indicators in water reclamation systems. They were constantly detected in both influent and effluent wastewaters at a relatively high concentration and showed no seasonal variation (Figure 1), suggesting a high abundance and persistence during wastewater treatment. AdV and EV have been proposed as indicators of human fecal contamination because of their high prevalence in sewage contaminated water (Albinana-Gimenez 2009, Hot 2003, Silva 2011). Our results, however, demonstrated that the concentration and reduction of AdV was more variable than PMMoV and AiV, and EV was less abundant than PMMoV and AiV.

PMMoV has been proposed as a novel viral indicator for fecal pollution in marine waters (Rosario 2009) and river water (Hamza 2011). PMMoV was also the most abundant in wastewater among the viruses we tested (Figure 1), which is in agreement with a previous study demonstrating that PMMoV is the most abundant virus in human feces when analyzed through a metagenomic survey of RNA (Zhang 2006). It has been reported that PMMoV is excreted in human feces at concentrations of 10^6 to 10^9 viruses per g (dry weight) (Zhang 2006). PMMoV is more abundant in wastewater than viruses that cause human disease, most likely because PMMoV in human feces is of dietary origin (from peppers and their processed products such as hot sauce and curry) and the virus is excreted from large healthy human populations. The present study provides additional evidence on the prevalence of PMMoV in wastewater and reclaimed water; this is the first study showing quantitative data on the seasonal occurrence and reduction of PMMoV during treatment at a WWTP. Although the behavior of PMMoV in the environment is not necessarily similar to that of enteric viruses because of differences in morphology between PMMoV (an extremely stable rod-shaped virion with a length of more than 300 nm) and enteric viruses (round-structured virions with a diameter of 30~90 nm), PMMoV appears to be a useful conservative “viral tracer” in wastewater reuse systems.

AiV belongs to the family *Picornaviridae*, which includes epidemiologically important enteric viruses such as EVs and hepatitis A viruses. The structural properties (e.g., size of virion, structures of capsid and genome) of AiV are also similar to those of other human enteric viruses. The prevalence of AiV in aquatic environments has been reported in previous studies that detected AiV in raw and treated sewage (Sdiri-Lourizi 2010, Kitajima 2011), sewage sludge (Bibby 2013),

biosolids (Bibby 2011), sewage-polluted river water (Alcara 2011), and shellfish (Hansman 2008, Le Guyader 2008, Sdiri-Lourizi 2010). We recently developed a RT-qPCR assay for the quantification of AiV genomes and reported quantitative data on the prevalence of AiV in wastewater in Japan (Kitajima 2013). In the present study, this RT-qPCR assay was utilized to investigate the prevalence and reduction of AiV at a WWTP. The data suggest that AiV is highly prevalent among humans throughout the year and is resistant to wastewater treatment. AiV may be a cause of human gastroenteritis, but the prevalence of clinical cases has not been widely studied in the United States. This is the first study that has quantitatively detected AiV in wastewater samples outside of Japan. Future efforts should focus on the environmental persistence of AiV using both RT-qPCR and an infectivity assay. AiV can be easily propagated and assayed with routine cell culture using Vero cells (Yamashita 1993), thereby greatly facilitating the designing of experiments aimed to determine the effectiveness of disinfectants such as chlorine, UV, and ozone in inactivating AiVs.

Of the human enteric viruses tested, the caliciviruses (NoV and SaV) are the leading cause of nonbacterial gastroenteritis in all age groups worldwide and have been reported to be more prevalent in the winter season in developed countries located in the temperate climate area (Green 2007, Siebenga 2010). Previous studies demonstrated that the concentration of GI and GII NoV increases in colder months (Katayama 2008, Haramoto 2006, Kitajima 2012, Pérez-Sautu 2012). In our results, their concentrations in wastewater varied over the year but were not higher during the winter months (Figure 1A and 1B). GIV norovirus, which has rarely been identified from environmental samples (Kitajima et al., 2009, 2010, 2011, La Rosa 2008), was

detected in wastewater samples collected from both Plants A and B, and the detection rate was consistent with a previous study investigating the presence of GIV NoV in wastewater in Japan (Kitajima et al. 2009). Although SaV has not been identified as often as NoV from gastroenteritis patients (Hansman 2007), it was detected at higher concentrations than NoV in the wastewater samples, indicating that SaV may be more prevalent in water environments than previously thought. Although the prevalence of GI and GII NoV and ARV in wastewater has been well studied, only a limited numbers of studies have investigated the seasonal occurrence of GIV NoV or SaV in wastewater (Haramoto 2008, Kitajima 2009, 2011, Sano 2011). This is the first study showing quantitative data on the occurrence of GIV NoV and SaV in the United States.

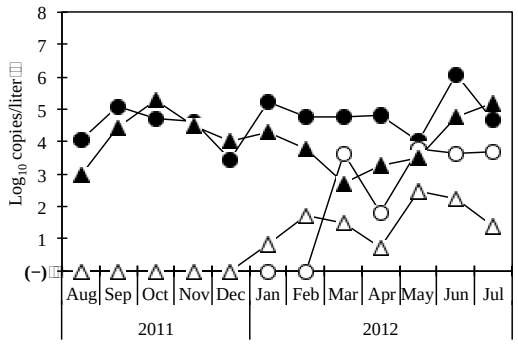
Although the removal efficiency of viruses by activated sludge process has been well studied (Haramoto 2006, Katayama 2008, Hata 2013), there have only been limited data on their removal by trickling filter wastewater treatment process (Ali 1997, Robertson 2000). We observed no statistically significant difference of $\log_{10}$ reduction (*t*-test, $P > 0.05$) between Plants A and B, which utilize activated sludge process and biological trickling filter, respectively, for all types of viruses except for SaV that showed a significantly higher $\log_{10}$ reduction at Plant A (*t*-test, $P = 0.042$). This observation suggests that activated sludge and trickling filter processes studied in the present work behave similarly with respect to virus removal.

In the present study, we quantified nine types of viruses in wastewater in Arizona; this is the first study quantitatively showing the seasonal occurrence of GIV NoV, SaV, AiV, and PMMoV in wastewater in the United States. We found that PMMoV

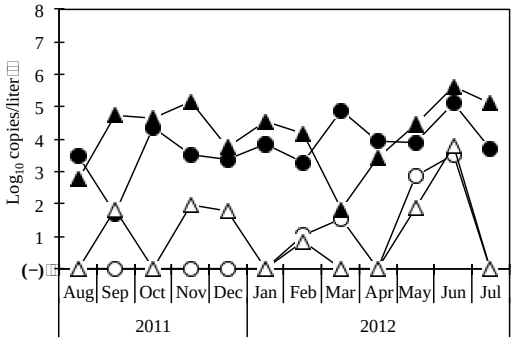
and AiV were constantly abundant in both influent and effluent wastewater, strongly suggesting that they are promising indicators for human enteric viruses and human fecal pollution in aquatic environments and for the performance efficacy of advanced water reclamation systems. These viruses were abundant year round in wastewater and showed less removal during wastewater treatment than other viruses. This more comprehensive analysis of the relative abundance, occurrence, and reduction of viruses in wastewaters may allow for the development of more conservative viral tracers and indicators to further ensure the microbial safety of wastewater reclamation systems.

FIGURES

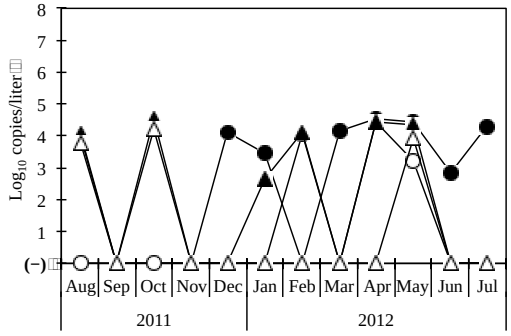
(A) GI NoV



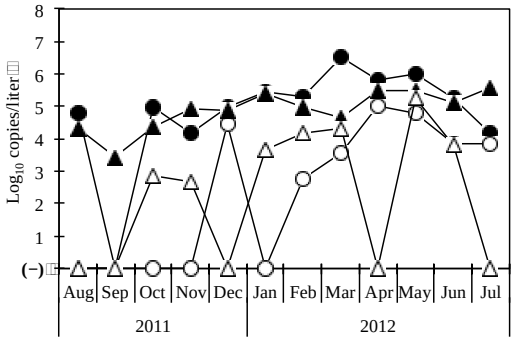
(B) GII NoV



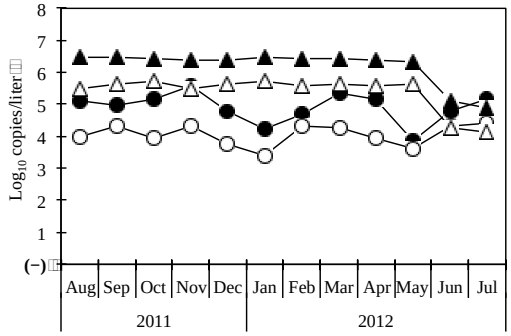
(C) GIV NoV



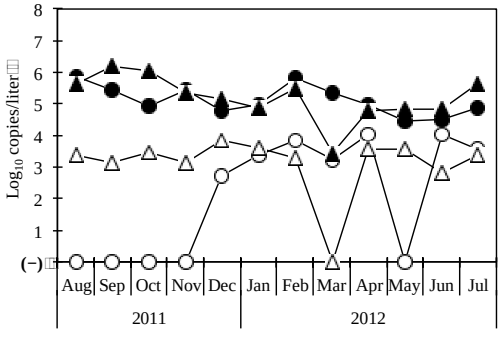
(D) SaV



(E) AiV



(F) EV



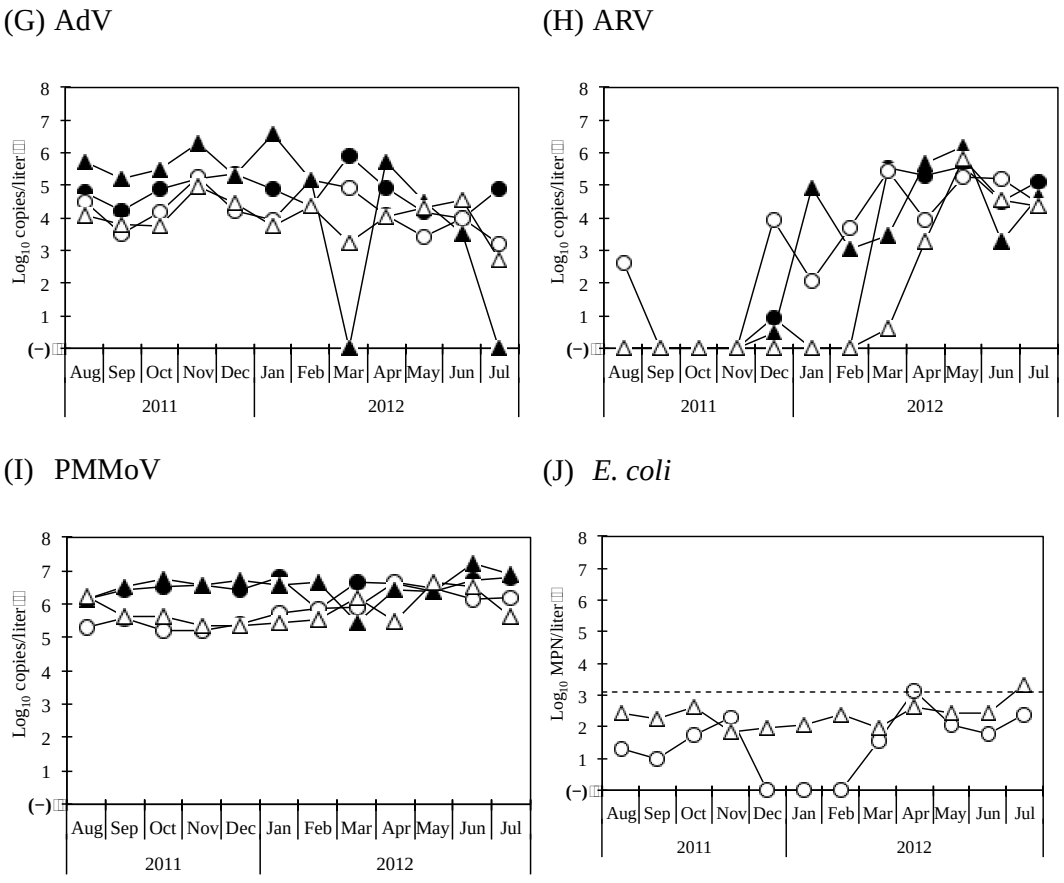


Figure 1. Concentration of viruses and *E. coli* in influent (●, Plant A; ▲, Plant B) and effluent (○, Plant A; △, Plant B) wastewater. *E. coli* was determined only for the effluent samples; the broken line indicates the criteria for *E. coli* in recreational water recommended by the US Environmental Protection Agency (i.e., 126 MPN/100 ml) (USEPA 1986). (−), not detected.

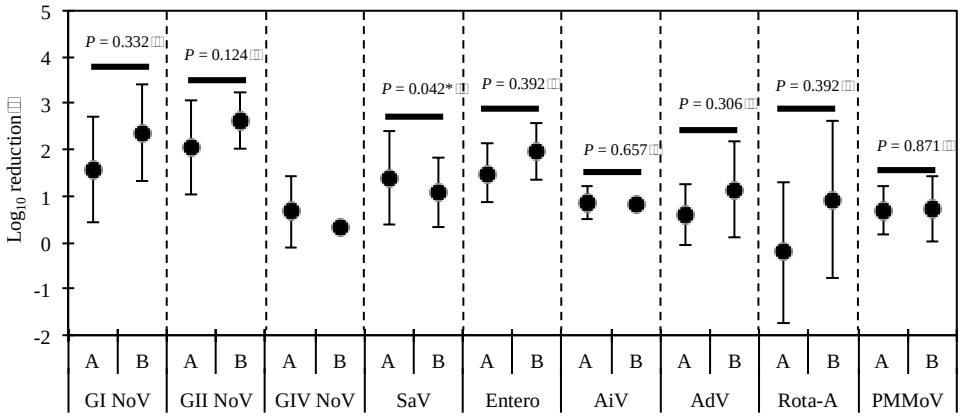


Figure 2. Reduction of viruses at two WWTPs (Plant A and B). The plots represent mean values, and error bars represent standard deviations. *P* values for the Student's *t*-test comparing $\log_{10}$ reductions at Plant A and that at Plant B are also presented; a value with statistical significant difference ($P < 0.05$) is indicated with an asterisk (*).

Table 1. Recovery efficiency of MNV spiked in wastewater concentrates.

WWTP	Sample	*n*	Geometric mean ± standard deviation
A	Influent	12	83.4 ± 14.1%
	Effluent	12	100.2 ± 47.7%
B	Influent	12	75.4 ± 18.6%
	Effluent	12	114.9 ± 69.6%

ACKNOWLEDGMENTS

The authors would like to thank Dr. Kelly Reynolds for her laboratory contributions and two anonymous wastewater treatment plants in southern Arizona for providing wastewater samples and *E. coli* data.

This study was supported by the National Science Foundation (NSF) Water and Environmental Technology (WET) Center and Water Resources Research Center (WRRC), The University of Arizona. We also wish to acknowledge the support of the Japan Society for the Promotion of Science (JSPS) to Masaaki Kitajima under the JSPS Postdoctoral Fellowships for Research Abroad.

REFERENCES

1. Levine, A. D.; Asano, T. Recovering sustainable water from wastewater. *Environ. Sci. Technol.* **2004**, *38* (11), 201A–208A; DOI 10.1021/es040504n.
2. Toze, S. Water reuse and health risks — real vs. perceived. *Desalination* **2006**, *187* (1–3), 41–51; DOI 10.1016/j.desal.2005.04.066.
3. Gerba, C. P.; Kitajima, M.; Iker, B. C. Viral presence in waste water and sewage and control methods. In *Viruses in Food and Water: Risks, Surveillance and Control*; N. Cook, Ed.; Woodhead Publishing Ltd.: Philadelphia 2013; pp 293–315.
4. Harwood, V. J.; Levine, A. D.; Scott, T. M.; Chivukula, V.; Lukasik, J.; Farrah, S. R.; Rose, J. B. Validity of the indicator organism paradigm for pathogen reduction in reclaimed water and public health protection. *Appl. Environ. Microbiol.* **2005**, *71* (6), 3163–3170; DOI 10.1128/AEM.71.6.3163.
5. Baggi, F.; Demarta, A.; Peduzzi, R. Persistence of viral pathogens and bacteriophages during sewage treatment: lack of correlation with indicator bacteria. *Res. Microbiol.* **2001**, *152* (8), 743–751; DOI 10.1016/S0923-2508(01)01255-4.
6. Gerba, C. P.; Goyal, S. M.; LaBelle, R. L.; Cech, I.; Bodgan, G. F. Failure of indicator bacteria to reflect the occurrence of enteroviruses in marine waters. *Am. J. Public Health* **1979**, *69* (11), 1116–1119.
7. AWPRC Study Group on Health Related Water Microbiology. Bacteriophages as model viruses in water quality control. *Water Res.* **1991**, *25* (5), 529–545. DOI 10.1016/0043-1354(91)90126-B.
8. Hot, D.; Legeay, O.; Jacques, J.; Gantzer, C.; Caudrelier, Y.; Guyard, K.; Lange,

M.; Andréoletti, L. Detection of somatic phages, infectious enteroviruses and enterovirus genomes as indicators of human enteric viral pollution in surface water. **2003**, *Water Res. 37* (19), 4703–4710. DOI 10.1016/S0043-1354(03)00439-1.

9. Silva, H. D.; García-Zapata M. T. A.; Anunciação, C. E. Why the Use of Adenoviruses as Water Quality Virologic Marker? *Food Environ. Virol.* **2011**, *3* (3–4), 138–140; DOI 10.1007/s12560-011-9069-2.
10. Albinana-Gimenez, N.; Miagostovich, M. P.; Calgua, B.; Huguet, J. M.; Matia, L.; Girones, R. Analysis of adenoviruses and polyomaviruses quantified by qPCR as indicators of water quality in source and drinking-water treatment plants. *Water Res.* **2009**, *43* (7), 2011–2019; DOI 10.1016/j.watres.2009.01.025.
11. Hamza, I. A.; Jurzik, L.; Uberla, K.; Wilhelm, M. Evaluation of pepper mild mottle virus, human picobirnavirus and Torque teno virus as indicators of fecal contamination in river water. *Water Res.* **2011**, *45* (3), 1358–1368; DOI 10.1016/j.watres.2010.10.021.
12. Girones, R.; Ferrús, M. A.; Alonso, J. L.; Rodriguez-Manzano, J.; Calgua, B.; Corrêa Ade, A.; Hundesa, A.; Carratala, A.; Bofill-Mas, S. Molecular detection of pathogens in water--the pros and cons of molecular techniques. *Water Res.* **2010**, *44* (15), 4325–4339; DOI 10.1016/j.watres.2010.06.030.
13. Rodriguez-Manzano, J., Alonso, J. L.; Ferrús, M. A.; Moreno, Y.; Amorós, I.; Calgua, B.; Hundesa, A.; Guerrero-Latorre, L.; Carratala, A.; Rusiñol, M.; Girones, R. Standard and new faecal indicators and pathogens in sewage treatment plants, microbiological parameters for improving the control of reclaimed water. *Water Sci. Technol.* **2012**, *66* (12), 2517–2523; DOI

10.2166/wst.2012.233.

14. American Public Health Association; American Water Works Association; Water Environment Federation. In *Standard Methods for the Examination of Water and Wastewater, 21st edn.*; Eaton, A. D. et al. Eds.; United Book Press: Baltimore 2005, p 9-72–9-74.
15. Haramoto, E.; Kitajima, M.; Kishida, N.; Katayama, H.; Asami, M.; Akiba, M. Occurrence of viruses and protozoa in drinking water sources of Japan and their relationship to indicator microorganisms. *Food Environ. Virol.* **2012**, *4* (3), 93–101; DOI 10.1007/s12560-012-9082-0.
16. Katayama, H.; Shimasaki, A.; Ohgaki, S. Development of a virus concentration method and its application to detection of enterovirus and Norwalk virus from coastal seawater. *Appl. Environ. Microbiol.* **2002**, *68* (3), 1033–1039; DOI 10.1128/AEM.68.3.1033-1039.2002.
17. Hata, A.; Kitajima, M.; Katayama, H. Occurrence and reduction of human viruses, F-specific RNA coliphage genogroups and microbial indicators at a full-scale wastewater treatment plant in Japan. *J. Appl. Microbiol.* **2013**, *114* (2), 545–554. DOI 10.1111/jam.12051.
18. Kitajima, M.; Oka, T.; Takagi, H.; Tohya, Y.; Katayama, H.; Takeda, N.; Katayama, K. Development and application of a broadly reactive real-time reverse transcription-PCR assay for detection of murine noroviruses. *J. Virol. Methods*. **2010**, *169* (2), 269–273; DOI 10.1016/j.jviromet.2010.07.018.
19. *EPA's Ambient Water Quality Criteria for Bacteria –1986*; EPA440/5-84-002; United States Environmental Protection Agency: Washington, DC, 1986. water.epa.gov/scitech/swguidance/standards/upload/2001_10_12_criteria_ambie

ntwqc_bacteria1986.pdf.

20. Katayama, H.; Haramoto, E.; Oguma, K.; Yamashita, H.; Tajima, A.; Nakajima, H.; Ohgaki, S. One-year monthly quantitative survey of noroviruses, enteroviruses, and adenoviruses in wastewater collected from six plants in Japan. *Water Res.* **2008**, *42* (6–7), 1441–1448; DOI 10.1016/j.watres.2007.10.029.
21. Green, K. Y. Caliciviridae: The noroviruses. In *Fields Virology, 5th ed.*; Knipe, D., Howley, P., Eds.; Lippincott Williams and Wilkins: Philadelphia, 2007; pp 949–979.
22. Siebenga, J. J.; et al. Norovirus epidemiology. In *Caliciviruses: Molecular and Cellular Virology*; Hansman, G. S., Jiang, X. J., Green, K. Y., Eds.; Caister Academic Press: Norfolk, UK, 2010; pp 1–24.
23. Haramoto, E.; Katayama, H.; Oguma, K.; Yamashita, H.; Tajima, A.; Nakajima, H.; Ohgaki, S. Seasonal profiles of human noroviruses and indicator bacteria in a wastewater treatment plant in Tokyo, Japan. *Water Sci. Technol.* **2006**, *54* (11-12), 301–308; DOI 10.2166/wst.2006.888.
24. Kitajima, M.; Haramoto, E.; Phanuwan, C.; Katayama, H.; Furumai, H. Molecular detection and genotyping of human noroviruses in influent and effluent water at a wastewater treatment plant in Japan. *J. Appl. Microbiol.* **2012**, *112* (3), 605–613; DOI 10.1111/j.1365-2672.2012.05231.x.
25. Pérez-Sautu, U.; Sano, D.; Guix, S.; Kasimir, G.; Pintó, R. M.; Bosch, A. Human norovirus occurrence and diversity in the Llobregat river catchment, Spain. *Environ. Microbiol.* **2012**, *14* (2), 494-502; DOI 10.1111/j.1462-2920.2011.02642.x.
26. Kitajima, M.; Haramoto, E.; Phanuwan, C.; Katayama, H.; Ohgaki S. Detection

of genogroup IV norovirus in wastewater and river water in Japan. *Lett Appl Microbiol.* **2009**, *49* (5), 655–658; DOI 10.1111/j.1472-765X.2009.02718.x.

27. Kitajima, M.; Oka, T.; Haramoto, E.; Takeda, N.; Katayama, K.; Katayama, H. Seasonal distribution and genetic diversity of genogroups I, II, and IV noroviruses in the Tamagawa River, Japan. *Environ. Sci. Technol.* **2010**, *44* (18), 7116–7122; DOI 10.1021/es100346a.
28. Kitajima, M.; Oka, T.; Haramoto, E.; Phanuwan, C.; Takeda, N.; Katayama, K.; Katayama, H. Genetic diversity of genogroup IV noroviruses in wastewater in Japan. *Lett. Appl. Microbiol.* **2011**, *52* (2), 181–184; DOI 10.1111/j.1472-765X.2010.02980.x.
29. La Rosa, G.; Pourshaban, M.; Iaconelli, M.; Muscillo, M. Detection of genogroup IV noroviruses in environmental and clinical samples and partial sequencing through rapid amplification of cDNA ends. *Arch. Virol.* **2008**, *153* (11):2077–2083. DOI 10.1007/s00705-008-0241-4.
30. Hansman, G. S.; Oka, T.; Katayama, K.; Takeda, N. Human sapoviruses: genetic diversity, recombination, and classification. *Rev. Med. Virol.* **2007**, *17* (2), 133–141; DOI 10.1002/rmv.
31. Haramoto, E.; Katayama, H.; Phanuwan, C.; Ohgaki, S. Quantitative detection of sapoviruses in wastewater and river water in Japan. *Lett. Appl. Microbiol.* **2008**, *46* (3), 408–413; DOI 10.1111/j.1472-765X.2008.02330.x.
32. Sano, D.; Pérez-Sautu, U.; Guix, S.; Pintó, R. M.; Miura, T.; Okabe, S.; Bosch, A. Quantification and genotyping of human sapoviruses in the Llobregat river catchment, Spain. *Appl. Environ. Microbiol.* **2011**, *77* (3), 1111–1114; DOI 10.1128/AEM.01721-10.

33. Rosario, K.; Symonds, E. M.; Sinigalliano, C.; Stewart, J.; Breitbart, M. Pepper mild mottle virus as an indicator of fecal pollution. *Appl. Environ. Microbiol.* **2009**, *75* (22), 7261–7267; DOI 10.1128/AEM.00410-09.

34. Zhang, T.; Breitbart, M.; Lee, WH.; Run J. Q.; Wei, C. L.; Soh, S. W.; Hibberd, M. L.; Liu, E. T.; Rohwer, F.; Ruan, Y. RNA viral community in human feces: prevalence of plant pathogenic viruses. *PLoS Biol.* **2006**, *4* (1), e3; DOI 10.1371/journal.pbio.0040003.

35. Sdiri-Loulizi, K.; Hassine, M.; Aouni, Z.; Gharbi-Khelifi, H.; Sakly, N.; Chouchane, S.; Guédiche, M. N.; Pothier, P.; Aouni, M.; Ambert-Balay, K. First molecular detection of Aichi virus in sewage and shellfish samples in the Monastir region of Tunisia. *Arch. Virol.* **2010**, 155 (9), 1509–1513; DOI 10.1007/s00705-010-0744-7.

36. Kitajima, M.; Haramoto, E.; Phanuwan, C.; Katayama, H. Prevalence and genetic diversity of Aichi viruses in wastewater and river water in Japan. *Appl. Environ. Microbiol.* **2011**, *77* (6), 2184–2187; DOI 10.1128/AEM.02328-10.

37. Bibby, K.; Viau, E.; Peccia J. Viral metagenome analysis to guide human pathogen monitoring in environmental samples. *Lett. Appl. Microbiol.* **2011**, *52* (4), 386–392; DOI 10.1111/j.1472-765X.2011.03014.x.

38. Alcalá, A.; Vizzi, E.; Rodríguez-Díaz, J.; Zambrano, J. L.; Betancourt, W.; Liprandi, F. Molecular detection and characterization of Aichi viruses in sewage-polluted waters of Venezuela. *Appl. Environ. Microbiol.* **2010**, *76* (12), 4113–4115; DOI 10.1128/AEM.00501-10.

39. Hansman, G. S.; Oka, T.; Li, T. C.; Nishio, O.; Noda, M.; Takeda, N. Detection of human enteric viruses in Japanese clams. *J. Food. Prot.* **2008**, *71* (8), 1689–

1695;

40. Le Guyader, F. S.; Le Saux, J. C.; Ambert-Balay, K.; Krol, J.; Serais, O.; Parnaudeau, S.; Giraudon, H.; Delmas, G.; Pommepuy, M.; Pothier, P.; Atmar, R. L. Aichi virus, norovirus, astrovirus, enterovirus, and rotavirus involved in clinical cases from a French oyster-related gastroenteritis outbreak. *J. Clin. Microbiol.* **2008**, *46* (12), 4011–4017; DOI 10.1128/JCM.01044-08.
41. Yamashita, T.; Sakae, K.; Ishihara, Y.; Isomura, S.; Utagawa, E. Prevalence of newly isolated, cytopathic small round virus (Aichi strain) in Japan. *J. Clin. Microbiol.* **1993**, *31* (11), 2938–2943.
42. Ali, M. A.; El-Senousy, W. M.; El-Hawaary, S. E. Enteroviruses in sewage: comparison of different technologies for wastewater treatment and reuse. *J. Egypt. Public Health Assoc.* **1997**, *72* (5-6), 441-456.

APPENDIX C

EVALUATION OF COMMERCIAL KITS FOR THE EXTRACTION AND PURIFICATION OF VIRAL NUCLEIC ACIDS FROM ENVIRONMENTAL AND FECAL SAMPLES

Brandon C. Iker [a], Kelly R. Bright [a], Ian L. Pepper [a], Charles P. Gerba [a], Masaaki Kitajima [a,] *

[a] *Department of Soil, Water and Environmental Science, The University of Arizona, Tucson, Arizona 85721, USA*

*Corresponding author. Tel.: +1-520-626-8094; Fax: +1-520-621-6163

E-mail address: kitajima@email.arizona.edu (Masaaki Kitajima)

Running title: Comparison of viral nucleic acid extraction kits

Published: Journal of Virological Methods

ABSTRACT

The extraction and purification of nucleic acids is a critical step in the molecular detection of enteric viruses from environmental or fecal samples. In the present study, we assessed the performance of three commercially available kits, the MO BIO PowerVirus Environmental DNA/RNA Extraction kit, the Qiagen QIAamp Viral RNA Mini kit, and the Zymo ZR Virus DNA/RNA Extraction kit. Viral particles of adenovirus 2 (AdV), murine norovirus (MNV), and poliovirus type 1 (PV1) spiked in molecular grade water and three different types of sample matrices (i.e., biosolids, feces, and surface water concentrates) were extracted with the kits, and the yields of the nucleic acids were determined by quantitative PCR (qPCR). The MO BIO kit performed the best with the biosolids, which were considered to contain the highest level of inhibitors, and provided the most consistent detection of spiked virus from all of the samples. A qPCR inhibition test using an internal control plasmid DNA and a nucleic acid purity test using an absorbance at 230 nm for the nucleic acid extracts demonstrated that the MO BIO kit was able to remove qPCR inhibitors more effectively than the Qiagen and Zymo kits. Our results suggest that the MO BIO kit is appropriate for the extraction and purification of viral nucleic acids from environmental and clinical samples that contain high levels of inhibitors.

Key words: Virus; Nucleic acid; Quantitative PCR; Inhibitor; Extraction; Purification

1. INTRODUCTION

Molecular detection techniques have become an increasingly effective means for the rapid and sensitive detection of fecal pollution in the environment (Abbaszadegan et al., 1999; Bosch et al., 2008; Lemerchand et al., 2004; Rodríguez et al., 2009; Wong et al., 2012). Specifically, enteric viruses have been recommended as a molecular marker for water quality and are considered to be important molecular targets for protecting public health and improving risk assessment models, as compared to more traditional indicator organisms such as bacteriophages and *Escherichia coli* (Harwood et al., 2005). However, there are major challenges for the detection of viruses present at low concentrations in environmental samples. In such situations, virus particles must be first concentrated in order to effectively assay them (Cashdollar et al., 2013; Ikner et al., 2012). Although quantitative polymerase chain reaction (qPCR) detection cannot distinguish between infective and non-infective viruses in a sample, the use of rapid molecular assays provides researchers a highly sensitive prescreening tool prior to the more costly and time-consuming cell culture assays that are essential for exposure assessments of infectious virus.

Molecular methods for virus detection are often not reliable because of the presence of interfering substances such as humic and fulvic acids, RNases and DNases, and other polymerase enzyme inhibiting compounds, which can produce false negative results (Abbaszadegan et al., 1993; Gibson et al., 2012; Griffin et al., 2003; Hata et al., 2011; Rock et al., 2010; Wyn-Jones et al., 2011). The co-concentration of PCR inhibitors from virus concentrates and extracts and the subsequent need for dilution or removal of PCR inhibitors has been one of the more difficult challenges to overcome when trying to detect viral pathogens within

environmental samples (Borchardt et al., 2004; Gibson et al., 2012; Hata et al., 2011, Rock et al., 2010). Under optimal reaction conditions (i.e., no inhibitors present), qPCR can reliably detect as few as 10 gene copies per PCR reaction; however, the presence of PCR inhibitory substances in the sample can greatly reduce the sensitivity of detection.

In addition to environmental inhibitors, some virus concentration and extraction methods require the use of elution buffers such as beef extract that also contain inhibitory compounds. One approach to mitigate such PCR inhibition is to dilute raw nucleic acid extracts by 1:10 or greater, thereby significantly decreasing the concentration of inhibitors, but also the likelihood of detection, since nucleic acid concentrations are also decreased (Gibson et al., 2012; Moreira, 1998; Wilson et al., 1997). Additionally, the use of pre- or post-nucleic acid extraction procedures to remove or mitigate PCR inhibitors may further degrade viral genomes. Such additional procedures also potentially provide more opportunities for laboratory contamination and increase the overall cost and time of the assays as well.

Environmental virus concentration and extraction methods commonly cite the use of a commercially available virus extraction kit to extract viral nucleic acids in favor of in-house extraction methods that lack reproducibility and quality assurance from lab to lab. Many extraction kits currently used for viral nucleic acid isolation were originally developed for use in a clinical setting according to manufacturers' recommendations and provide excellent virus recovery for samples low in inhibitory compounds. An ideal environmental virus extraction kit should provide similar recoveries to established extraction methods and kits, while simultaneously providing

effective inhibitor removal to provide the cleanest, highest quality nucleic acid extracts for subsequent molecular detection and quantification.

In the present study, we assessed the performance of a newly developed commercially available kit, the MO BIO PowerVirus DNA/RNA Extraction Kit (MO BIO, San Diego, CA) in comparison to two other commercially available kits, the QIAamp Viral RNA Mini Kit (Qiagen, Hilden, Germany) and the ZR Virus DNA/RNA Extraction Kit (Zymo Research, Irvine, CA), for the molecular detection of adenovirus (AdV), murine norovirus (MNV), and poliovirus type 1 (PV1) in three different sample matrices commonly found to contain high concentrations of PCR inhibitors. Kit comparisons were evaluated on their relative recovery and quantification of viral particles from spiked environmental sample matrices when compared to a spiked molecular grade water control. The removal of PCR inhibitors was assessed with the use of an internal MNV plasmid (pMNV) control to estimate the level of inhibition found in each nucleic acid extraction. In addition, an absorbance of 230 nm by NanoDrop (Nano-drop Technologies, Wilmington, DE) was used to estimate humic-like substances present in the nucleic acid extracts.

2. MATERIALS AND METHODS

2.1. Viral stock preparation

AdV (type 2, ATCC VR-846) was propagated on human primary liver carcinoma (PLC/PRF/5) cells (ATCC CRL-8024) to obtain an initial stock concentration of approximately 10^5 50 % tissue culture infectious dose ($TCID_{50}$)/mL. PV1 (LSc 2ab Sabin strain) was obtained from the Baylor College of Medicine, and propagated on buffalo green monkey (BGM) kidney cells (provided by D. Dahling from the U.S. Environmental Protection Agency) to obtain an initial stock concentration of approximately 10^8 plaque forming units (PFU)/mL. MNV (S7-PP3 strain, isolated in Japan) was kindly provided by Y. Tohya (Nihon University, Kanagawa, Japan) and propagated on RAW 264.7 (ATCC TIB-71) cells to obtain an initial stock concentration of approximately 10^6 PFU/mL. All stock cultures were stored at −80°C.

2.2. Environmental and fecal sample preparations

2.2.1. Biosolid samples

Class A biosolid samples were prepared according to method ASTM D 4994-89 (**American Society for Testing and Materials, 2002)**. Briefly, beef extract was added to 10-20 g (dry weight) biosolids and stirred for 30 min to elute the viruses. The solids were then pelleted by centrifugation at 2,500 × g for 15 min; the supernatant was then flocculated by adjusting the pH to 3.5 and re-centrifuged at 10,000 × g for 30 min to form a pellet. The pellet was then dissolved in phosphate-buffered saline (PBS; pH 7.4; Sigma-Aldrich, St. Louis, MO) and filtered through a 0.22-μm pore size filter (Millex; Millipore, Bedford, MA). The final eluates were stored at −80°C.

2.2.2. Surface water samples

Surface river water concentrates from Oak Creek Canyon, Arizona of ~400 L were previously concentrated using a 1MDS filter (Cuno, Meriden, CT) and eluted with 3% beef extract followed by secondary concentration via organic flocculation (Sobsey and Jones 1979). Samples were kept at −80°C for long-term storage.

2.2.3. Fecal samples

For the assessment of viral detection directly from human feces, three fecal samples were selected from laboratory archives and prepared by suspending 1 g (wet weight) in 9 mL of sterile PBS and centrifugation at 2,500 × g for 15 minutes to pellet the fecal solids. The supernatants were decanted in a fresh 15 mL conical tube and stored at −20°C for later use.

2.3. Sample spiking and viral nucleic acid extraction

The characteristics of each kit are shown in Table 1. Molecular grade water was used as our spiked “pure” water control in order to estimate the virus recovery of each kit without the effects of any environmental parameters and to provide a baseline for the qPCR enumeration of viral genomes for each virus type.

For the control and sample matrix preparation, 1940 µl of each sample was spiked with 20 µl each of the three virus stocks described above and vortexed for 5 sec. This volume was then divided into 200-µl aliquots following the manufacturers’ suggested sampling volumes of 200 µl for the MO BIO and Zymo kits, and 140 µl for the Qiagen kit. All samples were stored at −80°C until needed.

For each round of extractions, the aliquots of spiked sample matrices and control samples were brought to room temperature and extracted following the manufacturers’ instructions. Nucleic acids were eluted in 100 µl of molecular grade

water for the MO BIO and Zymo kits, whereas 60 µl of elution buffer AVE was used for the Qiagen kit, according to the manufacturers' instructions. Nucleic acid extracts were frozen at −20°C for 24 h prior to quantification by qPCR.

2.4. Determination of nucleic acid purity

A NanoDrop ND-1000 spectrophotometer (Nano-drop Technologies, Wilmington, DE) was used to estimate the level of humic acid-like substances remaining in the nucleic acid extracts as a relative measure of potential PCR inhibitors using an absorbance of 230 nm. This wavelength was used over the more traditional 230/260 ratio because samples found to be very low in DNA concentrations can be skewed by the 230/260 nm ratio, making comparisons of DNA purity difficult.

2.5. Quantification of viral genomes and MNV plasmid control by qPCR

2.5.1. Reverse transcription

For the detection of the PV1 and MNV genomes, reverse transcription (RT) was performed using the High Capacity cDNA Reverse Transcription Kit with RNase inhibitors (Applied Biosystems, Foster City, CA). Briefly, 20 µl of undiluted RNA/DNA extract was added to 20 µl of RT mixture containing 4 µl of 10× reverse transcription buffer, 1.6 µl of 25× dNTPs, 4 µl of 10× random hexamers, 100 units of MultiScribe™ reverse transcriptase (Applied Biosystems), and 40 units of RNase inhibitor. The RT reaction mixture was incubated at 25°C for 10 min, 37°C for 120 min, and finally 85°C for 5 min to inactivate the enzyme.

2.5.2. Preparation of the standard plasmids

The plasmid standards for the qPCR assay for MNV, PV1, and AdV were prepared as previously described (Kitajima et al., 2010). Briefly, partial genomic

regions of MNV, PV1, and AdV that encompass the qPCR targets were amplified by (RT-)PCR, and the PCR products were cloned into the TOPO vectors (Invitrogen, Carlsbad, CA, USA). The plasmid concentration (ng/µl) was determined by measuring the optical density at 260 nm using a NanoDrop ND 1000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA) and the copy numbers of the plasmid DNA molecule were calculated.

2.5.3. qPCR

For the quantification of the viral genomes and plasmid controls, TaqMan-based qPCR assays were performed in 25 µl reaction volumes containing 2.5 µl of template (DNA or cDNA), 12.5 µl of iQ Supermix (Bio-Rad Laboratories, Hercules, CA), a primer set, and a probe. The sequences of primers and probes were derived from previous studies (Table 2). The PCR amplification was performed with the iQ5 Real-time PCR Detection System (Bio-Rad Laboratories) and amplification data were collected and analyzed with the iQ5 Optical System Software Version 2.1 (Bio-Rad Laboratories). Serial tenfold dilutions of the standard plasmid DNA containing inserts of the amplification region were used to generate a standard curve; thus, quantitative data on the DNA or cDNA copy numbers were obtained. All qPCR reactions were performed in duplicate; namely, two PCR tubes were used for all samples and standards and the average copy of the numbers obtained from the two tubes were used for subsequent calculations. Positive and negative controls were included in the qPCR reaction plates to ensure that false-negative and -positive results were avoided.

2.6. pMNV internal control

In order to estimate the level of PCR inhibition still present in each extract, a pMNV control was spiked at 10^4 plasmid copies per PCR tube along with 2.5 µl of

each raw nucleic acid extract and quantified under the conditions described above and in Table 2. The primer and probe sets used for the quantification of MNV and pMNV were identical. The pMNV quantification was performed in the presence of the raw nucleic acid extract that had not undergone the RT step, therefore allowing us to estimate the level of inhibition using the same molecular targets.

3. RESULTS

3.1. Extraction efficiency of viruses spiked in deionized water

In order to directly compare the virus extraction efficiency of each kit under optimum conditions, AdV, MNV, and PV1 were spiked into molecular grade water and the relative extraction efficiencies were determined by qPCR. For all three viruses tested, the Qiagen and Zymo kits provided the highest and lowest relative extraction efficiencies, respectively (Figure 1A). Less than a 1-$\log_{10}$ difference in detection could be observed between all three kits for AdV and PV1, while the MNV data set suggests that the Qiagen kit had a better recovery of MNV in the control than either of the other two kits.

3.2. Detection of viruses spiked in sample matrices

To estimate the extraction/qPCR efficiencies of AdV, MNV, and PV1 in environmental sample matrices by each extraction kit, virus particles were spiked into three biosolid extracts, three fecal suspensions, and three 1MDS surface water concentrates and quantified under the conditions described previously.

The biosolid extracts presumably had the highest level of inhibitors present and were therefore expected to be the most challenging sample matrices to analyze with respect to molecular detection and the effective removal of PCR inhibitors (Figure 1B). For the biosolid extracts, the Zymo kit had the poorest performance with no observed detection for all three viruses tested. The Qiagen kit was unable to provide a signal for AdV with no detection in the biosolid extracts, but provided a similar qPCR signal to that of the MO BIO kit for the two RNA viruses. The MO BIO kit had the best overall performance with the biosolid extracts, with similar

extraction/qPCR efficiencies to the control samples for all three viruses (Figure 1A). This was particularly true for AdV, the only DNA virus included in the study.

The recovery and detection of viral genomes from AdV, MNV, and PV1 in the three human fecal suspensions was highly variable for the Zymo and Qiagen kits (Figure 1C). This was most likely due to the fecal samples being from different sources. In addition, these two extraction kits exhibited a significant (~3-$\log_{10}$ or greater) loss in the extraction/qPCR efficiency from fecal samples for AdV.

All of the kits performed well for the recovery and detection of viral genomes from 1MDS surface water concentrates with little variation in the extraction/PCR detection efficiencies (<1-$\log_{10}$ difference in the virus detection between each of the kits) (Figure 1D). Overall, the MO BIO kit provided the most consistent detection of spiked virus from all of the samples, with the only real variability observed between the fecal suspension samples with AdV (Figure 1C).

3.3. Assessment of qPCR inhibition

The integrated effects of the nucleic acid extraction, the RT, and the qPCR efficiencies most likely affect the overall detection efficiency of the viruses spiked in the various sample matrices. To directly assess the effects of potential qPCR inhibition, control qPCR reactions targeting the pMNV (~10^{-5} copies/PCR tube) were performed in the presence of each nucleic acid extract (from the samples shown in Figure 1). We observed a >5-$\log_{10}$ loss in the detection in all three biosolid extract samples and in one of the fecal samples (sample number 1) for both the Zymo and Qiagen kit extracts (Figure 2A). This indicates the likelihood of a substantial amount of PCR inhibitors in these extracts. No substantial decrease in plasmid detection was observed with the MO BIO sample set in comparison to the positive control with ~10^{5}

copies per PCR tube, suggesting that the MO BIO kit effectively removed qPCR inhibitors from the virus concentrates.

In order to estimate the purity of the nucleic acid extracts from each kit, we measured the ultraviolet absorbance at a wavelength of 230 nm. Each sample was blanked using the suggested nucleic acid elution buffer (i.e., molecular grade water for the MO BIO and Zymo kit extracts and buffer AVE for the Qiagen kit extracts). Traditionally, the reading of 260/230 nm is taken as a ratio of DNA (260 nm) and humic-like substances (230 nm); however, according to the manufacturer (Nano-drop Technologies), this ratio can be skewed in samples with very low DNA concentrations. We therefore applied only the 230 nm reading as a rough estimate of the level of potential inhibitors (e.g., humic-like substances) present in our nucleic acid extracts. A relatively high absorbance could be observed in both the Qiagen and Zymo extracts, particularly for the biosolid samples and fecal sample number 1, whereas the MO BIO kit extracts had a low absorbance of less than 1.0 for all samples (Figure 2B). These data indicate that the Qiagen and Zymo kit extracts still contain qPCR inhibitors, which is in agreement with the results of the qPCR inhibition test (Figure 2A). Interestingly, though the amount of humic-like substances were fairly low for biosolid samples 1 and 2 for all three kits (similar for instance to the surface water extracts), there was still significant qPCR inhibition observed for these samples with the Zymo and Qiagen extracts, but not for the MO BIO extracts. This suggests that the MO BIO kit is also more effective at removing other PCR inhibitors that are not detected at an absorbance of 230 nm.

3.4 Effects of beta mercaptoethanol (BME) on qPCR detection

Both the MO BIO kit and Zymo kit protocols suggest the use of the addition of beta mercaptoethanol (BME) (at 1% and 0.5%, respectively) to the lysis buffer during the sample preparation as a means of inactivating potential RNases in the samples. We were curious as to the effects of this compound on the qPCR detection in our spiked sample matrices. Each extraction was performed with and without BME for each of our samples using these two kits. No significant improvement in detection could be observed with the addition of BME with either extraction kit (data not shown).

3.5 Effects of bead beating step on virus recovery

The MO BIO PowerVirus Kit employs the use of a 10-minute mechanical glass bead lysis step. A direct comparison of spiked AdV control samples was included in triplicate performed with and without this bead-beating step. There was an approximately 30% better recovery (data not shown) of spiked AdV genomes obtained when the bead beating step was omitted.

4. DISCUSSION

The mitigation of qPCR inhibition for the detection of viral genomes in environmental sample matrices is essential for improving virus detection methods. Ideally, a good extraction kit will provide the user with consistent recovery of both viral DNA and RNA genomes, will be applicable to a wide variety of environmental sample matrices, will have efficient removal of inhibitory substances that are widely present in environmental matrices, and will be cost-efficient and not overly time-consuming. A good extraction kit should also preferentially be scalable from low to high volume workflows. In the current study, the MO BIO PowerViral Environmental

DNA/RNA Extraction Kit with its integrated Inhibitor Removal Technology (IRT) appears to be an effective means of removing inhibitors while simultaneously providing pure and concentrated viral DNA and RNA for subsequent molecular analyses that provides comparable results to established virus extraction kits. In general, the MO BIO extraction kit provided enhanced PCR inhibitor removal from the biosolid and fecal samples when compared to the Qiagen and Zymo kits, respectively, for MNV, PV1, and AdV (Figure 1). However, for the surface water concentrates used in this study, there was no observable difference for the recovery of viral genomes between the three kits. Thus, if the environmental sample is relatively clean with respect to PCR inhibitory substances, then any of the three kits would be suitable for effective nucleic acid extraction.

We investigated the relevance of the addition of BME as an optional RNase inhibitor in the MO BIO and Zymo protocols. Under our experimental conditions, the BME did not significantly improve the detection and quantification of viral genomes. Nonetheless, we theorize that the use of BME may play a more substantial role in the long-term storage of nucleic acid extracts.

We also looked at the effects of bead beating on virus recovery in spiked viral control samples in an attempt to identify the cause of the lower recovery in the MO BIO extractions in comparison to those of the Qiagen kit. This was accomplished by evaluating the recovery of AdV with and without the bead-beating step. The bead beating resulted in approximately a 30% lower recovery than the extractions that did not include this step. This loss in recovery when using the beads is most likely due to the loss of the lysate left behind in the bead beating tube during the transfer of supernatant to the next step and not due to genomic shearing. The MO BIO kit

requires 200 µl of the sample to be added to 600 µlof the extraction buffer to bring the total volume of the lysate to 800 µl; however, after bead beating and subsequent centrifugation of beads and debris, only approximately 600 µl of lysate was recovered from the supernatant. We believe that although the bead beating step may therefore result in a loss of a portion of the lysate, it may still be a necessary step to ensure the complete lysis of viral capsids in complex, hard-to-amplify sample matrices. This step may possibly be omitted when recovering nucleic acid from "clean" water matrices; however, similar recoveries were observed for all three viruses in spiked surface water concentrates, suggesting that the relatively small loss caused by the beads does not significantly decrease virus quantification in more realistic sample matrices. This phenomenon is not well understood but we theorize that the loss of virus particles spiked into molecular grade water (i.e., under non-buffering conditions) might also be attributed to virus adsorption to the microcentrifuge tubes during the extraction process (Patel et al., 2007).

The advantages of this study over previous kit assessments are that multiple virus models including both DNA and RNA viruses were evaluated in order to estimate the virus extraction efficiencies and qPCR detection from three sample matrices. To our knowledge, the MO BIO PowerViral Environmental DNA/RNA Extraction Kit is the first virus extraction kit that is specifically designed for the isolation of viral nucleic acids from environmental sample matrices that are commonly found to be high in PCR inhibitors. Although not included in the current study, other difficult sample matrices such as oyster guts, meats, and leafy greens may be more thoroughly surveyed using extraction methodologies that integrate inhibitor removal technologies into their workflows. Additionally, companies like MO BIO and

Qiagen, have begun to integrate extraction formats into 96-well kits that are optimized for robot assisted work flows, decreasing the turnaround time, the consumable costs, and the labor requirements as regulatory agencies and related industries begin to look at the feasibility of such widespread surveillance. Both domestic and imported food items, irrigation waters, and biofilms, sediment samples, composts, and manures will need to be more thoroughly surveyed for the presence of environmental and fecal contamination and associated pathogens. Currently, traditional culture-based detection methods for pathogens and indicators like *Escherichia coli* and other enteric viruses take days to weeks to complete, are prohibitively costly to perform on a large scale, and are too exceedingly labor intensive to be an effective tools for widespread surveillance and should remain as a confirmatory steps once such contamination is identified using more sensitive, rapid, and cost effective molecular detection methods.

In summary, extraction methods that effectively remove PCR inhibitors can provide investigators with an effective tool that allows for the molecular detection and quantification of low levels of viruses in environmental matrices that routinely contain PCR inhibitors. Despite the success of the MO BIO PowerViral Environmental DNA/RNA Extraction Kit in providing high quality nucleic acid extracts for the detection of viral pathogens by qPCR, the effectiveness of this kit with additional sample matrices needs to be further evaluated in order to better assess the kit's overall performance in recovering nucleic acids from highly variable environmental samples.

Acknowledgements

The authors would like to thank The University of Arizona National Science Foundation Water and Environmental Technology (WET) Center and Water Resources Research Center for their financial support and Dr. Kelly Reynolds for her laboratory contributions. We also wish to acknowledge the support of the Japan Society for the Promotion of Science (JSPS) to Masaaki Kitajima, under the JSPS Postdoctoral Fellowships for Research Abroad.

REFERENCES

Abbaszadegan, M., Huber, M.S., Gerba, C.P., Pepper, I.L., 1993. Detection of enteroviruses in groundwater with the polymerase chain reaction. Appl. Environ. Microbiol. 59, 1318–1324.

Abbaszadegan, M., Stewart, P., LeChevallier, M., 1999. A strategy for detection of viruses in groundwater by PCR. Appl. Environ. Microbiol. 65, 444–449.

American Society for Testing and Materials. 2002. Standard practice for recovery of viruses from wastewater sludges, ASTM D4994-89. **American Society for Testing and Materials**, West Conshohocken, Pennsylvania.

Borchardt, M.A., Haas, N.L., Hunt, R.J., 2004. Vulnerability of drinking-water wells in La Crosse, Wisconsin, to enteric-virus contamination from surface water contributions. Appl. Environ. Microbiol. 70, 5937–5946.

Bosch, A., Guix, S., Sano, D., Pinto, R.M., 2008. New tools for the study and direct surveillance of viral pathogens in water. Curr. Opin. Biotechnol. 19, 295–301.

Cashdollar, J.L., Brinkman, N.E., Griffin, S.M., McMinn, B.R., Rhodes, E.R., Varughese, E.A., Grimm, A.C., Parshionikar, S.U., Wymer, L., Fout, G.S., 2013. Development and evaluation of EPA Method 1615 for detection of enterovirus and norovirus in water. Appl. Environ. Microbiol. 79, 215–223.

Gibson, K.E., Schwab, K.J., Spencer, S.K., Borchardt, M.A., 2012. Measuring and mitigating inhibition during quantitative real time PCR analysis of viral nucleic acid extracts from large-volume environmental water samples. Water Res. 46, 4281–4291.

Gregory, J.B., Litaker, R.W., Noble, R.T., 2006. Rapid one-step quantitative reverse transcriptase PCR assay with competitive internal positive control for detection of

enteroviruses in environmental samples. Appl. Environ. Microbiol. 72, 3960–3967.

Griffin, E.W., Donaldson, K.A., Paul, J.H., Rose, J.B., 2003. Pathogenic human viruses in coastal waters. Clin. Microbiol. Rev. 16, 129–143.

Harwood, V.J., Levine, A.D., Scott, T.M., Chivukula, V., Lukasik, J., Farrah, S.R., Rose, J.B., 2005. Validity of the indicator organism paradigm for pathogen reduction in reclaimed water and public health protection. *Appl. Environ. Microbiol.* 71, 3163–3170.

Hata, A., Katayama, H., Kitajima, M., Visvanathan, C., Nol, C., Furumai, H., 2011. Validation of internal controls for extraction and amplification of nucleic acids from enteric viruses in water samples. Appl. Environ. Microbiol. 77, 4336–4343.

Heim, A., Ebnet, C., Harste, G., Pring-Akerblom, P., 2003. Rapid and quantitative detection of human adenovirus DNA by real-time PCR. J. Med. Virol. 70, 228–239.

Ikner, L.A., Gerba, C.P., Bright, K.R., 2012. Concentration and recovery of viruses from water: a comprehensive review. Food Environ. Virol. 4, 41–67.

Kitajima, M., Oka, T., Takagi, H., Tohya, Y., Katayama, H., Takeda, N., Katayama, K., 2010. Development and application of a broadly reactive real-time reverse transcription-PCR assay for detection of murine noroviruses. J. Virol. Methods 169, 269–273.

Lemarchand, K., Masson, L., Brousseau, R., 2004. Molecular biology and DNA microarray technology for microbial quality monitoring of water. Crit. Rev. Microbiol. 30, 145–172.

Moreira, D., 1998. Efficient removal of PCR inhibitors using agarose-embedded DNA preparations. Nucleic Acids Res. 26, 3309–3310.

Patel, A., Noble, R.T., Joshua, A.S., Schwalbach, M.S., Hewson, I., and Fuhrman, J.A., 2007. Virus and prokaryote enumeration from planktonic aquatic environments by epiflourescence microscopy with SYBR Green I. Nat. Protoc. 2, 269-276.

Rock, C., Alum A., Abbaszadegan, M., 2010. PCR inhibitor levels in concentrates of biosolid samples predicted by a new method based on excitation-emission matrix spectroscopy. Appl. Eviron. Microbiol. 76, 102–109.

Rodríguez, R.A., Pepper, I.L., Gerba, C.P., 2009. Application of PCR-based methods to assess the infectivity of enteric viruses in environmental samples. Appl. Environ. Microbiol. 75, 297–307.

Sobsey, M.D., Jones, B.L., 1979. Concentration of poliovirus from tap water using positively charged microporous filters. Appl. Environ. Microbiol. 37, 588–595.

Wilson, I.G., 1997. Inhibition and facilitation of nucleic acid amplification. Appl. Environ. Microbiol. 63, 3741–3751.

Wong, K., Fong, T.T., Bibby, K., Molina, M., 2012. Application of enteric viruses for fecal pollution sources tracking in environmental waters. Environ. Int. 45, 151–164

Wyn-Jones, A.P., Carducci, A., Cook, N., D'Agostino, M., Divizia, M., Fleischer, J., Gantzer, C., Gawler, A., Girones, R., Höller, C., de Roda Husman, A.M., Kay, D., Kozyra, I., López-Pila, J., Muscillo, M., Nascimento, M.S., Papageorgiou, G., Rutjes, S., Sellwood, J., Szewzyk, R., Wyer, M., 2011. Surveillance of

adenoviruses and noroviruses in European recreational waters. Water Res. 45, 1025–1038.

TABLES

Table 1. Characteristics of virus nucleic acid extraction kits evaluated in this study.

Extraction kit	Manufacturer	Cost per unit (U.S.)[a]	Loading volume (µl)	Extraction principle	Final extraction volume (µl)	Approximate processing time (min)[d]	Additional reagents	Additional equipment
PowerViral Environmental DNA/RNA Extraction Kit	MO BIO	$6.00	□200	Bead beating[bc] Column	50~100	40	Ethanol, BME	Bead beater[c]
ZR Viral DNA/RNA Kit	Zymo Research	$5.16	□200	Column	50~100	20	Ethanol, BME	None
QIAamp Viral RNA Mini Kit	Qiagen	$4.40	□140	Column	60	30	Ethanol	None

[a] Approximate free market price.

[b] Mechanical lysis.

[c] Optional.

[d] Per sample.

Table 2. Primers and probes used for qPCR assays for the detection and quantification of spiked adenovirus (AdV), poliovirus type 1 (PV1), and murine norovirus (MNV) genomes, and murine norovirus plasmid (pMNV).

Target	Function	Name	Sequence (5'→3')	Product size (bp)	Reference
AdV	Forward primer	AQ2	GCCCCAGTGGTCTTACATGCACATC	132	Heim et al. 2003
	Reverse primer	AQ1	GCCACGGTGGGGTTTCTAAACTT		
	Probe	AP	FAM-TGCACCAGACCCGGGCTCAGGTACTCCGA-BHQ1		
PV1	Forward primer	EV1F	CCCTGAATGCGGCTAAT	143	Gregory et al. 2006

	Reverse primer	EV1R	TGTCACCATAAGCAGCCA		
	Probe	EV probe	FAM-ACGGACACCCAAAGTAGTCGGTTC-BHQ1		
MNV, pMNV	Forward primer	MNV-S	CCGCAGGAACGCTCAGCAG	129	Kitajima et al. 2010
	Reverse primer	MNV-AS	GGYTGAATGGGGACGGCCTG		
	Probe	MNV-TP	FAM-ATGAGTGATGGCGCA-MGB-NFQ		

FIGURE LEGENDS

Figure 1 Observed detection (log_{10} copies/ml determined by qPCR) of adenovirus 2 (AdV), murine norovirus (MNV), and poliovirus type 1 (PV1) extracted using MO BIO, Zymo, or Qiagen kits from spiked A) molecular grade water (extraction efficiency controls), B) biosolid extracts, C) fecal suspensions, and D) 1MDS surface water concentrates. Three samples (1, 2, and 3 on the horizontal axis) for each type of environmental sample matrix were spiked with each virus. To adjust for the difference between the three extraction kits in their loading and elution volumes, the concentrations of the original spiked viruses were calculated from the qPCR data and expressed as log_{10} copies/ml.

Figure 2 Assessment of the presence of qPCR inhibitors in the nucleic acid extracts from three separate extraction kits via A) a qPCR inhibition test using an internal murine norovirus (pMNV) standard plasmid control (1.0×10^5 copies/PCR tube) in the presence of 2.5 μl of nucleic acid extract and B) a test for the presence of humic acid-like substances with an absorbance of 230 nm to estimate the presence of potential PCR inhibitors. Three samples (1, 2, and 3 on the horizontal axis) for each type of environmental sample matrix were tested.

APPENDIX D

MICROBIAL COMMUNITY STRUCTURE OF BIOFOAM: IDENTIFYING THE CORE BIOFILM COMMUNITY OF A NOVEL POU BIOLOGICAL WATER TREATMENT MEDIUM.

Brandon C. Iker[1], Mark D. Sobsey[4], Anne K. Camper[2], Joan B. Rose[3], Anne L. Hoyt[5], Matthew K. Runyon[5], Roy W. Kuennen[5] and Ian L. Pepper[1]

[1]Department of Soil, Water and Environmental Science, The University of Arizona, Tucson, AZ 85721, [2]Department of Civil Engineering, Montana State University, Bozeman, MT 59717, [3]Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI 48824, [4]Department of Environmental Science Engineering, University of North Carolina Chapel Hill, NC 27599, [5]Access Business Group, Ada, MI 49355.

*Corresponding author. Tel.: (520) 626-3328; Fax: (520) 573-0852

E-mail address: ipepper@ag.arizona.edu

Running title: The core microbiome of biofoam

Submitted to: International Society of Microbial Ecology

ABSTRACT

Small scale biosand filters (BSFs) utilized as point of use (POU) water treatment technologies can provide quality drinking water in arid regions. In this current study, biofilms within a novel biologically active POU technology that removes pathogens from water were studied. The biofilms develop across porous foam cartridge filters termed "biofoam," and were assayed for community membership. Biofilms within the POU filter medium were developed at three different locations in the US using three different surface waters. Biofilm microbial communities that developed on biofoam were analyzed utilizing 454 pyrosequencing of the 16S rDNA genes, and showed a remarkable degree of shared community membership among the three locations. A large, diverse shared microbiome was found as defined by the top 100 shared operational taxonomic units (OTUs) at 0.03 cut off (97% similarity). This represented 280,000 of the 306,000 sequences (>90%). Of those, 25% were classified within the genus *Pseudomonas.* Members of the microbial communities found within the shared microbiome of the biofoam were closely associated with organisms commonly found in activated sludge, drinking water biofilms, rhizosphere, phyllosphere, and soil ecosystems. The biofoam provides a unique and effective porous matrix for biofilm formation, which appears to allow for the establishment of consistent microbial communities, even when developed at different locations utilizing different water sources. Improving our understanding of the biofoam's microbial ecology may provide insights into mechanisms of pathogen removal, and allow development of customized biofilms for targeted remediation projects, as part of a lightweight and energy efficient water filtration system.

Key words: Biofilm, biofoam, microbial ecology, pyrosequencing, POU water filter, pathogen removal.

1. INTRODUCTION

With increasing concerns over access to safe drinking water, especially in arid, undeveloped regions, or in the aftermath of natural disasters, the development of portable, low cost, efficient water treatment technologies continues to be a global need. One low cost approach for treating contaminated drinking water utilizes the development of a biofilm or “schmutzdecke” on porous media like that of sand or aragonite. The utilization of biofilms in biosand filters (BSFs) and slow sand filters (SSFs) was first used in European and American cities in the 1800’s and 1900’s (CAWST 2013). Larger biological particles such as protozoa, fungal spores, and large organic matter are retained when they pass through the media by physical exclusion (Sobsey et al. 2008). In addition, the development of a biologically active treatment layer, a schmutzdecke, or biofilm, is essential for efficient bacterial pathogen removal (Elliott et al. 2011). The mechanisms by which biofilms or biological treatment layers capture and remove pathogens to enhance drinking water quality remains unknown (Votano et al., 2004 and Stauber et al., 2012). Therefore, studying the ecology of developed biofilms used for water treatment may provide a greater understanding of how planktonic microbial communities assemble and maintain a biofilm that removes pathogens (Rittmann 2006).

To date, there have been few studies that have attempted to systematically characterize microbial communities within these biofilms. Some have utilized traditional

techniques like culture-based assays (Hunter et al. 2012). Others have used molecular techniques such as denaturing gradient gel electrophoresis (DGGE), clone libraries, phylochip, and qPCR (Haig et al. 2005; Wakelin et al. 2011). To our knowledge, no studies have utilized deep sequencing analysis to evaluate the microbial communities of biofilms developed for point-of-use (POU) water treatment systems. Insights into how biofilms develop and function in biological water treatment may prove useful in gaining an understanding of pathogen removal and inactivation mechanisms. This understanding may allow for the optimization of the biological mechanisms involved in low cost, low energy water treatment technologies.

In the current study, biofilms were developed within a novel foam matrix termed "biofoam" that is utilized as part of a POU water treatment system. Biofoam consists of synthetic flexible foam within which a biofilm develops. Previous work has shown that biofoam is capable of capturing and removing a wide range of microbial pathogens as water permeates through the treatment layer as part of a POU water treatment system (Rose et al 2013). Thus, biofoam works on the basic principle of BSFs or SSFs, where traditional BSFs utilize tens to a hundred pounds of sand to support the growth of biofilms, biofoam achieves similar performance across a thin (7 mm) permeable foam layer (Bauer et al. 2011; Sobsey et al. 2008; Elliott et al. 2011). This new media significantly improves the portability and ease-of-use of biofiltration-based water treatment. This collaborative study including researchers from four universities was designed to test the hypothesis that biofilms developed under identical conditions would contain a common microbiome, even if different water sources were utilized. In other

words, although it could be anticipated that every biofilm would be unique, a shared community with similar functional representatives would be created. This is desirable if the treatment devices are to have similar efficacy at different geographical locations with varying water quality. Biofilms were established at three diverse locations in the US: Bozeman, MT (Montana State University); East Lancing, MI (Michigan State University) and Chapel Hill, NC (University of North Carolina) using surface water sources from each of these locations. The processing and analysis of the biofilms occurred at the University of Arizona in Tucson, Arizona using Roche 454 pyrosequencing of the bacterial 16S rRNA gene libraries. The results were used to evaluate the microbial communities present in the biofoam to better understand the effect of geographically and ecologically different source waters on community composition.

2. METHODS

2.1. Reactor operation

Three parallel prototype POU units (replicates) were set up at each location. Each unit included a top bucket that could accommodate 15 L batches of raw water, a subsequent biofoam filtration unit with two foam filters, and an empty third bucket to catch the filtrate (Figure 1). A plunger was pulled to allow flow from the top bucket into the unit with the foam filters. Batches of water were passed through the unit twice per day with a set time of 6 hrs between batches. The biofoam remained wet throughout the study period.

2.2. Source water

Approximately 450 L of untreated surface water was used at each research location. Batches of water were refrigerated at 4° ± 2°C, and warmed to ambient temperature prior to use. Water quality parameters were measured for each new batch and collected on sampling days. The source water used in Chapel Hill, North Carolina was collected from University Lake. The lake is a surface water impoundment used as source water for the Orange Water and Sewer Authority drinking water system, which serves much of the urban population of southern Orange County, NC (primarily Chapel Hill and Carrboro, NC). The lake allows non-primary contact recreation such as fishing, boating, and rowing. Sampling began on August 6, 2012 and the biofilms were harvested on September 6, 2012. The source water used in Bozeman, Montana originated from two surface waters serving the drinking water treatment plant: Hyalite Reservoir and Sourdough Creek. Both are pristine sources at the head of the watershed. Water was procured every three days from the intake of the drinking water treatment plant. The experiment began on August 6, 2012 and the biofilms were harvested on August 30, 2012. The source water in Lansing, Michigan was untreated surface water collected from the Grand River at Francis Park, which is located downstream of the Upper Grand River watershed. The designated uses at this location include total and partial body contact recreation. The experiment began on July 12, 2012 and the biofilms were harvested on August 14, 2012.

2.3. Water quality analysis

Raw source waters and effluents following filtration through the biofoam were analyzed for turbidity, conductivity, pH, hardness, alkalinity, total and dissolved organic carbon (TOC and DOC), heterotrophic bacterial plate counts (HPCs), total coliforms, and total bacterial counts using standard methods (APHA, 2012). The sampling intervals were days 0 (start date), 3, 7, 10, and 14, and subsequently at weekly intervals until a pseudo-steady state biofilm was attained. The HPCs and total direct counts were used to determine the pseudo-steady state of the biofilm, which was defined as three consecutive sample events having less than 10% - 15% variation of HPCs. The filters were run for one additional week beyond the achievement of the steady state to allow for the evaluation of HPCs from the preceding week.

2.4. Foam biofilm sampling

After achieving a pseudo-steady state, the biofoam membranes were removed from each of the two filter cartridges in the three replicate filter units at each location. The biofilms were harvested from each biofoam utilizing a stomacher. Approximately 300 mL of filter-sterilized effluent was used as a medium into which the biofoam were stomached. The samples were processed in the stomacher for 2 minutes at high rotation (260 RPM) and water from the stomacher bag was removed by squeezing the foam to expel any absorbed water into the bottom of the bag. Approximately 100 mL of water containing the recovered biomass was centrifuged at 5000x g for 10 minutes to achieve pellets containing approximately 1 g of biomass. A total of three replicates were collected from each filter preparation. To stabilize microbial populations for storage and shipping,

LifeGuard Soil Preservation Solution (MO BIO San Diego, CA) was added to each tube following manufacturer's instructions. The samples were stored at -20°C and then shipped on ice to the University of Arizona.

2.5. Molecular methods

2.5.1 Nucleic acid extraction

The biofilm suspensions were centrifuged at 5000 x *g* for 15 minutes to separate the biomass from the preservation solution. For each biofilm sample, 0.2 g of the biofilm mass was utilized in duplicate extractions using MO BIO PowerBiofilm kits (MO BIO San Diego, CA) following the manufacturer's suggested protocol. The nucleic acid extracts were eluted in 100 μL of molecular grade water and placed on ice followed by DNA quantification using Qubit 2.0 (Invitrogen, Carlsbad, CA). Subsequently, 5 μL of raw extract was visualized on a 1% ethidium bromide gel to check for genomic shearing during the extraction process. The DNA extracts for all samples were normalized to 10 ng/μL using molecular grade water, then pooled 1:1 (vol/vol) for each technical replicate, and stored at -20°C until PCR amplified.

2.5.2. PCR amplification

The community DNA was amplified using the bacteria-specific primer pairs 515F (GTGYCAGCMGCCGCGGTA) and 909R (CCCCGYCAATTCMTTTRAGT) that target the V4 V5 16S rRNA gene (Tamaki et al. 2011) using Takara HS EX high fidelity Taq DNA polymerase 2.0x master mix (Takara Inc. Japan). The PCR for each sample was carried out in triplicate 50 μL reactions using approximately 10 ng of template per reaction in a Mastercycler Nexus thermocycler (Eppendorf, Hamburg, Germany) under

the following reaction conditions: initial denaturation at 94°C for 3 min, followed by 30 cycles of 94°C for 40 s, 56°C for 1 min, and 72°C for 1 min with a final extension at 72°C for 10 min. The PCR products were visualized using electrophoresis on a 1.5% agarose gel, and the DNA band with the correct size was excised and purified using a Zymoresearch gel extraction kit (Zymo Research, Irvine, CA). The purified PCR amplicons were then pooled and normalized to 500 ng for submission to the University of Arizona Genomic Center (UAGC, Tucson, AZ) for sequencing using Roche 454 Titanium XL chemistry (Roche, Indianapolis, IN).

2.5.2. Sequence processing

Raw sequences were processed using Mothur software version 1.29.0 (Schloss et al. 2009) following the standard operating procedure outlined on http://www.motur.org/wiki/Schloss_SOP (Schloss et. al., 2011). After de-noising and the removal of chimeras, 306,729 sequences were binned into operational taxonomic units (OTUs) under the unique, 0.01, 0.03, and 0.05 definitions. For diversity estimates, all data sets were normalized to 28,000 sequences. Taxonomic and phylogenetic identification of representative sequences from OTUs of interest were done using the modified Mothur version of RDP's Bayesian classifier.

3. RESULTS

3.1. Water quality of source waters and POU effluents

The quality of all of the source waters (influents) and water collected following passage through the biofoam (effluents) at the 3 geographical locations are shown in Tables 1 and 2, respectively. The Michigan source water had increased levels of conductivity, hardness, and alkalinity relative to the Montana and North Carolina waters, indicating a higher total dissolved solids content. The Montana water was notable for its lower TOC and DOC content as compared to the other two source waters. The microbial water quality of the Michigan and North Carolina source waters were relatively similar in terms of culturable HPCs, but the Montana source water had one order of magnitude lower HPCs.

The chemical water quality of the effluents at all 3 sites was similar to the influent water quality, indicating that the water quality was not significantly altered by the biofoam (Table 2). The lack of increase in total and cultural (HPC) counts implies that if the biofilms became detached during the study, they contained microbial loads similar to those of the influent and therefore no increases in HPC were detected over time.

3.2. Microbial community analysis of POU biofoams

The 454 pyrosequencing generated 306,729 effective reads per sample, which clustered at the unique 0.01, 0.03, and 0.05 sequence similarities. Only OTUs at the 0.03

(97% similarity) definitions were used in the following analysis and discussion. A summary of the observed OTUs (Sobs), their richness, an alpha diversity estimator, and Simpson's inverse diversity indices are shown in Table 3, which suggests that the diversity was similar at each site, but in general, diversity decreased in the following order: Michigan > Montana > North Carolina. This observation is also inferred by the rarefaction curves representing the three replicate biofilm samples from each of the three sites (shown in Figure 2) that provide a visual comparison of the relative OTU richness.

A class level taxonomic summary of all the major OTUs (Figure 3) indicated little variation between the replicate biofilms developed at each site in terms of relative taxonomic abundance. All sites were dominated by planctomycetes and alpha-, beta-, and gamma- proteobacteria, representing 6%, 15%, 16%, and 21% of the ~6,330 observed OTUs, respectively. The taxonomic classification at the genus level revealed that the Michigan and North Carolina biofilms were dominated by *Pseudomonas*, with approximately 66,000 or 25% of all sequences in the data set belonging to members of this genus. Members of the Firmicutes phyla like *Sporoscarcina* and *Bacillus* were more abundant in the Montana biofilms in comparison to the other two sites, with *Sporoscarcina* being the most abundant sequence. However, it is difficult to estimate the relative abundance of specific bacteria using the 16S gene sequence analysis due to the inherent differences in the gene copy number between different species.

The shared microbial community in the biofoam was defined as the top 400 shared OTUs at the 0.03 cutoff that had more than 10 sequences within each OTU definition. The top 100 shared OTUs (summarized in Table 4) represented ~280,000 of

the 306,000 effective sequences (>90%). The beta diversity of the shared communities was evaluated using Chao 1 and abundance-based coverage estimator (ACE) (Table 5). Chao 1, a less conservative estimate, uses singletons and doubletons or the "rare" community members for its calculation, while ACE, a more conservative estimate, uses OTUs with counts of individuals from one to 10 sequences. For this study, 10 sequences were chosen. Under these pairwise observations, the Michigan and Montana biofilms were considered to have higher shared diversity than that of Montana and North Carolina. The averages of nine pairwise diversity comparisons between each of the sites using these diversity estimators are shown in Table 5.

UniFrac, a distance metric for beta diversity estimates using phylogenetic information, was used for the community comparisons by estimating the differences in community composition between two environments (Hamady et al. 2010). Un-weighted UniFrac scores account for the fraction of total branch lengths that are different (Table 5). The weighted UniFrac scores account for the relative abundance of each taxon within the communities. Unifrac scores close to one suggest fewer shared branches, while scores approaching zero suggest a higher fraction of shared branches.

4. DISCUSSION

Unlike the low biodiversity found in drinking water biofilms (Hong et al. 2010), the microbial communities found in the biofoam in this study resembled microbial

communities found in activated sludge, waste water treatment plants, and environmental biofilms associated with soil environments such as rhizosphere populations. This similarity was true in terms of both the microbial membership and species diversity (Ye & Zhang 2012). Interestingly, a recent publication by Besemer and colleagues (2012) provides insight into the ecology of fresh water biofilms. Their work suggested that biofilm communities obtained from different natural streams shared a high degree of homology, even though there were clear differences in the planktonic communities of source waters. This observation suggests that only select microorganisms are capable of successfully integrating into a biofilm. Thus, as defined in classical microbial ecology, bacteria fulfill a specific function or niche within a specific habitat, which in this case is the biofilm community itself.

A major objective of this study was to identify the shared microbiome between biofilms generated from source waters from three separate geographical locations. The 454-pyrosequencing analyses revealed that a large, highly conserved core microbiome was shared among the three sites (Table 4 and 5). This shared microbiome was dominated by bacterial genera commonly found in diverse but functionally relevant environments, including activated sludge, the phyllosphere, the rhizosphere, soil, natural water, and drinking water biofilms (Pinto et al. 2012; Mußmann et al. 2013; Bengtsson & Øvreås 2010; Henne et al. 2012; Zhang et al. 2012; Ramey et al. 2004). The microbial communities from each site were strikingly similar despite the fact that they were developed using three geographically different source waters (Tables 1). These data agree

with a recent study, which found that source water microbial community had little effect on biofilm community membership (Besemer et al., 2012).

OTUs classified to the genus level belong to many taxa that are well characterized for functions such as the production of exopolysaccharides (EPS), external DNA (eDNA), quorum sensing molecules, proteases, and chitinases. Although inferring specific functions from 16S rRNA gene sequences is difficult, we theorize that the presence of numerous OTUs with taxonomical relevance reveals the functional potential for multiple biochemical transformations, especially those associated with the nitrogen cycle. These functions include ammonia oxidation including anaerobic ammonia oxidation (ANAMOX), EPS production, protease activity, and biopolymer and polyaromatic hydrocarbon (PAH) degradation (Table 5).

Pseudomonas spp. were among the most abundant shared sequences in the core biofoam community. This genus is found in most natural environments including soil, water, and marine environments, and it has been reported that *Pseudomonas spp.* have broad fundamental niches (Remold et al. 2011). In particular, pseudomonads have diverse enzyme systems and large genomes capable of multiple biochemical transformations (Spiers et al., 2000; Stanier,et al.,1966). The potential metabolic activities of the pseudomonads suggests why they may be prolific and perhaps even mandatory members of the biofilm communities examined in the present study.

Other abundant OTUs of interest in the core biofoam community are classified within a distinct phylum of the domain bacteria as members of *Planctomycetes, Verrucomicrobia,* and *Chlamidia* (PVC). In our study, ~6% of the shared sequences at all

sites were classified as *Planctomycetes*, including *Rhodopirellula*, *Singulisphaera*, *Schlesneria*, *Zavarzinella*, *Gemmata*, and other unclassified genera. *Planctomycetes* are particularly prevalent in soils, fresh waters, waste water treatment plants, plant microbiomes, and marine environments (Buckley et al. 2006; Besemer et al. 2012). Most members of *Planctomycetes* are capable of carrying out ANAMMOX reactions, which involve anaerobic oxidation of ammonium to nitrogen gas using nitrite as an electron acceptor concurrent to the reduction of carbon dioxide. This chemoautotrophic metabolism occurs within an anaearobic membrane-bound cell compartment called the anammoxosome. Thus, biofilms with abundant *Planctomycetes* communities could have potential for the efficient removal of nitrogen and may play an important role in the nutrient cycling and maintenance of the biofilm community (Kartal et al. 2010).

Interestingly, the genus *Sporosarcina* was found in much higher relative abundance than *Pseudomonas* in the Montana biofilms (Table 5). *Sporosarcina* are aerobic bacteria commonly found in soil and water environments, which are known to produce high levels of urease (McCoy et al. 1992; Tobler et al. 2011). Due to *Sporosarcina* species' ability to readily break down urea, these organisms may play a major role in the nitrogen cycle (Tobler et al. 2011). Although it is unclear why there is a relatively high abundance of this sequence observed in the Montana samples, the biomass yields from the Montana biofilms were approximately $1/5^{th}$ the mass of the other two sites. This lower biomass recovery may have been due to the lower TOC in the Montana source water. Despite these differences, the Montana source water still resulted in a biofilm community similar to the other two sites.

Several theories have been developed to explain the mechanisms of pathogen reduction in biosand filters, including amoebic grazing, exopolysaccharide (EPS) production, protease activity, extracellular DNA (eDNA), amoebic grazing, and starvation (Flemming & Wingender 2010). Recently, biofilms developed within biofoam have been shown to successfully remove pathogens from surface water as a component of a POU water treatment system (Rose et al. 2013). The reductions of *Cryptosporidium parvum, R. terrigena,* and MS2 phage throughout the POU system were shown to be 6.45 ± 0.86, 8.36 ± 0.69, and 4.55 ± 1.04 logs, respectively. Specifically, the removal across the biofoam filter layer were measured to be 3.71 ± 0.51, 2.74 ± 1.30, and 2.23 ± 0.69, respectively. Some of the aforementioned mechanisms are likely to be at least partially responsible for the pathogen reduction across the biofoam matrix. Our future research will attempt to examine whether manipulating the core microbial communities found within the foam can enhance pathogen reduction and to identify which of the above mechanisms of action may be responsible for the reduction in pathogens across the biofoam matrix.

Pseudomonads are well documented as being producers of EPS and eDNA, and in this current study, the genus *Pseudomonas* was among the most dominant shared community members at all three sites. Biofilm communities are typically embedded within a complex mixture of macromolecules including both proteins and EPS. EPS has been implicated as essential for biofilm architecture, including for the aggregation of bacterial cells, cell-to-cell recognition and communication, and gene transfer (Flemming & Wingender 2010). EPS matrix stabilization can also be enhanced by bacterial

appendages such as pili, fimbriae, and flagella (Branda et al. 2005). In addition, membrane vesicles that are derived from the outer membrane of the Gram-negative bacterium *Pseudomonas aeruginosa* are thought to intertwine with pili and flagella and to be important components of the biofilm matrix (Schaik et al. 2005; Schooling & Beveridge 2006). EPS is most likely constructed and maintained by members of the pseudomonads and other EPS producing organisms as a defense mechanism against protozoan grazing. Some bacteria including *P. aeruginosa* are known to produce substantial amounts of eDNA within biofilms, and such eDNA may be a requirement for biofilm formation (Whitchurch et al. 2002). eDNA is also known to have antimicrobial activity, causing cell lysis by chelating cations that normally stabilize the lipopolysaccharides on the outer membranes of bacteria (Mulcahy et al. 2008).

In addition, many of the shared OTUs observed in this study belong to members of bacterial taxonomic groups known to produce high levels of extracellular proteases, chitinases, endonucleases, and lipases (Kim et al. 2010; Molobela et al. 2010; Schloss & Handelsman 2003; van Frankenhuyzen et al. 2011). We speculate that such enzymes are most likely present and are a functional component of the biofilm community (Table 5). Both bacterial and protozoan proteases most likely play a large role in nutrient acquisition from source water for microbiota that are unable to integrate into the biofilm environment. Although eukaryotic communities in this study were not addressed, we theorize that the presence of eukaryotic protozoa and other highly abundant saprophytic bacteria are associated with surface grazing and nutrient cycling and may also provide much of the nitrogen that ultimately ends up within the biofoam (Thomas & Ashbolt

2011; Koh et al. 2012; Valster et al. 2009). This theory is supported by the presence of the diverse genus *Planctomycete* and other highly abundant diazotrophic organisms like *Sporocarcina* that could aid in detoxifying ammonia, allowing for biofilm protection and maintenance. Ammonia could be produced as the result of protozoa digesting microorganisms that are not able to seek refuge inside the biofilm, or that are not equipped with mechanisms to survive protozoan digestion. Future microbial surveys will involve thorough investigations into the role of eukaryotic organisms as functional members of the biological treatment layer.

The physical structure of the foam matrix may has major advantages over traditional sand media (Figure 4) for the development of a microbial biofilm matrix. One can think of the space found inside the porous foam as the inverse of the spaces found in sand media. These pores may play a role in shaping effective microbial communities that assemble on and within the biofoam and result in theefficient filtering of water for biological pathogen removal. One model that describes the structure and function of biofilms suggests that they consist of a number of microcolonies,which are spatially separated and heterogeneous with respect to their physical and chemical structure (Johnson, 2008). This model suggests that the microcolonies are building blocks for the biofilm. In our study, the porous nature of the biofoam (Figure 4) with variable intestinal voids and a large surface area may be a key factor in the production of biofilms that were similar for the three different locations. The pore sizes are highly variable in the biofoam. The large pores may allow for the efficient flow of water through the foam. The aeration,

pore space architecture, and large surface area provided by the foam appear to provide an ideal habitat for microorganisms to develop a biologically active biofilm matrix.

In conclusion, members of the microbial communities found within the core microbiome of the biofoam from the three locations were closely associated with microorganisms commonly found in activated sludge, drinking water biofilms, the rhizosphere, the phylloshere, and soil ecosystems. Our data suggests that biofoam may provide a unique matrix for biofilm formation that enables the development of select microbial communities shown to be efficient at removing viral, bacterial, and protozoa pathogens. Improving our understanding of the biofoam's microbial ecology could potentially provide insights into the mechanisms of action for the removal of microorganisms, the conditions that would result in expedited biofilm ripening times, and in how to develop customized biofilms for targeted remediation projects. Future studies on the biofilm structure and function will focus on the use of functional genomics including transcriptomics and proteomics. Additionally, providing a greater understanding of the spatial and temporal components of biofilm development should prove essential if one wishes to capitalize on more rapidly developed, resilient, and functionally diverse microbial communities for efficient biological water treatment. Thus, the biofoam may have novel applications in a lightweight, low cost, low energy water filtration technology for varioius matrices such as the treatment of drinking water, irrigation water, and gray water.

Acknowledgments

The authors would like to thank Ms. Becca Ives at the Water Quality, Environmental, and Molecular Microbiology Laboratory (WQEMM) located at Michigan State University in East Lansing, Michigan, Steve Hernandez from the University of Arizona for help with the SEM imaging, Douglas Wait, Emily Bailey and Tucker Witsil from the University of North Carolina for their laboratory support, andThe University of Arizona's National Science Foundation Water and Environmental Technology (WET) Center and Amway Corporation for their funding and support.

REFERENCES

APHA. (2012). Standard Methods for the Examination of Water and Wastewater 22nd Ed. American Public Health Association, Washington DC.

Bauer R, Dizer H, Graeber I, Rosenwinkel K-H, López-Pila JM. (2011). Removal of bacterial fecal indicators, coliphages and enteric adenoviruses from waters with high fecal pollution by slow sand filtration. Water research 45:439–52.

Bengtsson MM, Øvreås L. (2010). Planctomycetes dominate biofilms on surfaces of the kelp Laminaria hyperborea. BMC Microbiology 10:261.

Besemer K, Peter H, Logue JB, Langenheder S, Lindström ES, Tranvik LJ, et al. (2012). Unraveling assembly of stream biofilm communities. The ISME journal 6:1459–68.

Branda SS, Vik S, Friedman L, Kolter R. (2005). Biofilms: the matrix revisited. Trends in microbiology 13:20–6.

Buckley DH, Huangyutitham V, Nelson T a., Rumberger a., Thies JE. (2006). Diversity of planctomycetes in soil in relation to soil history and environmental heterogeneity. Applied and environmental microbiology 72:6429–6429.

CAWST. (2013). Center for Affordable Water and Sanitation Technology. http://www.cawst.org/en/resources/biosand–filter.

Elliott MA, DiGiano FA, Sobsey MD. (2011). Virus attenuation by microbial mechanisms during the idle time of a household slow sand filter. Water Research 45:4092–4102.

Flemming H-C, Wingender J. (2010). The biofilm matrix. Nature reviews. Microbiology 8:623–33.

Van Frankenhuyzen JK, Trevors JT, Lee H, Flemming C a, Habash MB. (2011). Molecular pathogen detection in biosolids with a focus on quantitative PCR using propidium monoazide for viable cell enumeration. Journal of microbiological methods 87:263–72.

Haig S, Dorea C, Collins G, Davies R, Quince C. (2005). Validating Laboratory Slow Sand Filtration Studies Through Water Quality and Molecular analysis. userweb.eng.gla.ac.uk 2005–2006.

Hamady M. (2010). Fast UniFrac: Facilitating high-throughput phylogenetic analysis of microbial communities including analysis of pyrosequencing and Phylochip data. 4:17–27.

Henne K, Kahlisch L, Brettar I, Höfle MG. (2012). Analysis of structure and composition of bacterial core communities in mature drinking water biofilms and bulk water of a citywide network in Germany. Applied and environmental microbiology 78:3530–3538.

Hong P-Y, Hwang C, Ling F, Andersen GL, LeChevallier MW, Liu W-T. (2010). Pyrosequencing analysis of bacterial biofilm communities in water meters of a drinking water distribution system. Applied and environmental microbiology 76:5631–5635.

Hunter PJ, Calvo-Bado LA., Parsons NR, Pettitt TR, Petch GM, Shaw E, et al. (2012). Variation in microbial communities colonizing horticultural slow sand filter beds: implications for filter function. Irrigation Science. DOI 10.1007/s00271-012-0339-z.

Kartal B, Kuenen JG, van Loosdrecht MCM. (2010). Engineering. Sewage treatment with anammox. Science (New York, N.Y.) 328:702–3.

Kim EH, Cho KH, Lee YM, Yim JH, Lee HK, Cho J-C, et al. (2010). Diversity of cold-active protease-producing bacteria from arctic terrestrial and marine environments revealed by enrichment culture. Journal of microbiology (Seoul, Korea) 48:426–32.

Koh KS, Matz C, Tan CH, Le HL, Rice S a, Marshall DJ, et al. (2012). Minimal increase in genetic diversity enhances predation resistance. Molecular ecology 21:1741–53.

McCoy D, Cetin A, Hausinger R. (1992). Characterization of urease from Sporosarcina ureae. Archives of microbiology 157:411–416.

Molobela IP, Cloete TE, Beukes M. (2010). Protease and amylase enzymes for biofilm removal and degradation of extracellular polymeric substances (EPS) produced by Pseudomonas fluorescens bacteria. African journal of microbiology research 4:1515–1524.

Mulcahy H, Charron-Mazenod L, Lewenza S. (2008). Extracellular DNA chelates cations and induces antibiotic resistance in Pseudomonas aeruginosa biofilms. PLoS pathogens 4:e1000213.

Mußmann M, Ribot M, von Schiller D, Merbt SN, Augspurger C, Karwautz C, et al. (2013). Colonization of freshwater biofilms by nitrifying bacteria from activated sludge. FEMS microbiology ecology. DOI 10.1111/1574-6941.12103.

Pinto AJ, Xi C, Raskin L. (2012). Bacterial community structure in the drinking water microbiome is governed by filtration processes. Environmental science & technology 46:8851–9.

Ramey BE, Koutsoudis M, von Bodman SB, Fuqua C. (2004). Biofilm formation in plant-microbe associations. Current opinion in microbiology 7:602–9.

Remold S, Brown C, Farris J, Hundley T, Perpich J, Purdy M. (2011). Differential habitat use and niche partitioning by pseudomonas species in human homes. Microbial Ecology 62:505–517.

Schloss PD, Gevers D, Westcott SL. (2011). Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. PloS one 6:e27310.

Schloss PD, Handelsman J. (2003). Biotechnological prospects from metagenomics. Current Opinion in Biotechnology 14:303–310.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Applied and environmental microbiology 75:7537–41.

Sobsey MD, Stauber CE, Casanova LM, Brown JM, Elliott M a. (2008). Point of use household drinking water filtration: A practical, effective solution for providing sustained access to safe drinking water in the developing world. Environmental science & technology 42:4261–4267.

Spiers AJ, Buckling A, Rainey PB. (2000). MINI REVIEW The causes of Pseudomonas diversity. Microbiology 146:2345–2350.

Stanier RY, Palleroni NJ, Doudoroff M. (1966). The aerobic pseudomonads: a taxonomic study. Journal of general microbiology 43:159–271.

Tamaki H, Wright CL, Li X, Lin Q, Hwang C, Wang S, et al. (2011). Analysis of 16S rRNA amplicon sequencing options on the Roche/454 next-generation titanium sequencing platform. PloS one 6:e25263.

Thomas JM, Ashbolt NJ. (2011). Do free-living amoebae in treated drinking water systems present an emerging health risk? Environmental science & technology 45:860–9.

Tobler DJ, Cuthbert MO, Greswell RB, Riley MS, Renshaw JC, Handley-Sidhu S, et al. (2011). Comparison of rates of ureolysis between Sporosarcina pasteurii and an

indigenous groundwater community under conditions required to precipitate large volumes of calcite. Geochimica et Cosmochimica Acta 75:3290–3301.

Valster RM, Wullings B a, Bakker G, Smidt H, van der Kooij D. (2009). Free-living protozoa in two unchlorinated drinking water supplies, identified by phylogenic analysis of 18S rRNA gene sequences. Applied and environmental microbiology 75:4736–46.

Wakelin S, Page D, Dillon P, Pavelic P, Abell GCJ, Gregg a. L, et al. (2011). Microbial community structure of a slow sand filter schmutzdecke: a phylogenetic snapshot based on rRNA sequence analysis. Water Science & Technology: Water Supply 11:426.

Whitchurch CB, Tolker-Nielsen T, Ragas PC, Mattick JS. (2002). Extracellular DNA required for bacterial biofilm formation. Science (New York, N.Y.) 295:1487.

Ye L, Zhang T. (2012). Bacterial communities in different sections of a municipal wastewater treatment plant revealed by 16S rDNA 454 pyrosequencing. Applied microbiology and biotechnology DOI 10.1007/s00253-012-4082-4.

Zhang T, Shao M-F, Ye L. (2012). 454 Pyrosequencing Reveals Bacterial Diversity of Activated Sludge From 14 Sewage Treatment Plants. The ISME journal 6:1137–47.

TABLES

Table 1. Influent water quality[(a)]

	Montana	Michigan	North Carolina
Number of samples (n)[(b)]	8(4)	7(4)	9(5)
HPC (LOG CFU/mL)	4.2±0.5	5.2±0.7	5.3±0.7
Total Coliforms (LOG CFU/mL)	2.7±0.7	1.3±2.8	1.0±0.2
Total Direct Counts (LOG cells/mL)	7.1±1.1	6.7±0.6	6.79±0.3
Turbidity (at 600 nm)	ND	3.0±4.0	5.8±1.8
Conductivity (µS/cm)	136.4±24.1	698.3±82.6	106.2±5.2
pH	8.05±0.6	8.0±0.7	7.8±0.37
Hardness (mg/L CaCO3)	105.0±20.0	223.3±48.6	56.0±17.9
Alkalinity (mg/L CaCO3 methyl orange)	85.0±14.1	157.5±61.9	51.7±7.5
TOC (mg/L)	1.2±2.8	5.8±1.8	6.7±2.3
DOC (mg/L)	2.1±4.3	5.4±2.2	6.4±2.5

[(a)]Mean water quality values ± standard deviations.

[(b)]n is the sample number for all parameters except Conductivity, pH, Hardness, and Alkalinity which were measured weekly. The numbers for those four parameters are shown in parentheses.

Table 2. Effluent water quality[(a)]

	Montana	Michigan	North Carolina
Number of samples[(b)]	24 (12)	21 (12)	27 (15)
HPC (LOG CFU/mL)	4.8±0.3	5.03±0.4	5.1±0.3
Total Coliforms (LOG CFU/mL)	3.5±0.5	1.2±1.3	0.8±0.1
Total Direct Counts (LOG cells/mL)	6.9±0.4	6.8±0.2	6.6±0.2
Turbidity (at 600 nm)	ND	1.8±1.1	3.7±0.6
Conductivity (µS/cm)	139.5±7.3	700.5±42.2	107.2±2.2
pH	8.01±0.3	8.07±0.3	7.7±0.1
Hardness (mg/L CaCO3)	105.0±9.1	225.4±21.5	52.0±10.1
Alkalinity (mg/L CaCO3 methyl orange)	87.1±5.4	165.0±39.2	51.7±3.5
TOC (mg/L)	1.92±1.4	5.9±0.9	6.9±1.7
DOC (mg/L)	1.8±1.9	5.7±1.0	7.00±1.6

[(a)]Mean water quality of effluent following POU treatment ± standard deviations.

[(b)]n is the sample number for all parameters except Conductivity, pH, Hardness, and Alkalinity which were measured weekly. The number for those four parameters is shown in parentheses.

Table 3 Summary of raw sequence data, cleaned sequences, observed OTUs, and alpha diversity estimators.

BIofoam sample IDs		Total raw 454 pyrosequencing reads	Effective sequences	Estimated coverage (Good's average)	Observed OTUs (0.03%)	Alpha diversity (Inverse Simpson's)	Standard Deviation (+/-)
Michigan State	MI1	72,218	31,797	94%	3,208	23.4	1.8
	MI2	71,693	33,015	95%	2,519	4.6	0.2
	MI3	66,700	31,449	94%	2,802	9.4	0.5
Montana State University	MT1	75,170	37,724	95%	2,401	16.1	1.2
	MT2	77,009	34,207	94%	2,804	39.0	3.3
	MT3	69,052	32,791	94%	3,044	94.1	6.1
University of North Carolina	NC1	74,959	39,527	96%	1,749	10.1	0.4
	NC2	72,744	35,480	95%	2,230	17.7	1.0
	NC3	69,237	39,527	96%	1,781	11.0	0.4
	Total=	648,782	306,729				

□

Table 4 Rank abundance heatmap of the top 100 shared OTUs showing the number of sequences in each OTU definition and their closest taxonomic identification to the genus level

Table 4 Taxonomic hierarchichical heatmap of top 100 shared OTUs (0.03 cutoff) represening the number of sequences in OTU definition and taxonomic identification to the genus level (89% CI).

OTU #	Size	Phylum	Genus	MI 1	MI 2	MI 3	MT 1	MT 2	MT 3	NC 1	NC 3	NC 3
Otu035	1167	Actinobacteria	Arthrobacter	0	0	0	741	26	14	33	81	272
Otu080	277	Actinobacteria	Nocardioides	14	5	7	5	3	13	61	65	104
Otu086	226	Actinobacteria	Marmoricola	0	1	0	82	46	90	3	0	4
Otu050	674	Actinobacteria	unclassified	93	53	55	105	91	146	26	61	44
Otu065	428	Actinobacteria	unclassified	48	14	25	99	71	127	13	18	13
Otu082	253	Actinobacteria	unclassified	19	11	6	45	80	76	4	5	7
Otu009	7426	Bacteroidetes	Flavobacterium	248	289	256	231	386	531	1343	2226	1916
Otu011	5934	Bacteroidetes	unclassified	969	975	923	455	556	259	53	1721	23
Otu022	2940	Bacteroidetes	unclassified	153	162	169	449	434	617	135	684	137
Otu026	2050	Bacteroidetes	unclassified	410	296	311	291	516	145	12	56	13
Otu038	1066	Bacteroidetes	Ferruginibacter	109	109	108	98	92	135	90	264	61
Otu039	1002	Bacteroidetes	unclassified	3	0	2	341	479	176	0	1	0
Otu051	671	Bacteroidetes	Ohtaekwangia	14	13	7	232	225	52	6	121	1
Otu053	588	Bacteroidetes	Terrimonas	72	73	91	105	81	102	7	37	20
Otu057	500	Bacteroidetes	Sediminibacterium	80	98	75	85	100	46	3	9	4
Otu059	480	Bacteroidetes	Pedobacter	7	7	5	0	12	9	33	265	142
Otu061	463	Bacteroidetes	Runella	121	94	109	52	57	30	0	0	0
Otu081	254	Bacteroidetes	unclassified	82	60	69	16	17	9	0	0	1
Otu090	195	Bacteroidetes	Emticicia	1	1	2	65	90	33	0	3	0
Otu023	2897	Acidobacteria	Gp4	739	806	822	135	105	205	9	53	23
Otu075	306	Acidobacteria	Gp6	68	51	38	19	54	49	3	20	4
Otu040	967	Armatimonadetes	Armatimonadetes_gp5	129	78	92	137	176	304	9	19	23
Otu060	465	Armatimonadetes	Armatimonas_gp1	68	29	13	70	122	155	1	4	3
Otu100	147	Chloroflexi	Caldilinea	39	35	26	10	11	10	5	7	4
Otu002	20032	unclassified*	unclassified	3690	1950	2326	2046	2957	3449	812	1875	927
Otu004	12951	Firmicutes	Sporosarcina	2	0	0	8376	4500	73	2	0	0
Otu033	1399	Firmicutes	Exiguobacterium	3	2	1	1344	5	44	0	0	0
Otu063	436	Firmicutes	Pasteuria	92	80	70	36	63	44	3	34	14
Otu066	416	Firmicutes	Clostridium	36	18	23	11	27	29	106	70	96
Otu088	206	Firmicutes	unclassified	79	51	50	1	5	9	3	4	4
Otu031	1797	Nitrospira	Nitrospira	230	103	287	36	56	87	57	906	35
Otu019	3554	Planctomycetes	unclassified	474	489	431	455	593	712	150	129	121
Otu032	1530	Planctomycetes	Planctomyces	376	258	292	77	122	90	127	93	95
Otu044	846	Planctomycetes	Zavarzinella	64	72	68	167	200	220	18	19	18
Otu067	411	Planctomycetes	Schlesneria	12	14	8	64	124	104	24	32	29
Otu089	197	Planctomycetes	Singulisphaera	69	26	33	14	11	17	7	5	15
Otu092	189	Planctomycetes	Rhodopirellula	8	12	14	39	77	39	0	0	0
Otu001	66281	Proteobacteria	Pseudomonas	7486	17414	12787	2382	398	261	10538	3883	11132
Otu005	12596	Proteobacteria	Arenimonas	8	3	3	696	765	1016	1739	7073	1293
Otu015	4495	Proteobacteria	unclassified	739	462	534	438	534	623	232	488	445
Otu017	3984	Proteobacteria	Haliea	311	178	221	748	775	1677	6	40	28
Otu018	3589	Proteobacteria	Legionella	92	53	67	810	851	1126	178	170	242
Otu028	1940	Proteobacteria	unclassified	55	38	47	365	403	539	46	383	64
Otu076	303	Proteobacteria	Stenotrophomonas	0	0	0	23	20	14	35	173	38
Otu091	194	Proteobacteria	unclassified	19	16	18	34	45	28	8	13	13
Otu094	170	Proteobacteria	Aquicella	25	12	28	7	16	20	12	21	29
Otu079	280	Proteobacteria	unclassified	63	46	42	12	13	18	16	50	20
Otu083	249	Proteobacteria	unclassified	59	41	27	28	32	27	4	26	5
Otu007	10126	Proteobacteria	Janthinobacterium	0	0	0	13	38	50	4821	2322	2882
Otu010	6104	Proteobacteria	unclassified	1	0	3	6	21	10	1011	734	4318
Otu013	5630	Proteobacteria	unclassified	1075	585	748	550	434	470	265	1162	341
Otu014	4864	Proteobacteria	Massilia	0	0	0	1	3	3	903	1605	2349
Otu024	2854	Proteobacteria	Albidiferax	80	40	49	1077	762	576	60	141	69
Otu027	2010	Proteobacteria	Methylophilus	140	61	91	554	545	607	2	8	2
Otu029	1883	Proteobacteria	unclassified	258	181	190	363	248	211	57	311	64
Otu030	1811	Proteobacteria	Hydrogenophaga	529	447	480	134	81	82	17	22	19
Otu034	1371	Proteobacteria	unclassified	299	218	301	171	125	97	16	109	35
Otu041	959	Proteobacteria	Polaromonas	3	2	0	109	88	96	230	326	105
Otu046	832	Proteobacteria	Acidovorax	192	142	143	80	38	49	48	45	95
Otu047	791	Proteobacteria	unclassified	277	158	148	20	31	16	13	94	34
Otu058	498	Proteobacteria	Polynucleobacter	235	89	131	7	13	14	4	5	0
Otu064	433	Proteobacteria	Azospira	214	120	96	0	0	0	1	2	0
Otu068	407	Proteobacteria	Naxibacter	0	0	0	0	0	1	93	28	285
Otu070	387	Proteobacteria	Dechloromonas	160	76	96	20	18	8	1	5	3
Otu071	387	Proteobacteria	Deefgea	0	0	1	110	161	115	0	0	0
Otu078	285	Proteobacteria	Sulfuritalea	132	47	83	1	0	1	3	11	7
Otu084	241	Proteobacteria	Variovorax	31	16	17	49	94	18	7	7	2
Otu087	210	Proteobacteria	Rhodoferax	1	2	0	77	74	52	2	2	0
Otu093	182	Proteobacteria	Sphaerotilus	1	5	3	77	46	43	2	2	3
Otu096	156	Proteobacteria	Aquabacterium	12	5	9	56	34	31	0	9	0
Otu098	152	Proteobacteria	Pelomonas	0	0	0	0	0	0	18	133	1
Otu003	18065	Proteobacteria	Sphingomonas	474	431	504	1326	1812	1560	3784	4320	3854
Otu006	11078	Proteobacteria	unclassified	919	486	608	2156	2840	3657	102	219	91
Otu008	8073	Proteobacteria	unclassified	1246	644	989	926	1380	2371	135	174	208
Otu012	5776	Proteobacteria	unclassified	579	330	389	916	1259	1026	268	643	366
Otu016	4018	Proteobacteria	unclassified	610	376	447	307	540	521	364	587	266
Otu020	3226	Proteobacteria	Novosphingobium	83	40	49	1029	959	938	32	44	52
Otu021	3026	Proteobacteria	unclassified	70	38	46	786	856	825	83	146	176
Otu025	2230	Proteobacteria	Rhodobacter	839	369	499	83	102	130	71	92	45
Otu036	1105	Proteobacteria	unclassified	199	119	136	138	241	265	3	0	4
Otu037	1089	Proteobacteria	unclassified	20	13	13	175	268	560	12	15	13
Otu042	945	Proteobacteria	Sphingopyxis	2	2	3	222	407	194	35	56	24
Otu043	847	Proteobacteria	unclassified	110	58	72	180	139	136	24	106	22
Otu045	842	Proteobacteria	Devosia	144	65	84	91	127	199	47	80	5
Otu049	747	Proteobacteria	Hyphomicrobium	85	57	65	114	170	244	1	2	9
Otu055	535	Proteobacteria	Brevundimonas	35	26	38	44	60	63	32	72	165
Otu056	527	Proteobacteria	unclassified	181	118	111	16	27	16	8	42	8
Otu062	448	Proteobacteria	Caulobacter	73	47	47	24	38	19	40	132	28
Otu069	390	Proteobacteria	unclassified	55	24	58	55	55	118	8	8	9
Otu073	367	Proteobacteria	Roseomonas	90	58	40	36	38	36	18	26	25
Otu077	296	Proteobacteria	Phenylobacterium	8	1	5	52	67	81	24	28	30
Otu085	229	Proteobacteria	Rickettsia	30	11	9	34	66	64	12	3	0
Otu097	154	Proteobacteria	unclassified	27	21	14	24	23	29	2	8	6
Otu099	148	Proteobacteria	unclassified	7	1	6	28	45	43	6	8	4
Otu048	764	Verrucomicrobia	unclassified	158	121	165	69	126	118	0	5	2
Otu052	610	Verrucomicrobia	unclassified	57	31	44	133	183	143	1	17	1
Otu054	547	Verrucomicrobia	Spartobacteria	26	17	16	39	46	76	114	64	149
Otu072	369	Verrucomicrobia	3_genus_incertae_sedis	86	59	65	13	41	19	10	63	13
Otu074	360	Verrucomicrobia	Luteolibacter	97	43	63	43	52	62	0	0	0
Otu095	158	Verrucomicrobia	unclassified	66	46	37	1	3	0	2	3	0

*OTU 2 is only top 100 shared OTUs to not be taxenomically placed within 80% CI to taxonomical phyllum. Ranked closest to FIrmicutes at (65%) but not likely accurate placement.

Table 5 Summary of average pairwise beta diversity estimators sobs, Chao 1, ACE, and weighted and un-weighted UniFrac scores from each site (n=9).

Table 5 Average pairwise Unifrac score and beta diveristy estimates between each site

Groups	Unifrac		Beta diversity estimates		
	weighted	un-weighted	Sobs	Chao 1	ACE
Michigan-Montana	0.46 ±0.04	0.77 ±<0.01	533 ±27	835 ±83	791 ±25
Michigan-NorthCarolina	0.36 ±0.05	0.80 ±<0.01	426 ±53	648 ±77	644 ±68
Montana-NorthCarolina	0.38 ±0.01	0.82 ±<0.01	366 ±47	601 ±116	550 ±57

Figures

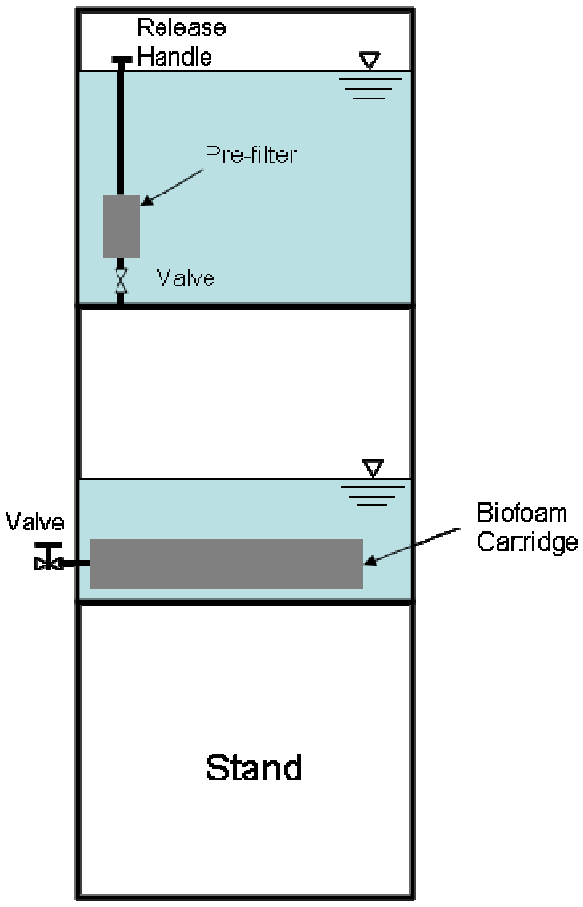


Figure 1. Schematic of biofoam test unit.

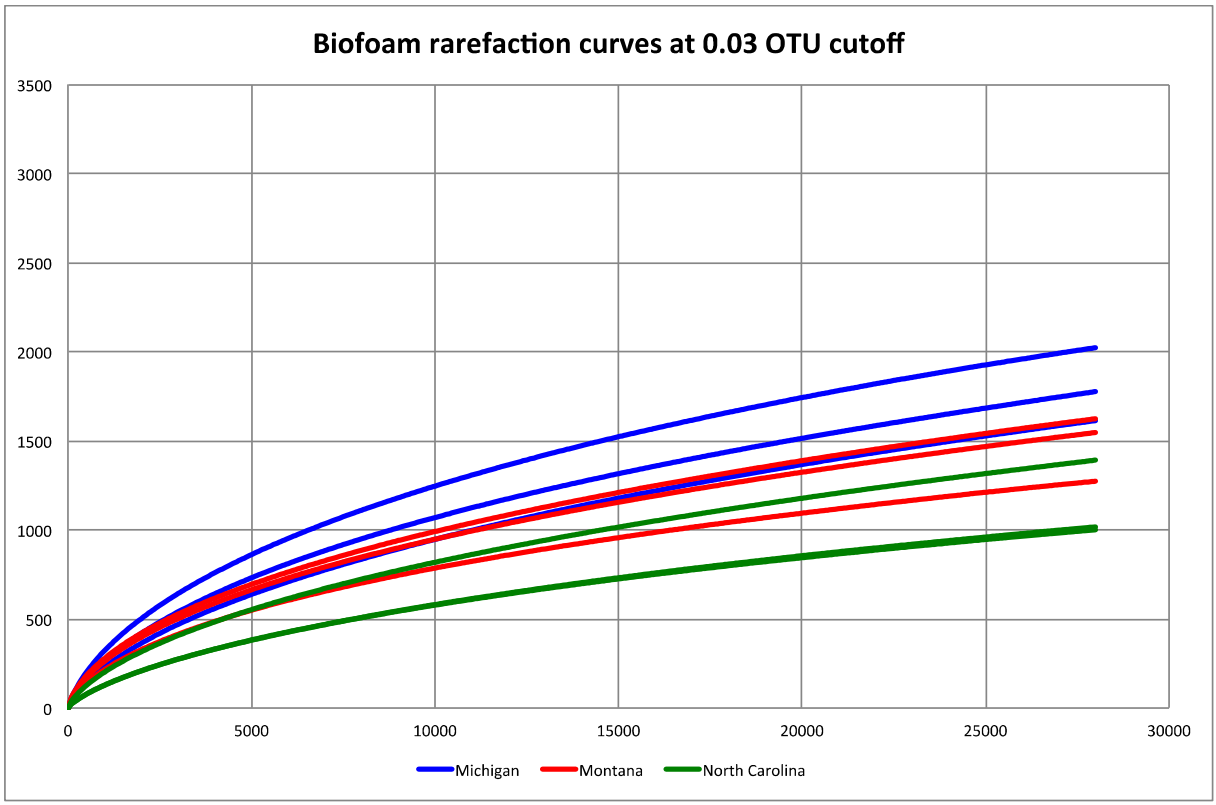


Figure 2. Observed OTU richness and estimated community coverage per sampling effort. Rarefaction curves indicating the expected OTU richness from each replicate with increased sampling efforts. Each library was normalized to 28000 sequences and OTUs were defined at 0.3% cutoff.

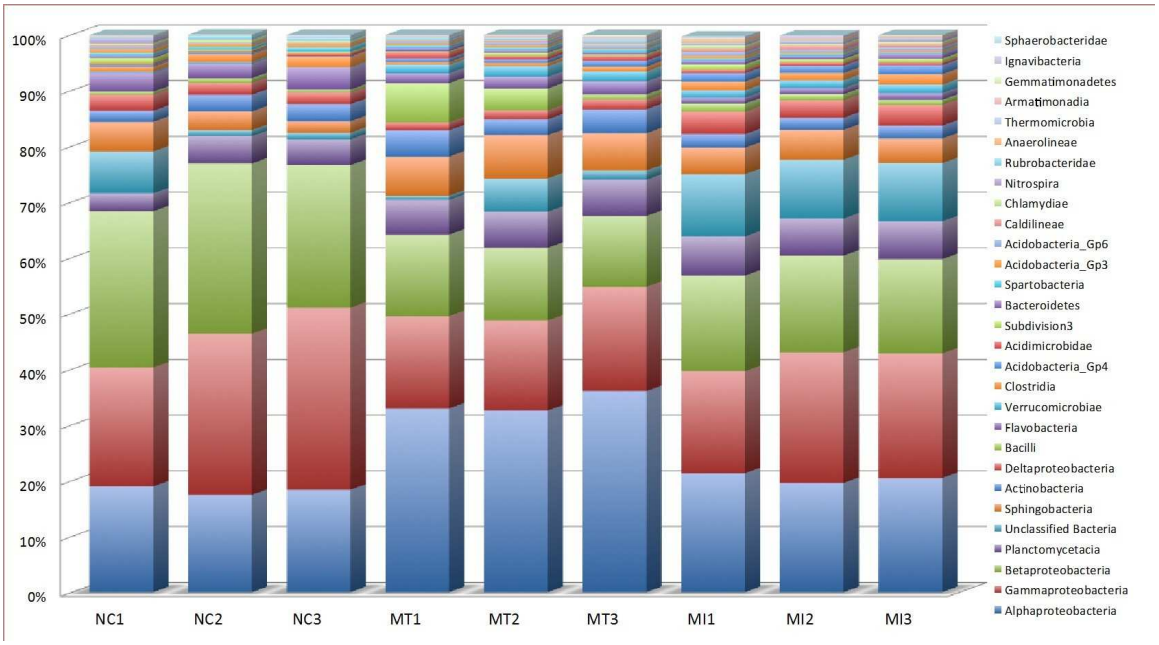


Figure 3. Class level taxonomic summaries of OTUs at 97% similarity.

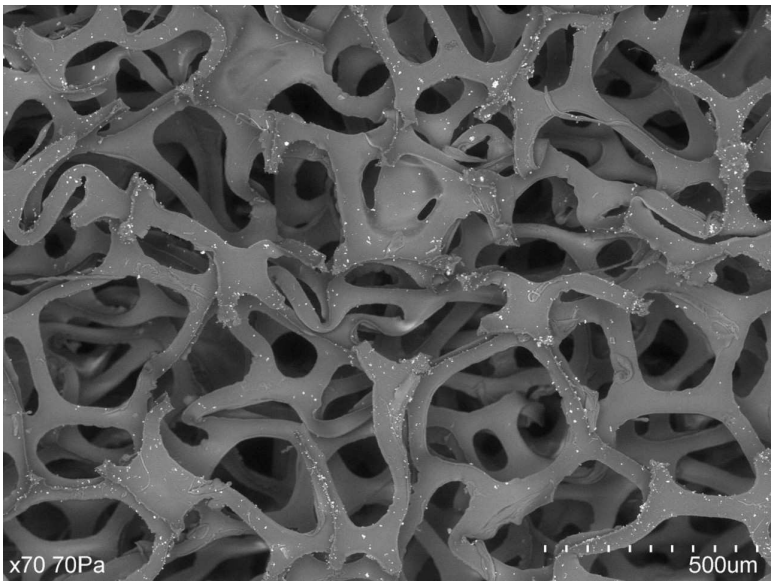


Figure 4. SEM image of the structure of biofoam.