**Diversity of DNA viruses in effluents of membrane bioreactors in Traverse City, MI (USA)
and La Grande Motte (France)**

**Evan O'Brien¹, Mariya Munir¹, Terence Marsh², Marc Heran³, Geoffroy Lesage³,
Volodymyr V. Tarabara¹, Irene Xagorarakis¹**

¹Department of Civil and Environmental Engineering, Michigan State University, East Lansing,
MI 48823 USA

²Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing,
MI 48823 USA

³Institut Européen des Membranes, IEM; UMR-5635, Université de Montpellier, ENSCM,
CNRS; Place Eugène Bataillon 34095 Montpellier cedex 5, FRANCE

*Corresponding Author: Irene Xagorarakis

Mailing address:

A124 Engineering Research Complex,

Michigan State University,

East Lansing, MI 48824

Phone: (517) 353-8539. Fax: (517) 355-0250.

E-mail: xagorara@msu.edu

Submitted to: Water Research

Abstract

This study assesses diversity of DNA viruses in the effluents of two membrane bioreactor (MBR) wastewater treatment plants (WWTPs): an MBR in the United States and an MBR in France. Viral diversity of these effluents is compared to that of a conventional activated sludge WWTP in the U.S. Diversity analysis indicates *Herpesvirales* to be the most abundant order of potentially pathogenic human DNA viruses in wastewater treated effluent in all utilities. Other potentially pathogenic human viruses detected include *Adenoviridae*, *Parvoviridae*, and *Polyomaviridae*. Bacteriophage order *Caudovirales* comprises the majority of DNA virus sequences in the effluent of all utilities. The choice of treatment process (MBR versus activated sludge reactor) utilized had no impact on effluent DNA viral diversity. In contrast, the type of disinfection applied had an impact on the viral diversity present in the effluent.

Keywords: Membrane bioreactor, metagenomics, viruses, wastewater treatment

1. Introduction

Viruses are potentially the most hazardous pathogens among those found in wastewater (Sidhu et al., 2008; Toze, 1997). They are also generally more difficult to detect in environmental samples. A high diversity of human viral pathogens is present in the environment (approximately 200 recognized human viral pathogen species) and is further elevated in samples affected by pollution. Moreover, additional species are continuously discovered (Bibby, 2013). It has been estimated that 2 to 12 million people die every year from waterborne diseases. While the majority of the outbreaks are caused by unidentified agents, it has been suggested that most agents in question are enteric viruses in groundwater and surface water bodies (US EPA, 2006).

Despite recent advances in water and wastewater treatment technology, waterborne diseases still pose a serious threat to public health across the world (Toze, 1997).

The Contaminant Candidate List (CCL) compiled and periodically updated by the U.S. Environmental Protection Agency includes contaminants that are known or anticipated to occur in public water systems, and which may require regulation under the Safe Drinking Water Act. Included on the CCLs are numerous viruses, such as the double-stranded DNA adenoviruses. Double-stranded DNA viruses have been shown to be more resistant to UV disinfection when compared with other virus types (Gerba et al., 2002). Adenoviruses have been investigated and detected in wastewater in prior studies via conventional methods (Bofill-Mas et al., 2010, 2006; Cantalupo et al., 2011; Francy et al., 2012; Hewitt et al., 2011; Jacangelo et al., 2003; Katayama et al., 2008; Kuo et al., 2010; La Rosa et al., 2010; Petrinca et al., 2009; Prado et al., 2014, 2011; Schlindwein et al., 2010; Sedmak et al., 2005; Sima et al., 2011; Simmons et al., 2011; Simmons and Xagorarakis, 2011; Toze, 1999) and determined to be among the most abundant human viruses in WWTP effluent (Hata et al., 2013; La Rosa et al., 2010). It has been concluded that adenoviruses may serve as indicators for general viral contamination (La Rosa et al., 2010; Okoh et al., 2010).

In addition to human viral pathogens, bacteriophages may also have a significant impact on the natural water that receives treated effluent from a wastewater treatment facility. It has been shown that bacteriophages are strong regulators of microbial diversity within a WWTP (Shapiro et al., 2009). Bacteriophages can have an effect on the microbial community as well as on eukaryotic members of an ecosystem that rely on bacteria (Díaz-Muñoz and Koskella, 2014), making phages all the more important in maintaining ecological health of the aquatic environment. The majority of bacteriophages contain a DNA genome, making DNA

bacteriophages by far the most prevalent group of viruses impacting the microbial community (McGrath and Sinderen, 2007).

Both conventional activated sludge (AS) WWTPs and MBRs release pathogenic viruses into surface water (Kuo et al., 2010; Simmons et al., 2011; Simmons and Xagorarakis, 2011). It has been demonstrated that MBRs can remove several viruses (including adenovirus) at higher efficiencies compared to conventional AS utilities (Francis et al., 2012; Xagorarakis et al., 2014); for example, log removal of adenovirus ranged from 1.3 to 2.4 in conventional AS utilities (Haramoto et al., 2007; Hewitt et al., 2011; Katayama et al., 2008), whereas MBRs accomplished 3.4 to 5.6 log removals (Kuo et al., 2010; Simmons et al., 2011; Simmons and Xagorarakis, 2011). Though removal of specific viruses in these utilities has been investigated, there has been little application of metagenomics in studying the viral diversity of MBR effluent and how it compares to that of a conventional AS plant.

Next-generation DNA sequencing has recently been applied to study viral metagenomes (viromes) in environmental samples (Alhamlan et al., 2013; Gomez-Alvarez et al., 2012; Kim et al., 2015; Wommack et al., 2011) as well as at different stages of the wastewater treatment (Aw et al., 2014; Bibby et al., 2011; Bibby and Peccia, 2013; Cantalupo et al., 2011; Rosario et al., 2009; Tamaki et al., 2012). These methods have been shown to provide more conservative estimates of viral occurrence compared to the rates detected by qPCR (Bibby and Peccia, 2013). The advantage of metagenomics is that it allows comprehensive characterization of microbial communities. However, metagenomic methods do not assess infectivity and the sequence annotation is only as reliable and robust as the assembly methods and database used for analysis. Metagenomic analysis is presently only capable of identifying a fraction of the viruses present in the environment (Aw et al., 2014; Cantalupo et al., 2011).

Nonetheless, metagenomic methods are an effective tool for analyzing the microbial diversity of environmental samples. Though wastewater has been investigated with metagenomic methods, there has been little use of metagenomics to evaluate the microbial diversity of wastewater effluents, and to the authors' knowledge none comparing different types of wastewater utilities or the impact of disinfection. The specific objectives of this study are: 1) To investigate the diversity of human DNA viruses detected in effluents of MBR WWTPs equipped with membranes of different pore sizes; 2) To assess the diversity of DNA bacteriophages in MBR WWTP effluents; 3) To compare the diversity of DNA viruses in MBR WWTP effluents with that in a conventional WWTP effluent and a natural water; and 4) To investigate the impact of disinfection on DNA virus diversity in WWTP effluent.

2. Materials and Methods

2.1. Sampling locations

The selected MBR WWTPs were facilities located in Traverse City (Michigan, USA) and La Grande Motte (Languedoc-Roussillon-Midi-Pyrénées, France), which are both tourism destinations. Sampling was performed during the warmer seasons when the population of each location is increased due to the large number of vacationers. Effluent samples were collected at three wastewater treatment utilities. In Spring 2013, sampling was performed at the East Lansing WWTP (East Lansing, MI), which is a conventional activated sludge plant employing hypochlorite disinfection, and the Traverse City WWTP (Traverse City, MI), which employs MBR technology with ultrafiltration membranes of 0.04 μm nominal pore size and UV disinfection. A sample was taken from the treated effluent both before disinfection and after disinfection at each of the Michigan utilities. In summer 2015, sampling was performed at La

Grande Motte WWTP (La Grande Motte, France), which is also an MBR plant but employs microfiltration membranes of 0.45 μm nominal pore size and does not have an additional disinfection step. A sample was also taken from the treated effluent at this utility. Main operational parameters for these utilities are summarized in Table 1.

2.2. Sample collection

Approximately 300 L of sampled treated effluent was passed through a NanoCeram Virus Sampler filter (Argonide Corporation) at a rate of 11 to 12 L/min using a previously described method (Kuo et al., 2010). Samples from the two Michigan WWTPs were kept on ice and transported to Michigan State University (East Lansing, MI), while samples from France were kept on ice and transported to Université de Montpellier (Montpellier, France) for further processing.

2.3. Sample processing

All NanoCeram filters used to concentrate the treated effluent samples were eluted according to the standard method (US EPA, 2001) within 24 h of initial sampling. Briefly, a 1.5% w/v beef extract (0.05 M glycine, pH 9.0–9.5) solution was used as the eluent. The filters were submerged for a total of 2 min (two separate 1 min elutions) in filter housings with 1 L of beef extract added to the pressure vessel. After the beef extract was passed through each filter, pH of the 1 L of the eluate was adjusted to 3.5 ± 0.1 using 1 M HCl and flocculated for 30 min. Further concentration of the solution was performed by two stages of centrifugations for 15 min at 2500g and 4 °C. The supernatant was then decanted and the process was repeated until all the beef extract solution was centrifuged. The accumulated pellets were resuspended using 30 mL of 0.15 M sodium phosphate (pH 9.0–9.5) and mixed until the pellet was mostly dissolved. The pH was then adjusted to 9.0–9.5 using 1 M HCl. The solution was placed into a 50 mL centrifuge

tube and centrifuged for 10 min at 4 °C at 7000g. The supernatant was poured off into a separate 50 mL centrifuge tube, the pH was adjusted to 7.0–7.5 for stabilization of the virus particles, and the pellet was discarded. The supernatant was loaded into a 60 mL syringe and passed through a 0.22 µm sterilized filter for removal of bacteria, fungi and other contaminants. All samples were completely mixed and placed into 2 mL cryogenic tubes. Samples from France were shipped on dry ice to Michigan State University, where all samples were stored at –80 °C until further analysis.

2.4. Nucleic acid extraction

Viral DNA was extracted using a MagNA Pure Compact Instrument (Roche Applied Science) and a MagNA Pure Compact Nucleic Acid Isolation Kit according to the manufacturer's instructions. A 400 µL sample was loaded in the instrument and yielded an elution volume of 100 µL. DNase treatment is performed by the MagNA Pure Compact prior to extraction. The extracts were stored in a freezer at -20 °C. Following extraction the quantity of viral DNA extracts from all samples were verified for quality control purposes using the NanoDrop spectrophotometer (NanoDrop® ND-1000, Wilmington, DE).

2.5. Metagenomic analyses

Viral DNA extracts were sequenced on an Illumina platform (Illumina HiSeq, Roche Technologies) at the Research Technology Support Facility (RTSF) at Michigan State University. DNA-Seq libraries were prepared using the Rubicon Genomics ThruPLEX DNA-seq Kit. After preparation, libraries underwent quality control and were quantified using Qubit dsDNA, Caliper LabChipGX and Kapa Biosystems Library Quantification qPCR kit. The samples were pooled together and the pool was loaded on one lane of an Illumina HiSeq 2500 Rapid Run flow cell (v2) and sequencing was done in a 2x150bp paired end format using

Illumina Rapid SBS reagents. Base calling was performed by Illumina Real Time Analysis (RTA) v1.18.64 and output of RTA was demultiplexed and converted to FastQ format with Illumina Bcl2fastq v1.8.4. The raw sequencing files were then assessed for quality control using FastQC (Andrews, 2010). The flexible read trimming tool for Illumina NGS data called Trimmomatic was used for trimming the paired-end raw reads from the Illumina sequencer and removing adapters using ILLUMINACLIP (Bolger et al., 2014). The trimmed reads were assembled into contig files so as to reduce the chances of false positive detection using an iterative de Bruijn graph de novo assembler for short reads sequencing data with highly uneven sequencing depth called IDBA-UD using a minimum k-mer length of 40, maximum k-mer length of 120, and an interval of 10 (Peng et al., 2012).

2.6. MetaVir2 analyses

The assembled contig files for all samples were uploaded to the MetaVir2 web server for analysis. MetaVir2 is an online database designed to annotate viral metagenomics sequences (raw reads or assembled contigs) (Roux et al., 2014). The MetaVir server provides taxonomic affiliations of the viral sequence reads. Taxonomic composition is computed from a BLAST comparison with the RefSeq complete viral genomes protein sequences database from NCBI (release of 2014-07-10) using BLASTp. Open reading frames (ORFs) are predicted for each contig using MetaGeneAnnotator (Noguchi et al., 2008) and are compared to RefSeq through BLASTp, and each predicted translated ORF is affiliated to its best BLAST hit (if any), i.e. to the affiliation of the predicted protein with the highest BLAST score. The number of hits is defined as the occurrences of the input sequence in the database. Best hit ratio is defined as the number of hits for one category divided by total number of hits. Metavir only selects for sequences longer than 300 bp. A maximum E-value cutoff of 1E-5 was used. Sample

metagenomes are publicly available on MetaVir under the project titles “EastLansing/TraverseCity” (project IDs 3491, 3492, 3534, 3692) and “LaGrandeMotte” (project ID 7215).

2.7. Bowtie2 and SAMTools analyses

Following MetaVir analyses, the Bowtie2 and SAMTools modules were used to offer further confirmation of the presence of viruses of interest. Bowtie2 is a module that can align sequencing reads and SAMTools can provide further analyses of the alignment including statistical information regarding how well the two aligned sequences match. Thus, these tools were used to compare existing genomic data of viruses to the raw sequence data of the experimental samples. The Complete RefSeq Release of Viral and Viroid Sequences was downloaded from the NCBI website on 10 December 2015 and used in this analysis. This complete genome was aligned with the paired-end sequencing data of our experimental samples using Bowtie2. SAMTools was then used to provide output data regarding the alignment performed by Bowtie2 in the form of a percentage of coverage of base pairs of the selected genomes as well as statistical and quality control information regarding the alignment.

2.8. qPCR analyses

Additional confirmation using real-time quantitative PCR was performed on each sample using a Roche LightCycler 1.5 instrument (Roche Applied Sciences) for the detection of Human Adenovirus 40/41 using a previously described assay (Xagorarakis et al., 2007). Briefly, for each sample, a 20- μ L PCR mixture was created in triplicate containing 4 μ L of 5x LightCycler TaqMan Master Mix, 0.8 μ L of 10 μ M forward primer (final concentration, 400 nM), 0.4 μ L of each 10 μ M reverse primer (final concentration, 200 nM), 0.6 μ L of 10 μ M TaqMan probe (final concentration, 300 nM), 8.8 μ L of PCR-grade water, and 5 μ L of DNA extract. The real-time

PCR program used a denaturation step for 15 min at 95 °C, followed by an amplification step of 45 cycles at 95 °C for 15 s, 60 °C for 30 s, and 72 °C for 10 s, concluding with a cooling step at 40 °C for 30 s.

3. Results

Sequencing data were analyzed using MetaVir2. The number of contigs in the assembled files for all samples was in the range of 76970 to 256064 as computed by MetaVir. A summary of the affiliation of these contig sequences is presented in Table 2. A comparison of the ratio of affiliated sequences to unaffiliated sequences is shown in Figure 1. The effluent samples before disinfection for the two MBR utilities contained 19.88% (Traverse City WWTP) and 17.36% (La Grande Motte WWTP) affiliated sequences, whereas the effluent sample after disinfection from Traverse City WWTP contained 16.36% affiliated sequences. A similar proportion is seen in the East Lansing conventional WWTP utility, which contained 19.39% affiliated sequences in the effluent sample before disinfection and 11.60% affiliated sequences in the effluent sample after disinfection.

The taxonomic composition of all affiliated sequences for all samples is calculated by MetaVir. The large majority of the viruses detected in this manner are DNA viruses (>96% of affiliated sequences for all samples), which are results consistent with the analysis methods employed in the study targeting DNA viruses. The table in the supplemental information presents the detected DNA viruses (both double-stranded and single-stranded).

Of the detected DNA viruses, the significant majority corresponded to bacteriophages, as shown in Figure 2. Similar relative abundances of bacteriophages are observed between the treated effluent samples before disinfection from the MBR utilities (86.15% for Traverse City

WWTP and 82.13% for La Grande Motte WWTP of annotated DNA viral sequences), and this bacteriophage abundance is comparable to that observed in the East Lansing conventional WWTP (86.06% of annotated DNA viral sequences). The treated effluent samples after disinfection, however, are different for the two Michigan plants. Traverse City WWTP, which uses UV disinfection, had a bacteriophage relative abundance of 83.13% after the application of disinfection. East Lansing WWTP, which uses hypochlorite disinfection, showed a 64.55% relative abundance of bacteriophage.

Only a small percentage of detected viruses in the MBR samples before disinfection infect vertebrates, including those that infect humans; 0.53% in Traverse City WWTP and 0.56% in La Grande Motte WWTP of affiliated DNA sequences fall into this category. These numbers are similar to that (0.58%) for the sample from East Lansing WWTP. A smaller relative abundance of vertebrate viruses was present in the effluent before disinfection compared to the relative abundance after disinfection in each case (0.89% in Traverse City WWTP and 1.21% in East Lansing WWTP).

The first table in the supplementary information presents classification of virus families according to the NCBI taxonomy. The first column shows the viral orders detected and each subsequent column shows each sample. The table provides the number of hits for each sample for each viral order. For both MBR samples before disinfection, the order of bacteriophages *Caudovirales* was by far the most abundant; it accounted for 77.62% in Traverse City WWTP and 74.84% in La Grande Motte WWTP of DNA virus sequences. The sample from East Lansing WWTP again proved similar, with *Caudovirales* comprising 77.71% of DNA sequences. The abundance of *Caudovirales* for the after disinfection samples correspond to the

relative abundance of bacteriophage; the order comprises 74.52% of sequences in Traverse City WWTP but 60.67% of sequences in East Lansing WWTP.

Within the *Caudovirales* order, the family *Siphoviridae* was the most dominant in both MBR effluent samples before disinfection, comprising just over 40% of *Caudovirales* sequences in these samples, as well as in the conventional sample, as shown in Figure 3. The other families, *Myoviridae* and *Podoviridae*, also have similar relative abundances across all three effluent samples before disinfection. The most common bacterial hosts within the *Caudovirales* order are *Bacillus*, *Mycobacterium*, and *Pseudomonas*. Bacteria of human concern, including *Vibrio*, *Salmonella*, and *Escherichia*, also show a high number of hits for their respective phages across all samples. The second table in the supplementary information displays the number of hits as reported by MetaVir for each the *Caudovirales* order separated by bacterial host family.

The most abundant order of vertebrate viruses detected in the samples is *Herpesvirales*. The majority of sequences detected in this order belong to the family *Herpesviridae* which is known to infect humans and other mammals. Among the many mammalian herpesviruses detected were many human herpesviruses, including human herpesvirus 1 (herpes simplex virus-1), human herpesvirus 3 (varicella zoster virus, also known as chickenpox), human herpesvirus 4 (Epstein-Barr virus), and human herpesvirus 8. Other orders of vertebrate viruses detected by Metavir include the dsDNA virus orders *Adenoviridae*, and *Polyomaviridae*, as well as the ssDNA virus order *Parvoviridae*. To add confidence to the annotation of these pathogenic viruses, the contigs associated with *Herpesvirales* and *Adenoviridae* in each sample were further investigated to assess the annotation of individual genes within each contig using the contig annotations from MetaVir. MetaVir uses a BLAST comparison with the RefSeq complete viral genomes protein sequences database from NCBI (release of 2014-07-10) using BLASTp and a

maximum E-value cutoff of 1E-5. In each of the associated contigs, the gene with the best BLAST hit bitscore was a putative protein associated with the respective pathogenic virus (*Herpesvirales* or *Adenoviridae*); these bitscores ranged from 51.6 to 92 with a median of 70.5 for the annotated *Adenoviridae* contigs and 50.8 to 195 with a median of 71.2 for the human *Herpesvirales* contigs.

Quality control analysis from FastQC indicates that the raw sequencing files are of high quality. No sequences were flagged as poor quality in any of the samples. Statistics from Bowtie2 show similar results, with no reads in any sample having been flagged by Bowtie2 as failing quality control.

Results from Bowtie2 showed detection of human herpesvirus in all samples, as well as detection of human adenovirus in the Traverse City WWTP effluent after disinfection. Real-time quantitative PCR was used to assess the presence of human adenovirus 40/41 in all samples; detection mirrored the results as computed by MetaVir. HAdV40/41 was detected in the effluent after disinfection samples for East Lansing and Traverse City WWTPs at a concentration of 7.44×10^2 and 1.47×10^2 copies/L, respectively. This corresponds to the samples that returned multiple hits for *Adenoviridae* as reported by MetaVir.

4. Discussion

Despite thousands of affiliated sequences corresponding to known virus genomes in our samples, the vast majority of sequences (between 80% and 89%) were unaffiliated with any genome in the MetaVir database. While a large portion of these unaffiliated sequences could be associated with bacteria, as genetic material from organisms aside from viruses may have passed through the sample isolation procedure, a portion of the sequences in the samples may be derived

from uncharacterized viral genomes. These findings are consistent with previous studies utilizing next generation sequencing which also found significant proportions of unaffiliated sequences (Aw et al., 2014; Bibby and Peccia, 2013; Cantalupo et al., 2011; Tamaki et al., 2012). Consequently, annotation results are biased by the available sequences in the RefSeq database; this presents a significant hurdle in the use of metagenomic techniques. More robust genomic data and comparison databases are required to truly assess the full diversity of metagenomic samples.

The majority of viruses released in wastewater effluent are bacteriophages, which can affect the microbial community in receiving streams. Bacteriophages have been shown to be important factors in maintaining phage-host systems, with a sufficient concentration of a bacterial host required to begin the production of phage (Wommack and Colwell, 2000). A heavy influx of phages into a natural water system could potentially disrupt this phage-host system and microbial ecosystem. The figure in the supplementary information shows the most commonly detected bacteriophage hosts (*Pseudomonas*, *Bacillus*, *Mycobacterium*, and *Burkholderia*) for each of the samples. *Bacillus* is widely used as a model organism due to its prevalence in nature, whereas *Pseudomonas* is a group with significant metabolic diversity. All of these groups contain species known to be pathogenic to humans and have been widely studied, perhaps resulting in high representation in the MetaVir database and high detection rates in the samples.

In addition to bacteriophages, many different virus types exist among the human DNA viruses detected in this study: icosahedral, non-enveloped adenoviruses and polyomaviruses, enveloped herpesviruses, and the much larger enveloped poxviruses. Nonetheless, as shown in Figures 2 and 3, similar relative abundances were observed for all viruses (regardless of structure and type) in each of the wastewater effluents without disinfection. While the numbers of hits as

presented by MetaVir are low for human pathogens (including *Adenoviridae*, *Herpesviridae*, *Parvoviridae*, and *Polyomaviridae*), they are nonetheless detected indicating that metagenomic methods can be used to screen for human pathogens in wastewater effluent. Figure 4 describes the abundance of human viruses in each wastewater sample, with *Herpesviridae* being the most commonly detected viral order, significantly more than *Adenoviridae*, suggesting that *Herpesviridae*, like *Adenoviridae*, could also be used as an indicator of viral contamination. However, it should be noted that herpesviruses provide challenges in metagenomic analyses due to the fact that they have been shown to be capable of integrating into host genomes (Arbuckle et al., 2010), introducing uncertainty into the veracity of the metagenomic detection of herpesviruses in environmental samples. More thorough research should be conducted to assess the presence of herpesviruses in wastewater in order to use herpesviruses as indicators. These results are also consistent with prior metagenomic studies of human pathogens in studies that investigated samples from raw wastewater influent (Aw et al., 2014; Cantalupo et al., 2011) or biosolids and sewage sludge (Bibby et al., 2011; Bibby and Peccia, 2013). Another study that explored all stages of wastewater including effluent (Tamaki et al., 2012) reported data only on bacteriophages, the most commonly detected viruses.

This is the first study to compare viral diversity in the effluent of three wastewater utilities with different treatment technologies and disinfection techniques. Because the high cost of sequencing limited the analysis to only one sample per location, quantitative comparison among the samples is difficult. However, striking comparisons can be drawn when analyzing the diversity of the samples. Perhaps the most striking finding from these results is the similarity in diversity between the conventional activated sludge utility (East Lansing) and the two MBR plants (Traverse City WWTP and La Grande Motte WWTP). There is no clear difference among

the three facilities in terms of viral diversity in the samples before disinfection. The three effluent samples before disinfection all have practically identical diversity by viral host, and the diversity of the order *Caudovirales* is also very similar among the samples before disinfection.

Although MBRs have been shown to have higher virus removal efficiencies than conventional utilities, the diversity profile in these results indicate that MBRs are not more adept at removing particular kinds of viruses, but rather exhibit relatively equal removal of all DNA virus types. These results hold for both of the MBR WWTPs even though the membranes used at these two facilities have very different nominal pore sizes: 0.04 μm (ultrafiltration) in Traverse City MBR and 0.45 μm (microfiltration) in La Grande Motte MBR. However, the difference in the removal of human adenovirus from model feeds by these two membrane types has been shown to be significant (2.3 log for membranes with 0.04 μm pores and 0.7 logs for membranes with 0.45 μm pores) (Yin et al., 2015).

It is not until disinfection is performed that a divergence in diversity appears. The disinfection by hypochlorite at EL WWTP greatly reduces the relative abundance of bacteriophages in the effluent, whereas the UV disinfection at the TC WWTP does not affect viral diversity. Analyzing the three *Caudovirales* families, it appears that chlorination reduces the relative abundance of *Siphoviridae* and *Myoviridae*, allowing *Myoviridae* to become the most abundant *Caudovirales* family, whereas UV disinfection affects all families with relative equivalence. These results suggest that from a metagenomic standpoint the diversity of DNA viruses is insensitive to the choice of secondary treatment. Rather, the method of disinfection employed is the treatment process impacting the eventual viral diversity in the wastewater effluent.

Quality control analyses of the samples indicate high sequencing quality. No reads in any sample failed quality control or were flagged as poor quality. Similarly favorable quality control results were generated from the Bowtie2/SAMTools analyses. Nevertheless, it should be noted that due to only having one sample from each location, issues regarding reproducibility should be acknowledged. While sampling, preparation, and analysis methods were consistent across all samples, further investigation would be necessary to verify the reproducibility of these results.

5. Conclusions

Metagenomic analyses were performed using MetaVir2 as a basic tool to determine viral diversity of wastewater effluent in a conventional activated sludge utility and two membrane bioreactor utilities in the United States and France. This study is the first to evaluate diversity of DNA viruses in wastewater effluent using metagenomics and to compare the viral diversity for wastewater utilities of different types and in different locations. The study demonstrates that the majority of viruses released in wastewater effluent are bacteriophages, which can affect the microbial community of receiving streams. The study also reveals that both conventional activated sludge plant and membrane bioreactor utilities have a similar diversity of DNA viruses in their wastewater effluents prior to disinfection. Moreover, the type of disinfection process utilized has an impact in on the diversity of bacteriophages. Further research is required to determine how different disinfection methods impact bacteriophage diversity. This study also demonstrates that potentially human pathogenic DNA viruses are released into the environment via wastewater effluent and the most abundant potential human pathogen observed belongs to the taxonomic order *Herpesvirales*. The observed abundance of herpesviruses in the effluent of treatment utilities prompts further studies to investigate the fate of herpesviruses in wastewater.

Other potentially pathogenic human viruses detected in this study include *Adenoviridae*, *Parvoviridae*, and *Polyomaviridae*. Even with the thousands of affiliated sequences in this study, they remain a small fraction compared to the unaffiliated sequences. While metagenomic analysis has progressed significantly, more robust genomic databases are required to fully assess the biological diversity of a sample. Additionally, while the samples in this study were prepared using consistent methods, there is a significant degree of variation in methods used in published papers making it difficult to draw concrete comparison-based conclusions. To truly unlock the potential of metagenomics analysis, there must be standardization of sample preparation methods, sample analysis, as well as more robust available genomic data to accommodate analysis.

Acknowledgments

This material is based upon work supported in part by the National Science Foundation under Grant CBET- 1236393 and in part by a strategic partnership grant from the Center for European, Russian, and Eurasian Studies at Michigan State University. We would like to thank the managers of the wastewater treatment plants in East Lansing (Michigan, USA), Traverse City (Michigan, USA), and La Grande Motte (Languedoc-Roussillon-Midi-Pyrénées, France) for providing samples and information needed for this study. A very special thanks to the Research Technology Support Facility for assistance in sequencing and the Bioinformatics Center for Education and Productivity at the Institute for Cyber-Enabled Research at Michigan State University for the bioinformatics support and assistance provided.

References

- Alhamlan, F.S., Ederer, M.M., Brown, C.J., Coats, E.R., Crawford, R.L., 2013. Metagenomics-based analysis of viral communities in dairy lagoon wastewater. *J. Microbiol. Methods* 92, 183–188. doi:10.1016/j.mimet.2012.11.016
- Andrews, S., 2010. FastQC: A quality control tool for high throughput sequence data. Ref. Source.
- Arbuckle, J.H., Medveczky, M.M., Luka, J., Hadley, S.H., Luegmayr, A., Ablashi, D., Lund, T.C., Tolar, J., Meirleir, K.D., Montoya, J.G., Komaroff, A.L., Ambros, P.F., Medveczky, P.G., 2010. The latent human herpesvirus-6A genome specifically integrates in telomeres of human chromosomes in vivo and in vitro. *Proc. Natl. Acad. Sci.* 107, 5563–5568. doi:10.1073/pnas.0913586107
- Aw, T.G., Howe, A., Rose, J.B., 2014. Metagenomic approaches for direct and cell culture evaluation of the virological quality of wastewater. *J. Virol. Methods* 210, 15–21. doi:10.1016/j.jviromet.2014.09.017
- Bibby, K., 2013. Metagenomic identification of viral pathogens. *Trends Biotechnol.* 31, 275–279. doi:10.1016/j.tibtech.2013.01.016
- Bibby, K., Peccia, J., 2013. Identification of Viral Pathogen Diversity in Sewage Sludge by Metagenome Analysis. *Environ. Sci. Technol.* 47, 1945–1951. doi:10.1021/es305181x
- Bibby, K., Viau, E., Peccia, J., 2011. Viral metagenome analysis to guide human pathogen monitoring in environmental samples. *Lett. Appl. Microbiol.* 52, 386–392. doi:10.1111/j.1472-765X.2011.03014.x
- Bofill-Mas, S., Albinana-Gimenez, N., Clemente-Casares, P., Hundesa, A., Rodriguez-Manzano, J., Allard, A., Calvo, M., Girones, R., 2006. Quantification and Stability of Human Adenoviruses and Polyomavirus JCPyV in Wastewater Matrices. *Appl. Environ. Microbiol.* 72, 7894–7896. doi:10.1128/AEM.00965-06
- Bofill-Mas, S., Rodriguez-Manzano, J., Calgua, B., Carratala, A., Girones, R., 2010. Newly described human polyomaviruses Merkel Cell, KI and WU are present in urban sewage and may represent potential environmental contaminants. *Virol. J.* 7, 141. doi:10.1186/1743-422X-7-141
- Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 170. doi:10.1093/bioinformatics/btu170
- Cantalupo, P.G., Calgua, B., Zhao, G., Hundesa, A., Wier, A.D., Katz, J.P., Grabe, M., Hendrix, R.W., Girones, R., Wang, D., Pipas, J.M., 2011. Raw Sewage Harbors Diverse Viral Populations. *mBio* 2, e00180-11. doi:10.1128/mBio.00180-11
- Díaz-Muñoz, S.L., Koskella, B., 2014. Chapter Four - Bacteria–Phage Interactions in Natural Environments, in: Gadd, S.S. and G.M. (Ed.), *Advances in Applied Microbiology*. Academic Press, pp. 135–183.
- Francy, D.S., Stelzer, E.A., Bushon, R.N., Brady, A.M.G., Williston, A.G., Riddell, K.R., Borchardt, M.A., Spencer, S.K., Gellner, T.M., 2012. Comparative effectiveness of membrane bioreactors, conventional secondary treatment, and chlorine and UV disinfection to remove microorganisms from municipal wastewaters. *Water Res.* 46, 4164–4178. doi:10.1016/j.watres.2012.04.044
- Gerba, C.P., Gramos, D.M., Nwachuku, N., 2002. Comparative Inactivation of Enteroviruses and Adenovirus 2 by UV Light. *Appl. Environ. Microbiol.* 68, 5167–5169. doi:10.1128/AEM.68.10.5167-5169.2002

- Gomez-Alvarez, V., Revetta, R.P., Domingo, J.W.S., 2012. Metagenomic Analyses of Drinking Water Receiving Different Disinfection Treatments. *Appl. Environ. Microbiol.* 78, 6095–6102. doi:10.1128/AEM.01018-12
- Haramoto, E., Katayama, H., Oguma, K., Ohgaki, S., 2007. Quantitative analysis of human enteric adenoviruses in aquatic environments. *J. Appl. Microbiol.* 103, 2153–2159. doi:10.1111/j.1365-2672.2007.03453.x
- Hata, A., Kitajima, M., Katayama, H., 2013. 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.* 114, 545–554. doi:10.1111/jam.12051
- Hewitt, J., Leonard, M., Greening, G.E., Lewis, G.D., 2011. Influence of wastewater treatment process and the population size on human virus profiles in wastewater. *Water Res.* 45, 6267–6276. doi:10.1016/j.watres.2011.09.029
- Jacangelo, J.G., Loughran, P., Petrik, B., Simpson, D., McIlroy, C., 2003. Removal of enteric viruses and selected microbial indicators by UV irradiation of secondary effluent. *Water Sci. Technol.* 47, 193–198.
- Katayama, H., Haramoto, E., Oguma, K., Yamashita, H., Tajima, A., Nakajima, H., Ohgaki, S., 2008. One-year monthly quantitative survey of noroviruses, enteroviruses, and adenoviruses in wastewater collected from six plants in Japan. *Water Res.* 42, 1441–1448. doi:10.1016/j.watres.2007.10.029
- Kim, Y., Aw, T.G., Teal, T.K., Rose, J.B., 2015. Metagenomic Investigation of Viral Communities in Ballast Water. *Environ. Sci. Technol.* 49, 8396–8407. doi:10.1021/acs.est.5b01633
- Kuo, D.H.-W., Simmons, F.J., Blair, S., Hart, E., Rose, J.B., Xagorarakis, I., 2010. Assessment of human adenovirus removal in a full-scale membrane bioreactor treating municipal wastewater. *Water Res.* 44, 1520–1530. doi:10.1016/j.watres.2009.10.039
- La Rosa, G., Pourshaban, M., Iaconelli, M., Muscillo, M., 2010. Quantitative real-time PCR of enteric viruses in influent and effluent samples from wastewater treatment plants in Italy. *Ann. Dell'Istituto Super. Sanità* 46, 266–273. doi:DOI: 10.4415/ANN_10_03_07
- McGrath, S.M., Sinderen, D. van, 2007. *Bacteriophage: Genetics and Molecular Biology*. Horizon Scientific Press.
- Noguchi, H., Taniguchi, T., Itoh, T., 2008. MetaGeneAnnotator: Detecting Species-Specific Patterns of Ribosomal Binding Site for Precise Gene Prediction in Anonymous Prokaryotic and Phage Genomes. *DNA Res.* 15, 387–396. doi:10.1093/dnares/dsn027
- Okoh, A.I., Sibanda, T., Gusha, S.S., 2010. Inadequately Treated Wastewater as a Source of Human Enteric Viruses in the Environment. *Int. J. Environ. Res. Public. Health* 7, 2620–2637. doi:10.3390/ijerph7062620
- Peng, Y., Leung, H.C.M., Yiu, S.M., Chin, F.Y.L., 2012. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28, 1420–1428. doi:10.1093/bioinformatics/bts174
- Petrinca, A. r., Donia, D., Pierangeli, A., Gabrieli, R., Degener, A. m., Bonanni, E., Diaco, L., Cecchini, G., Anastasi, P., Divizia, M., 2009. Presence and environmental circulation of enteric viruses in three different wastewater treatment plants. *J. Appl. Microbiol.* 106, 1608–1617. doi:10.1111/j.1365-2672.2008.04128.x
- Prado, T., Gaspar, A.M.C., Miagostovich, M.P., 2014. Detection of enteric viruses in activated sludge by feasible concentration methods. *Braz. J. Microbiol.* 45, 343–349. doi:10.1590/S1517-83822014000100049

- Prado, T., Silva, D.M., Guilayn, W.C., Rose, T.L., Gaspar, A.M.C., Miagostovich, M.P., 2011. Quantification and molecular characterization of enteric viruses detected in effluents from two hospital wastewater treatment plants. *Water Res.* 45, 1287–1297. doi:10.1016/j.watres.2010.10.012
- Rosario, K., Nilsson, C., Lim, Y.W., Ruan, Y., Breitbart, M., 2009. Metagenomic analysis of viruses in reclaimed water. *Environ. Microbiol.* 11, 2806–2820. doi:10.1111/j.1462-2920.2009.01964.x
- Roux, S., Tournayre, J., Mahul, A., Debroas, D., Enault, F., 2014. Metavir 2: new tools for viral metagenome comparison and assembled virome analysis. *BMC Bioinformatics* 15, 76. doi:10.1186/1471-2105-15-76
- Schlindwein, A.D., Rigotto, C., Simões, C.M.O., Barardi, C.R.M., 2010. Detection of enteric viruses in sewage sludge and treated wastewater effluent. *Water Sci. Technol.* 61, 537–544. doi:10.2166/wst.2010.845
- Sedmak, G., Bina, D., MacDonald, J., Couillard, L., 2005. Nine-Year Study of the Occurrence of Culturable Viruses in Source Water for Two Drinking Water Treatment Plants and the Influent and Effluent of a Wastewater Treatment Plant in Milwaukee, Wisconsin (August 1994 through July 2003). *Appl. Environ. Microbiol.* 71, 1042–1050. doi:10.1128/AEM.71.2.1042-1050.2005
- Shapiro, O.H., Kushmaro, A., Brenner, A., 2009. Bacteriophage predation regulates microbial abundance and diversity in a full-scale bioreactor treating industrial wastewater. *ISME J.* 4, 327–336. doi:10.1038/ismej.2009.118
- Sidhu, J.P.S., Hanna, J., Toze, S.G., 2008. Survival of enteric microorganisms on grass surfaces irrigated with treated effluent. *J. Water Health* 6, 255–262. doi:10.2166/wh.2008.029
- Sima, L.C., Schaeffer, J., Saux, J.-C.L., Parnaudeau, S., Elimelech, M., Guyader, F.S.L., 2011. Calicivirus Removal in a Membrane Bioreactor Wastewater Treatment Plant. *Appl. Environ. Microbiol.* 77, 5170–5177. doi:10.1128/AEM.00583-11
- Simmons, F.J., Kuo, D.H.-W., Xagorarakis, I., 2011. Removal of human enteric viruses by a full-scale membrane bioreactor during municipal wastewater processing. *Water Res.* 45, 2739–2750. doi:10.1016/j.watres.2011.02.001
- Simmons, F.J., Xagorarakis, I., 2011. Release of infectious human enteric viruses by full-scale wastewater utilities. *Water Res.* 45, 3590–3598. doi:10.1016/j.watres.2011.04.001
- Tamaki, H., Zhang, R., Angly, F.E., Nakamura, S., Hong, P.-Y., Yasunaga, T., Kamagata, Y., Liu, W.-T., 2012. Metagenomic analysis of DNA viruses in a wastewater treatment plant in tropical climate. *Environ. Microbiol.* 14, 441–452. doi:10.1111/j.1462-2920.2011.02630.x
- Toze, S., 1999. PCR and the detection of microbial pathogens in water and wastewater. *Water Res.* 33, 3545–3556. doi:10.1016/S0043-1354(99)00071-8
- Toze, S., 1997. Microbial Pathogens in Wastewater: Literature review for urban water systems multi-divisional research program. CSIRO Land and Water.
- US EPA, 2006. Prepublication of the ground water rule federal register notice.
- US EPA, 2001. Manual of Methods for Virology (Chapter 14).
- Wommack, K.E., Colwell, R.R., 2000. Virioplankton: Viruses in Aquatic Ecosystems. *Microbiol. Mol. Biol. Rev.* 64, 69–114. doi:10.1128/MMBR.64.1.69-114.2000
- Wommack, K.E., Srinivasiah, S., Liles, M.R., Bhavsar, J., Bench, S., Williamson, K.E., Polson, S.W., 2011. Metagenomic Contrasts of Viruses in Soil and Aquatic Environments, in:

Metagenomic Contrasts of Viruses in Soil and Aquatic Environments. John Wiley & Sons.

Xagorarakis, I., Kuo, D.H.-W., Wong, K., Wong, M., Rose, J.B., 2007. Occurrence of Human Adenoviruses at Two Recreational Beaches of the Great Lakes. *Appl. Environ. Microbiol.* 73, 7874–7881. doi:10.1128/AEM.01239-07

Xagorarakis, I., Yin, Z., Svambayev, Z., 2014. Fate of Viruses in Water Systems. *J. Environ. Eng.* 140, 4014020. doi:10.1061/(ASCE)EE.1943-7870.0000827

Yin, Z., Tarabara, V.V., Xagorarakis, I., 2015. Human adenovirus removal by hollow fiber membranes: Effect of membrane fouling by suspended and dissolved matter. *J. Membr. Sci.* 482, 120–127. doi:10.1016/j.memsci.2015.02.028

Tables and Figures

Table 1: Wastewater treatment plant characteristics.

	EAST LANSING Michigan, USA	TRAVERSE CITY Michigan, USA	LA GRANDE-MOTTE Languedoc-Roussillon- Midi-Pyrénées, France
Wastewater treatment process	Conventional Activated Sludge (CAS)	Membrane Biological Reactor (MBR)	Membrane Biological Reactor (MBR)
Sludge retention time, days	14	7.58	46.5
Capacity, MGD	18.8	17.0	3.2
Average flow, MGD	13.4	8.5	2.6
Disinfection	Hypochlorite	UV	None
Nominal pore size, μm	n/a	0.04	0.45
Sampling dates	Spring 2013	Spring 2013	Summer 2015

Table 2: Metagenome analysis statistics for viral samples (from MetaVir).

Metagenome Name	Number of contigs	Affiliated sequences	Unaffiliated sequences	No. of genes affiliated	No. of genes predicted
East Lansing Wastewater Treatment Plant (EL)					
Effluent Before Disinfection	151994	29479	122515	35856	258135
Effluent After Disinfection	256064	29706	226356	33975	365870
Traverse City Wastewater Treatment Plant (TC)					

Effluent Before Disinfection	151992	30182	121610	36781	258140
Effluent After Disinfection	197517	32316	165201	39649	309715
La Grande Motte Wastewater Treatment Plant (LGM)					
Treated Effluent (No Disinfection)	76970	13364	63606	15072	124682

Figure 1: Metagenome summary (from MetaVir).

Figure 2: Relative abundance (number of affiliated sequences for virus host group divided by total number of affiliated sequences for the sample) for each sample by virus host group (from MetaVir).

Figure 3: Relative abundance of Caudovirales order by family (number of affiliated sequences for each family divided by total number of affiliated Caudovirales sequences) (from MetaVir).

Figure 4: Relative number of hits for human viral orders for each sample, measured as the ratio of number of hits for the viral order to the total number of hits in the sample (from MetaVir).

**Effluent
Before
Disinfection**

Bacteriophages

Vertebrate viruses
(incl. humans)

Plant viruses

Amoeba viruses

Invertebrate viruses

Archaea viruses

**Effluent
After
Disinfection**

**East Lansing:
Hypchlorite
Disinfection**

**Traverse City:
UV Disinfection**