Journal of Virological Methods xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Journal of Virological Methods

journal homepage: www.elsevier.com/locate/jviromet

Metagenomic approaches for direct and cell culture evaluation of the virological quality of wastewater

Tiong Gim Aw[a,*], Adina Howe[b,c], Joan B. Rose[a]

[a] Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI 48824, United States
[b] Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI 48824, United States
[c] Computing, Environment, and Life Sciences, Argonne National Laboratory, Argonne, IL 60439, United States

Article history:
Received 27 June 2014
Received in revised form
15 September 2014
Accepted 16 September 2014
Available online xxx

Keywords:
Viral metagenomics
Next generation sequencing
Wastewater
Bacteriophages
Enteric viruses

A B S T R A C T

Genomic-based molecular techniques are emerging as powerful tools that allow a comprehensive characterization of water and wastewater microbiomes. Most recently, next generation sequencing (NGS) technologies which produce large amounts of sequence data are beginning to impact the field of environmental virology. In this study, NGS and bioinformatics have been employed for the direct detection and characterization of viruses in wastewater and of viruses isolated after cell culture. Viral particles were concentrated and purified from sewage samples by polyethylene glycol precipitation. Viral nucleic acid was extracted and randomly amplified prior to sequencing using Illumina technology, yielding a total of 18 million sequence reads. Most of the viral sequences detected could not be characterized, indicating the great viral diversity that is yet to be discovered. This sewage virome was dominated by bacteriophages and contained sequences related to known human pathogenic viruses such as adenoviruses (species B, C and F), polyomaviruses JC and BK and enteroviruses (type B). An array of other animal viruses was also found, suggesting unknown zoonotic viruses. This study demonstrated the feasibility of metagenomic approaches to characterize viruses in complex environmental water samples.

1. Introduction

Wastewater infrastructure is central to sustainable development in every modern society and provides a treated effluent that decreases pollution and waterborne disease risks prior to discharge and in many cases for reuse. However, waterborne disease outbreaks and beach closures still occur today, often due to poor wastewater infrastructure and inadequate treatment in both developed and developing regions of the world (Chalmers et al., 2010; Laine et al., 2011). Until recently, the microbial safety of water and wastewater has been determined exclusively by testing for bacterial indicators (e.g. coliforms and *Escherichia coli*). However, there is evidence that traditional bacterial indicators fail to adequately protect public health and often fail to indicate the presence of pathogenic viruses and parasitic protozoa which can survive for much longer periods than the bacterial indicators in water and wastewater environments (Scott et al., 2002; Fong and Lipp, 2005).

All living entities have at least one virus associated with them (Ackermann, 2003). Viruses are shed in extremely high numbers in fecal matter (including bacteriophages, enteric viruses from humans and animals), are found in high numbers in wastewater, and can contaminate water sources used for drinking water, recreational activities, aquaculture and irrigation (Fong and Lipp, 2005). Viruses have been shown to persist for extended periods in the water environment and are also more resistant to removal by wastewater treatment systems compared to bacteria (Allwood et al., 2003; Gomila et al., 2008). The host specificity of viruses also suggests that they could be promising library-independent tools to determine the microbial sources of fecal contamination in water environments (e.g. human versus bovine). A better understanding of viruses in wastewater will help to better inform treatment system stability and performance to achieve safe effluents.

Technology for microbial detection is changing constantly and keeping up with the latest developments can be challenging, particularly when monitoring the quality of water and wastewater. Current detection approaches for viruses using polymerase chain reaction (PCR) and reverse transcriptase (RT)-PCR (for the detection of RNA viruses) are limited by known sequence information, and investigators must select the range of viruses to be considered in a given assay (mainly a single-target approach), which does

* Corresponding author at: Department of Fisheries and Wildlife, 480 Wilson Road, 13 Natural Resources, Michigan State University, East Lansing, MI 48824, United States. Tel.: +1 517 432 1930; fax: +1 517 353 9807.
E-mail address: tgaw@msu.edu (T.G. Aw).

http://dx.doi.org/10.1016/j.jviromet.2014.09.017
0166-0934/

not allow for the analysis of all viruses in a sample (Girones et al., 2010). Molecular biology is being revolutionized by whole-genome sequencing of individual microorganisms, as well as entire microbial communities, a field known as metagenomics (Edwards and Rohwer, 2005; Mokili et al., 2012). Metagenomics is a sequence-based analysis of mixtures of genetic material corresponding to microbial communities (fungi, bacteria and viruses) and other living entities, which can be recovered directly from environmental samples. Leveraging high throughput next generation sequencing (NGS) technologies, metagenomics has become a powerful tool for characterizing viruses because viral nucleic acids are isolated and sequenced directly from environmental samples without requiring a priori knowledge of these communities (Roux et al., 2012; Holmfeldt et al., 2013). This advancement provides a novel opportunity to assess microbial water and wastewater quality characteristics by using genetic data from the viruses as a whole rather than only focusing on a few specific, well-known viral groups and viruses that are cultivatable (Aw and Rose, 2012). Although the rapid development in metagenomics has opened entire new perspectives to microbial communities in different environments, only a few studies have used metagenomics approaches to characterize viruses in wastewater environments (Cantalupo et al., 2011; Tamaki et al., 2012; Bibby and Peccia, 2013a,b). To date, most metagenomic studies of water and wastewater have focused mainly on bacteria or are limited to the lower sequencing yield of earlier technologies (e.g., 454 pyrosequencing) (Cantalupo et al., 2011; Kristiansson et al., 2011; Ye and Zhang, 2011; Tamaki et al., 2012).

In this paper, the use of Illumina high-throughput sequencing platform coupled with bioinformatics analysis to detect and characterize viral diversity in untreated sewage is described. In addition to identifying viruses directly in wastewater, viruses were also isolated in cell culture and characterized using the NGS. Challenges of applying the NGS technologies as a tool for evaluating the virological quality of wastewater are also discussed.

2. Materials and methods

2.1. Concentration and purification of viruses from sewage

Two untreated sewage samples (1 l each) were collected from the East Lansing Wastewater Treatment Plant, Michigan. Viral particles were concentrated and purified by polyethylene glycol (PEG) precipitation as previously described (Shieh et al., 1995). Briefly, the sewage sample (pH adjusted to 7.2) was mixed with 8% of PEG 8000 (Promega) and 0.3 M NaCl. The mixture was incubated at 4 °C for about 18 h before centrifugation at 6700 × *g* for 30 min. The pellet was dissolved in 20 ml of phosphate buffer saline (PBS). The sample was further extracted with an equal volume of chloroform and the supernatant was collected by centrifugation at 1700 × *g* for 30 min. The concentrated supernatant containing viruses was passed through 0.22-μm filters to remove any remaining bacterial and eukaryotic cells. The concentrates were combined and stored at −80 °C until processed on cell culture and viral nucleic acid extraction as described below. A 5-ml aliquot of sewage concentrate was used for viral cell culture. Virus particles from the remaining sewage concentrate were further concentrated to approximately 1 ml by centrifugation through an Amicon Ultra 100kDA centrifugation column (Millipore, Billerica, MA). An outline of the methods for generating viral metagenomes from sewage is shown in Fig. 1.

2.2. Viral cell culture

Ten-fold dilution of sewage concentrate was inoculated onto three individual tissue culture flasks containing a confluent monolayer of A549 human epithelial lung carcinoma cells (American Type Culture Collection, ATCC CCL-185). The A549 cell line was used because it is susceptible to a wide range of enteric viruses including enteroviruses and adenoviruses (Lee et al., 2004). Adenovirus type 2 (ATCC VR-846) was used as the positive control for A549 cell culture. Three negative control flasks (with PBS buffer) were used for each cell culture assay. Tissue culture flasks were incubated at 37 °C with 5% CO_2 and were examined for cytopathic effects (CPE) daily. Viral infections were allowed to proceed until the onset of CPE, which is typically within three days for sewage samples. To recover the viruses, the cells were freeze-thawed three times to disrupt the integrity of cells and release the viral particles into the supernatant. A second passage on a fresh monolayer was performed to confirm the CPE. The supernatant was filtered through 0.22-μm filters to remove any host cells. Virus particles were further concentrated to approximately 1 ml by centrifugation through an Amicon Ultra 100 kDA centrifugation column (Millipore, Billerica, MA).

2.3. Viral nucleic acid extraction, amplification and sequencing

DNase (Roche) treatment was used to remove any free nucleic acids from final viral concentrates with and without cell culture. After DNase treatment, each viral concentrate was split into two equal subsamples, of which one was used for DNA viruses and the other was used for RNA viruses. For DNA viruses, nucleic acids were extracted from viral particles using QIAGEN QIAamp MinElute Virus Spin Kit, followed by random amplification in triplicate reactions using GenomiPhi V2 DNA amplification Kit (GE Healthcare) according to the manufacturer's instructions. For RNA viruses, nucleic acids were extracted using QIAGEN RNeasy Plus Mini Kit, followed by randomly primed RNA amplification in triplicate reactions using a TransPlex Complete Whole Transcriptome Amplification Kit (WTA2, Sigma–Aldrich) according to the manufacturer's instructions. The amplification products were purified. Libraries from each sample were prepared using the Illumina TruSeq kit with unique Multiplex ID tags (barcodes). The samples were pooled and sequencing was performed at the Research Technology Support Facility, Michigan State University. 75 bp paired-end reads were generated on one lane of the Illumina Genome Analyzer IIx sequencer.

2.4. Bioinformatics

Prior to assembly and annotation of the metagenomic dataset, the quality of the Illumina sequencing data was checked using FastQC and quality trimmed (or filtered) using prinseq_lite (Schmieder and Edwards, 2011). The following parameters were used for quality trimming: overall mean Q score of 25; maximum number of ambiguous (N) basecalls of 2; 3′-end trimmed of low quality bases, mean Q score of 20 over a 5 base window; and reads less than 30 nucleotides after trimming were discarded. Filtered read files were processed to match pairs where both reads passed filtering; singleton reads were saved in a separate file. Longer, assembled sequences enable more specific and confident annotations in downstream data analysis (Wommack et al., 2008). Both the paired and singleton reads were de novo assembled into a longer contiguous sequence (contig) using the short read assembly algorithm, Velvet 1.2.06 (Zerbino, 2010). Velvet modules were run with different k-mer lengths (35–63 bp) with the following parameters: Velveth *K* = 35–63, -short, -shortPaired and velvetg-exp_cov auto, -cov_cutoff 0, -scaffolding no. The assembly for each sample with the highest N50 and the maximum percentage of reads assembled was used for annotation. Assembled contigs larger than 200 bp were queried against the National Center for Biotechnology Information (NCBI) non-redundant Viral Reference Sequence (RefSeq) database

Please cite this article in press as: Aw, T.G., et al., Metagenomic approaches for direct and cell culture evaluation of the virological quality of wastewater. J. Virol. Methods (2014), http://dx.doi.org/10.1016/j.jviromet.2014.09.017