Environmental Science & Technology

Article

pubs.acs.org/est

# Metagenomic Investigation of Viral Communities in Ballast Water

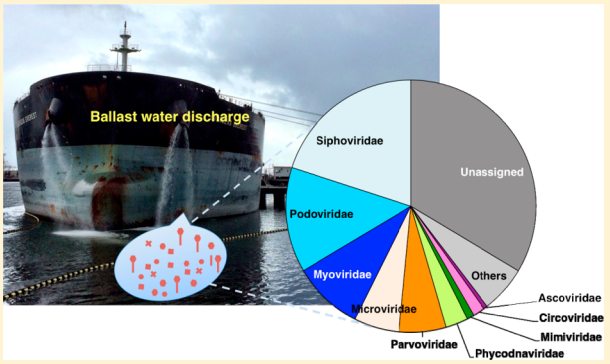
Yiseul Kim,*[,†] Tiong Gim Aw,[‡] Tracy K. Teal,[†] and Joan B. Rose[†,‡]

[†]Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, Michigan 48824, United States

[‡]Department of Fisheries and Wildlife, Michigan State University, East Lansing, Michigan 48824, United States

Supporting Information

**ABSTRACT:** Ballast water is one of the most important vectors for the transport of non-native species to new aquatic environments. Due to the development of new ballast water quality standards for viruses, this study aimed to determine the taxonomic diversity and composition of viral communities (viromes) in ballast and harbor waters using metagenomics approaches. Ballast waters from different sources within the North America Great Lakes and paired harbor waters were collected around the Port of Duluth—Superior. Bioinformatics analysis of over 550 million sequences showed that a majority of the viral sequences could not be assigned to any taxa associated with reference sequences, indicating the lack of knowledge on viruses in ballast and harbor waters. However, the assigned viruses were dominated by double-stranded DNA phages, and sequences associated with potentially emerging viral pathogens of fish and shrimp were detected with low amino acid similarity in both ballast and harbor waters. Annotation-independent comparisons showed that viromes were distinct among the Great Lakes, and the Great Lakes viromes were closely related to viromes of other cold natural freshwater systems but distant from viromes of marine and human designed/managed freshwater systems. These results represent the most detailed characterization to date of viruses in ballast water, demonstrating their diversity and the potential significance of the ship-mediated spread of viruses.




## INTRODUCTION

Ballast water has been used as an essential component of efficient and safe shipping operations dating back to the 1870s.[1] Globally, as high as 12 billion tons of ballast water are transported and exchanged by more than 45 000 ocean-going vessels each year.[2] In the United States (U.S.) alone, an estimated 79 million tons of ballast water are annually discharged into coastal areas from international ports.[3] A significant volume, 6.6 million tons of ballast water per year, is also discharged into the freshwaters of the North America Great Lakes.[4] It is known that biological materials are discharged and exchanged with ballast water, and therefore this global and widely used practice brings with it potential ecological, economic, and public health problems including invasive species and the disruption of native ecosystems in major ports worldwide.[5,6]

The U.S. Environmental Protection Agency (U.S. EPA) has established ballast water discharge standards that align with the International Maritime Organization (IMO) and U.S. Coast Guard (U.S.C.G.) rules based on organism size classes.[7] At the state level, California has the most stringent ballast water management criteria with a state-specific standard of zero detectable living organisms for all organism size classes including virus-like particles (VLPs) in the final discharge going into effect January 2020.[8] Efforts to comply with increasingly stringent regulatory demands will, however, be one of the most significant challenges for the shipping industry over the next few years as technologies for ballast water treatment are still in the research and development phase.[5]

A few studies have shown that bacteria and VLPs numerically dominate ballast water biota and are transferred globally in greater numbers than any other size classes of organisms.[9−13] For example, the ballast water of vessels arriving at Chesapeake Bay on the U.S. East Coast contained an average of $8.3 \times 10^8$ bacteria/L and $7.4 \times 10^9$ VPLs/L.[9] An interesting follow-up study showed that an estimated $3.9 \times 10^{18}$ bacteria and $6.8 \times 10^{19}$ VLPs (assuming a survival rate of 56% and applying estimates of ship traffic) in ballast water were annually discharged to and survived in the lower Chesapeake Bay.[11]

Viruses are small infectious agents that exist through parasitic relationships with a wide range of hosts, including humans, animals, bacteria, fungi, and plants—some of which are highly host specific. Viruses are of special interest because they are thought to be the most abundant and diverse biological entities on Earth with as many as $10^{10}$ VLPs/L of seawater, approximately 10 times more than the number of prokaryotes.[14,15] Moreover, viruses influence the structure and diversity of microbial communities by infection and lysis of host communities.[16] Examples of specific viruses that have been

Received: April 13, 2015
Revised: June 10, 2015
Accepted: June 24, 2015
Published: June 24, 2015

DOI: 10.1021/acs.est.5b01633
Environ. Sci. Technol. 2015, 49, 8396−8407

Table 1. Identification, Description, and Collection Information for the Ballast and Harbor Water Samples

| sample[a] | sample type | ballast water source port[b] | ballast water discharge port (sampling location) | voyage duration[c] (sampling date) | sampling method | ballast water treatment |
|---|---|---|---|---|---|---|
| AB | ballast water | Toledo (Lake Erie) | Duluth, Terminal C (Lake Superior) | 3 days (5/15/13) | ballast water pipeline | untreated |
| BB | ballast water | Open lake (Lake Ontario) | Duluth, Terminal A (Lake Superior) | 7 days (5/9/13) | ballast water pipeline | untreated |
| BH | harbor water | | Duluth, Terminal A (Lake Superior) | – (5/9/13) | bucket with rope | |
| IB | ballast water | Essexville (Lake Huron) | Duluth, Terminal C (Lake Superior) | 2 days (5/10/13) | ballast water pipeline | untreated |
| IH | harbor water | | Duluth, Terminal C (Lake Superior) | – (5/10/13) | bucket with rope | |
| MB | ballast water | Burns Harbor (Lake Michigan) | Duluth, Terminal B (Lake Superior) | 4–8 days (5/10/13) | sounding pipe | untreated |
| MH | harbor water | | Duluth, Terminal B (Lake Superior) | – (5/10/13) | bucket with rope | |
| PB | ballast water | Hamilton (Lake Ontario) | Duluth, Terminal C (Lake Superior) | 4 days (5/14/13) | ballast water pipeline | untreated |

[a]The ballast water was designated BB, IB, and MB, and their matching harbor water as BH, IH, and MH. [b]Ballast water source ports were where the vessels sampled for ballast water had undergone ballast water exchange prior to their arrivals at the Port of Duluth–Superior. [c]Voyage duration was the difference in days between date of ballast water uptake from the ballast water source port and date of ballast water sampling from the Port of Duluth–Superior.

identified as invasive species introduced via ships' ballast water are marine cyanophage and Viral Hemorrhagic Septicemia Virus (VHSV) in the Great Lakes and Infectious Salmon Anemia Virus (ISAV) in Chile, Europe, and Northwest Atlantic.[17−20]

While the introduction of a complex assemblage of microorganisms through ballast water is a growing concern globally, the microbial diversity of ballast water remains largely unknown. Moreover, the potential ecological impacts and public health risk are not well understood. The primary reasons for the lack of knowledge about the microbial communities, particularly viruses, in the ballast water system is related to the difficulty in collection of ballast water for virus analysis, the specificity of the viral-host systems used for identification, and the lack of universal genetic markers for viruses such as the 16S rRNA gene used for prokaryotes.[21,22] However, the development of modern genomic tools and the emergence of high-throughput sequencing (HTS) technologies have overcome some of the limitations of classical methods for virus detection and characterization. Increasing capacity in HTS and improvements in bioinformatics analyses have had a major impact on the expansion of virus detection and characterization and discovery of novel viruses, including zooplankton and phytoplankton viruses.[23,24]

These advances have allowed us to learn more about taxonomic diversity and composition of viral communities (viromes) in water. Our ability to characterize the ballast water virome of ships in the Great Lakes is of particular interest as this is an economically important shipping area that is also very susceptible to external influences on its native freshwater communities and therefore a good model system for the study of viral transport. The Great Lakes has a unique shipping system in which ships can move through the St. Lawrence Seaway linking North America with ports throughout the world. Consequently, this large freshwater basin is particularly vulnerable to invasive species and has been invaded by more than 180 non-native species within the past two centuries.[25] To the best of our knowledge, no studies have been published to date on viromes in the Great Lakes ballast water. Therefore, the objectives of this study were (i) to investigate the composition and taxonomic diversity of viruses in ballast water collected around the Great Lakes and (ii) to understand the Great Lakes ballast water virome signatures by comparative virome analyses.

## MATERIALS AND METHODS

**Ethics Statement.** The ballast water sampling in the Port of Duluth—Superior was approved by the Wisconsin Department of Natural Resources (WDNR) and by captains of vessels whose ballast waters were sampled. The sampling was conducted under the guidance of a ballast water inspector from the WDNR for safety purposes. Names of vessels and port terminals were designated as random letters as part of the sample confidentiality agreement.

**Sample Collection.** Ballast waters were collected from five bulk carriers coming from different parts (Lakes Erie, Huron, Michigan, and Ontario) of the North America Great Lakes but all arriving in three terminals of the Port of Duluth—Superior over a one-week period on May 2013 (Figure S1, Table 1). Ballast waters (60 L) were collected from one ballast tank per vessel either through a ballast water pipeline or sounding pipe. Surface harbor waters from different port terminals were also collected with a bucket near the vicinity of only three vessels whose ballast waters were sampled. While the ballast waters were from different lakes, all harbor waters were from Lake Superior. The ballast (B) waters were designated BB, IB, and MB and their matching harbor (H) waters as BH, IH, and MH. The identification, description, and collection information for the ballast and harbor waters are summarized in Table 1. Dissolved oxygen and salinity of the samples were measured using a YSI 85 probe (YSI Inc., Yellow Springs, OH, USA), pH using a Ub-5 pH meter (Denver Instrument, Bohemia, NY, USA), and turbidity using a 2020we portable turbidity meter (LaMotte Company, Chestertown, MD, USA).

**Preparation and Sequencing of Viromes.** Samples were stored at 4 °C and processed within 24 h of sample collection in a laboratory of the University of Minnesota—Duluth. A pipeline for the generation of viromes from the ballast and harbor waters is depicted in Figure S2A. Briefly, viral particles in approximately 60 L of each water sample were concentrated by tangential flow ultrafiltration (30 kDa cutoff).[26] The concentrates were transported overnight to Michigan State University (MSU) at 4 °C and stored immediately at −80 °C


DOI: 10.1021/acs.est.5b01633
Environ. Sci. Technol. 2015, 49, 8396−8407

upon arrival for further processing. Viral particles in the ultrafiltration concentrates were further concentrated and purified by polyethylene glycol (PEG) precipitation.[27] Viral particles contained in the PEG concentrates were subjected to a combination of purification methods, including chloroform treatment, a series of 0.45 $\mu$m and 0.22 $\mu$m sterile syringe filtrations, and DNase I treatment to remove any nonviral cells. The viral DNA and RNA were simultaneously extracted using the QIAamp MinElute Virus Spin Kit (Qiagen, Valencia, CA, USA). To ensure the absence of any contaminating microbial DNA, viral nucleic acids were screened by 16S rDNA PCR with 27F/1492R universal primers.[28] The extracted viral nucleic acids were reverse transcribed and amplified as previously described to obtain a sufficient quantity of DNA and reverse transcribed RNA (cDNA) for metagenomic sequencing.[29,30] Resulting viral DNA and cDNA were purified using PCR Clean-Up System (Promega, Madison, WI, USA). The sequencing libraries with a 200 base pair (bp) insert size were prepared using the Illumina TruSeq DNA prep kit version 2 (Illumina, San Diego, CA, USA). Sequencing was performed on the Illumina HiSeq 2500 instrument in a rapid run mode at the Research Technology Support Facility (RTSF) at MSU, generating paired-end (PE) reads with a read length of 100 bp.

**Analysis of Viromes.** A pipeline for the analyses of the metagenomic sequencing data sets is depicted in Figure S2B. Briefly, raw reads from each virome were trimmed to remove the "Primer B" sequence used for random amplification and low quality reads were discarded (quality value lower than 30 in 50% of bases, shorter than 30 bp, or containing ambiguous nucleotides) using HOMER version 4.4 and FASTX-Toolkit version 0.0.13, respectively.[31,32] PE reads from each virome were *de novo* assembled into contiguous reads (contigs) using IDBA-UD and blasted against the NCBI Viral Reference Sequence (RefSeq) database (downloaded from ftp://ftp.ncbi.nih.gov/refseq/release/viral in September 2014) for taxonomic assignment using BLASTX (E-value < $10^{-5}$).[33,34] The BLASTX output was parsed using the MEtaGenome Analyzer (MEGAN) version 5.6.6 (Min Score = 50.0, Max Expected = 1.0E$^{-5}$, Top Percent = 10.0, Min Support Percent = 0.1, Min Support = 1, and LCA Percent = 100.0).[35] To assess relative abundance of a phylogenetic group in the viromes, read mapping to contigs was performed using default settings in Bowtie 2.[36] Relative abundance of a phylogenetic group was calculated based on eq 1, and its percentage was used to compare a particular group in a virome to the rest of the viromes.

$$\mathrm{Ri} = \sum (\mathrm{Ni}/\mathrm{Li}) \quad (1)$$

where Ri is the relative abundance of a phylogenetic group *i*, Ni is the number of reads aligned to a contig in a phylogenetic group *i*, and Li is the length (kbp) of a contig in a phylogenetic group *i*.

Read mapping to reference genomes of major emerging viral pathogens of fish and shrimp was performed using default settings in Bowtie 2 to verify the taxonomic assignment of contigs through the BLASTX search.[36] All Illumina sequencing reads from the ballast and harbor water viromes obtained in this study are available in the NCBI Sequence Read Archive (www.ncbi.nlm.nih.gov/sra) under accession number SRP048255.

**Comparison of Viromes.** Annotation-free approaches, which are independent from taxonomic assignment of contigs were used to compare individual viromes with each other. Analysis of shared homologues of contigs present in each virome was performed by pairwise comparisons using TBLASTX (E-value < $10^{-3}$). The percentage of a shared number of contigs in each direction was used to represent the similarity between viromes.[37] Furthermore, pairwise comparisons were performed by determining contig coverage between viromes using QUAST version 2.3.[38] An alignment of all contigs in a virome to all contigs in the other virome (used as a reference) was performed, and the ratio of aligned bases to the number of all bases in the reference was used to represent the contig coverage.[38] Similarities of the eight viromes were visualized with principal coordinate analysis (PCoA) using PAST statistical package version 3.[39] Viromes of two temperate freshwater lakes in France were downloaded from MetaVir version 2 (http://metavir-meb.univ-bpclermont.fr/) and included in the PCoA analysis as an outlier group.[40,41] Last, the Great Lakes ballast and harbor water viromes were compared with a set of previously published viromes from other aquatic environments using MetaVir based on sequence similarity with a cross-TBLASTX search.[41] The MetaVir workflow requires data sets containing at least 50 000 input sequences for the TBLASTX-based comparison, thus the analysis was not available for the two viromes in this study (35 819 and 15 887 contigs for BB and MB viromes, respectively).

## RESULTS

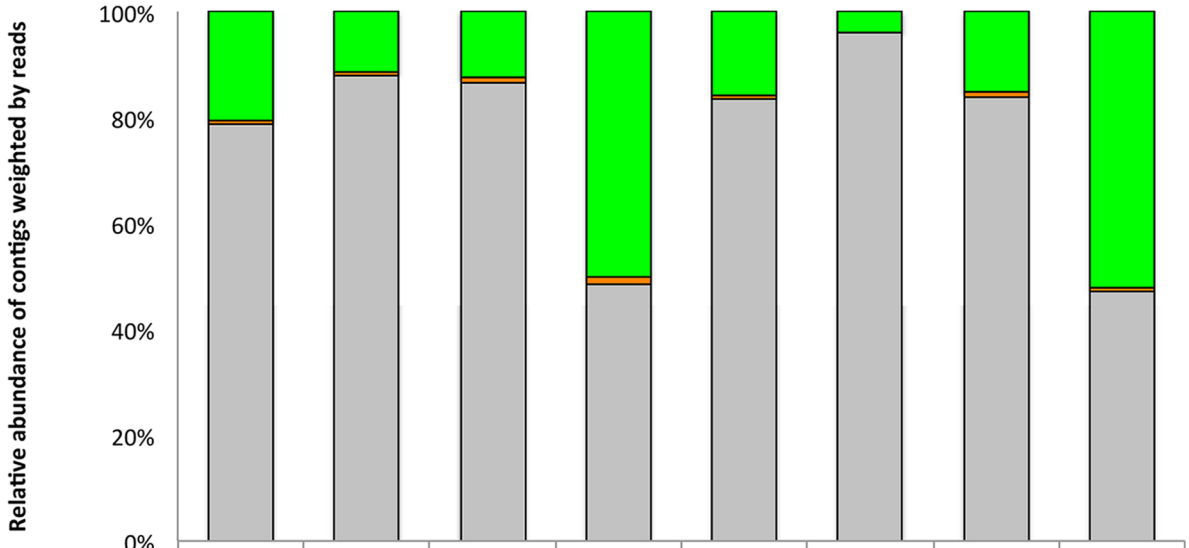
**General Water Quality.** The pH, dissolved oxygen, salinity, and turbidity of the ballast and harbor waters are summarized in Table S1. As expected, all samples were low in salinity. The harbor waters collected from Lake Superior were potentially influenced by shipping activities and had an average pH of 6.9, while the ballast water from vessels originating from Lakes Erie, Huron, Michigan, and Ontario had an average pH of 6.7. Ballast waters from Lakes Erie (AB) and Ontario (PB) had low dissolved oxygen (DO, 4.9 mg/L) resulting in a lower average of DO (6.8 mg/L) in comparison to harbor waters from Lake Superior (8.6 mg/L). The harbor waters generally had higher turbidity levels (average 9.7 NTU) compared with the ballast waters (average 3.9 NTU) with the exception of the ballast water from Lake Erie (AB, 10.0 NTU).

**Overview of the Virome Data Sets.** A pipeline for metagenomics analysis of the ballast and harbor waters was developed to address the identification of both DNA and RNA viruses. The use of Illumina Hiseq 2500 resulted in a total of 551 022 890 raw sequence reads with a read length of 100 bp. After quality trimming and filtering of reads, 501 015 363 reads (90.3% of raw reads) remained. Remaining reads were then split into 470 931 386 PE (94% of remaining reads) and 30 083 977 single-end (SE; 6% of remaining reads) reads prior to *de novo* assembly (Table S2). The PE reads were used for assembling reads into contigs, producing a total of 867 050 assembled contigs. The mean contig length among the contigs across eight viromes was 590 bp, and the mean N50 was 638 bp. A significant increase in contig length generated from *de novo* assembly improves annotation through homology searches against a reference database. Mapping PE reads to contigs showed that overall alignment rates ranged from 26.8 to 91.2% depending on the viromes (Table S3).

**Taxonomic Profile of the Great Lakes Ballast and Harbor Water Viromes.** A wide array of viruses was discovered in all ballast and harbor waters from the Great Lakes. Among all assembled contigs, about 22.8% were assigned to viral taxa, but 77.2% had no or low levels of amino acid

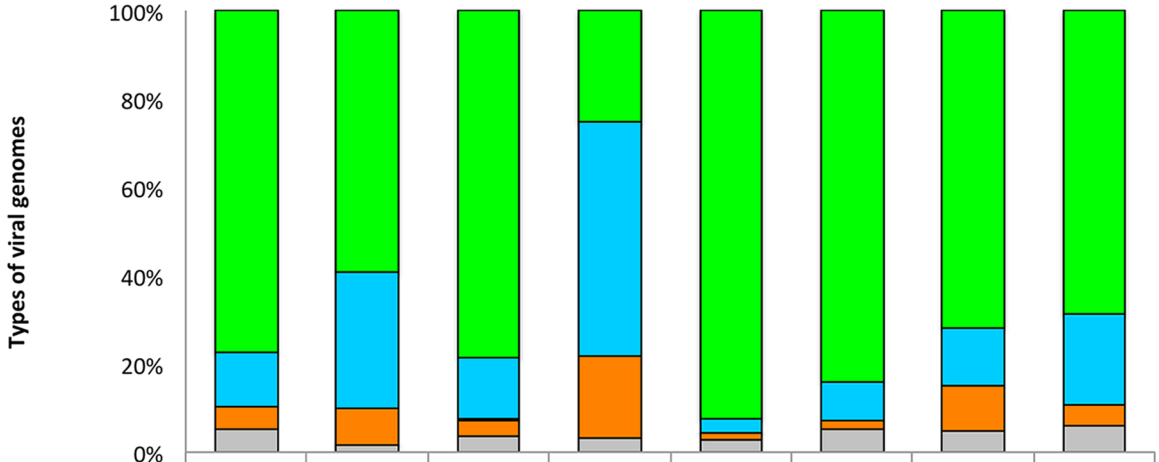

DOI: 10.1021/acs.est.5b01633
Environ. Sci. Technol. 2015, 49, 8396−8407

A




| | AB (Erie) | BB (Ontario) | BH (Superior) | IB (Huron) | IH (Superior) | MB (Michigan) | MH (Superior) | PB (Ontario) |
|---|---|---|---|---|---|---|---|---|
| Assigned | 20.4 | 11.6 | 12.6 | 50.3 | 15.7 | 3.9 | 15.3 | 52.3 |
| Not assigned | 0.9 | 0.4 | 1.0 | 1.1 | 0.8 | 0.1 | 0.8 | 0.7 |
| No hits | 78.7 | 88.0 | 86.4 | 48.6 | 83.5 | 96.0 | 83.9 | 47.0 |

B



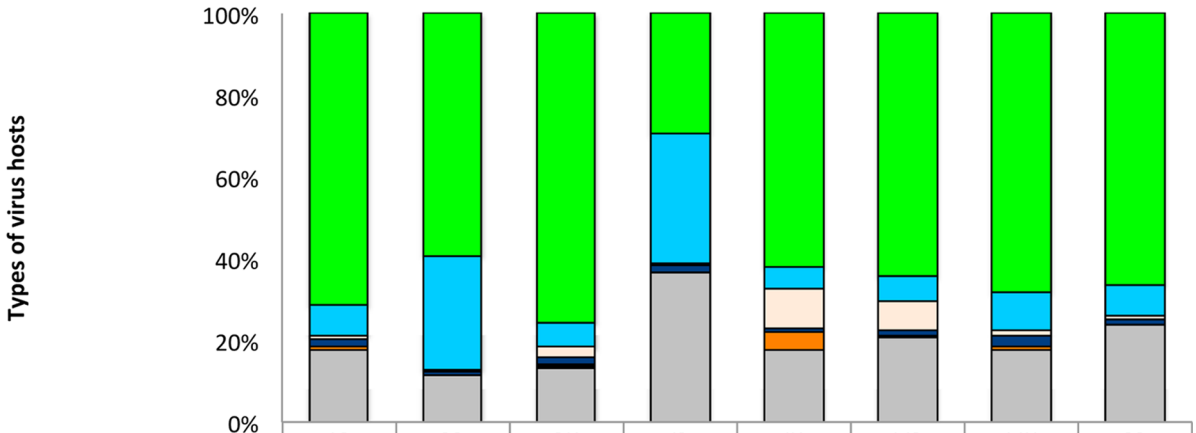

| | AB (Erie) | BB (Ontario) | BH (Superior) | IB (Huron) | IH (Superior) | MB (Michigan) | MH (Superior) | PB (Ontario) |
|---|---|---|---|---|---|---|---|---|
| dsDNA viruses | 77.6 | 59.3 | 78.5 | 25.3 | 92.4 | 84.1 | 71.7 | 68.9 |
| ssDNA viruses | 12.1 | 30.6 | 13.9 | 52.8 | 3.1 | 8.8 | 13.3 | 20.4 |
| dsRNA viruses | 0.0 | 0.0 | 0.3 | 0.1 | 0.0 | 0.0 | 0.2 | 0.0 |
| ssRNA viruses | 5.0 | 8.4 | 3.6 | 18.7 | 1.6 | 2.1 | 10.2 | 4.7 |
| Unassigned | 5.2 | 1.7 | 3.8 | 3.1 | 2.9 | 5.0 | 4.7 | 6.0 |

**Figure 1.** continued



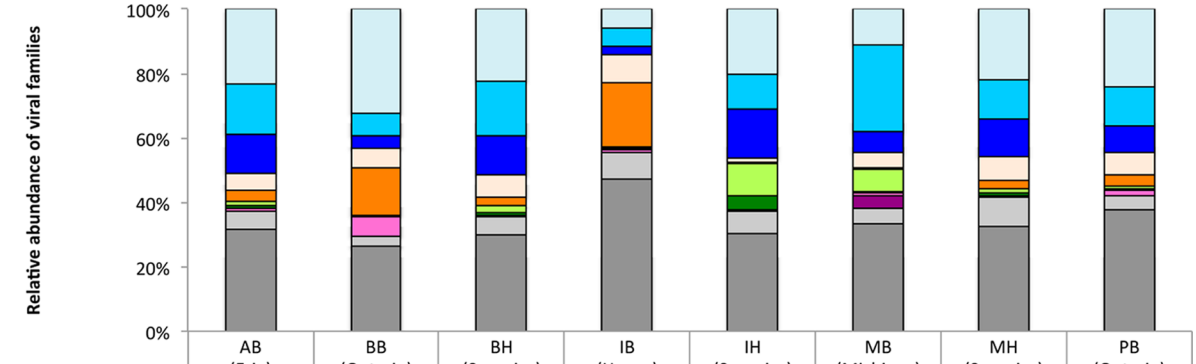
DOI: 10.1021/acs.est.5b01633
*Environ. Sci. Technol.* 2015, 49, 8396−8407

C




| | AB (Erie) | BB (Ontario) | BH (Superior) | IB (Huron) | IH (Superior) | MB (Michigan) | MH (Superior) | PB (Ontario) |
|---|---|---|---|---|---|---|---|---|
| Bacteria | 71.5 | 59.3 | 75.7 | 29.4 | 62.1 | 64.3 | 68.2 | 66.5 |
| Vertebrates/Invertebrates | 7.2 | 28.0 | 6.0 | 31.8 | 5.2 | 6.1 | 9.5 | 7.5 |
| Algae | 1.2 | 0.3 | 2.3 | 0.5 | 9.7 | 7.0 | 1.3 | 0.8 |
| Plants | 1.5 | 1.0 | 1.9 | 1.8 | 1.0 | 1.5 | 2.4 | 1.3 |
| Amoebae | 1.0 | 0.1 | 0.6 | 0.1 | 4.5 | 0.2 | 0.9 | 0.3 |
| Fungi/Protozoa | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Unknown host | 17.5 | 11.3 | 13.4 | 36.4 | 17.6 | 20.9 | 17.7 | 23.7 |

D




| | AB (Erie) | BB (Ontario) | BH (Superior) | IB (Huron) | IH (Superior) | MB (Michigan) | MH (Superior) | PB (Ontario) |
|---|---|---|---|---|---|---|---|---|
| Siphoviridae | 23.2 | 32.2 | 22.2 | 5.7 | 20.1 | 11.1 | 21.8 | 24.0 |
| Podoviridae | 15.6 | 6.9 | 17.3 | 5.9 | 10.8 | 27.0 | 12.3 | 12.2 |
| Myoviridae | 12.0 | 4.1 | 11.8 | 2.6 | 15.6 | 6.5 | 11.6 | 8.5 |
| Microviridae | 5.2 | 6.0 | 7.0 | 8.6 | 0.9 | 4.9 | 7.4 | 6.8 |
| Parvoviridae | 3.8 | 14.8 | 2.7 | 19.8 | 0.7 | 0.4 | 2.6 | 3.6 |
| Phycodnaviridae | 1.2 | 0.3 | 2.3 | 0.5 | 9.7 | 7.0 | 1.3 | 0.8 |
| Mimiviridae | 1.0 | 0.1 | 0.6 | 0.1 | 4.4 | 0.2 | 0.9 | 0.3 |
| Circoviridae | 0.7 | 6.0 | 0.6 | 1.1 | 0.3 | 0.9 | 0.6 | 1.9 |
| Ascoviridae | 0.0 | 0.0 | 0.0 | 0.0 | 0.3 | 3.7 | 0.0 | 0.0 |
| Others[a] | 5.7 | 3.3 | 5.5 | 8.3 | 6.6 | 5.0 | 9.2 | 4.5 |
| Others[b] | 31.6 | 26.3 | 29.9 | 47.1 | 30.5 | 33.4 | 32.4 | 37.5 |

**Figure 1.** Relative abundance of contigs weighted by sequence reads based on taxonomic assignment of contigs (A). Contigs that were assigned to viral taxa but did not meet the selected MEGAN parameters were placed under "Not assigned," and contigs that did not have any hits to known sequences in the databases were placed under "No hits." Types of viral genomes (B), types of virus hosts (C), and contigs assigned to viral families (D) in the ballast and harbor water viromes. Viral families whose maximum relative abundances across eight viromes less than 3% were represented as "Others[a]". Unassigned contigs at the family level were represented as "Others[b]."

similarity to known viral sequences in the NCBI RefSeq database (Figure 1A). Of the contigs with similarity to known viruses, 34 different viral families were identified, consisting of 15 double-stranded DNA (dsDNA, 69.7%), six single-stranded



DOI: 10.1021/acs.est.5b01633
Environ. Sci. Technol. 2015, 49, 8396−8407

Table 2. Contigs Identified As Viral Pathogens of Fish and Shrimp by BLASTX Search against the NCBI RefSeq Database

| virus | taxonomic classification | AB (Erie) | BB (Ontario) | BH (Superior) | IB (Huron) | IH (Superior) | MB (Michigan) | MH (Superior) | PB (Ontario) |
|---|---|---|---|---|---|---|---|---|---|
| fish | | | | | | | | | |
| infectious spleen and kidney necrosis virus | *Iridoviridae, Megalocytivirus* | 3 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| koi herpesvirus | *Alloherpesviridae, Cyprinivirus* | 2 | 0 | 1 | 0 | 5 | 0 | 3 | 0 |
| striped jack nervous necrosis virus | *Nodaviridae, Betanodavirus* | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| shrimp | | | | | | | | | |
| infectious myonecrosis virus | unclassified *Totiviridae* | 1 | 0 | 5 | 0 | 0 | 0 | 1 | 0 |
| macrobrachium rosenbergii nodavirus | *Nodaviridae, Alphanodavirus* | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| taura syndrome virus | *Dicistroviridae, Aparavirus* | 0 | 0 | 0 | 2 | 0 | 0 | 1 | 1 |
| white spot syndrome virus | *Nimaviridae, Whispovirus* | 3 | 1 | 0 | 13 | 2 | 0 | 4 | 23 |

DNA (ssDNA, 19.4%), one double-stranded RNA (dsRNA, 0.1%), and 12 single-stranded RNA (ssRNA, 6.8%) viruses (Figure 1B). These represented viruses infecting a wide range of hosts, including bacteria (62.1%), vertebrates/invertebrates (12.6%), algae (2.9%), plants (1.6%), amoebae (1.0%), and fungi/protozoa (0.01%; Figure 1C).

Relative abundance of viral families revealed that more than half (average 52.5%) of the assigned contigs in each virome were homologous to dsDNA phages, belonging to the order of *Caudovirales* (*Myoviridae*, *Podoviridae*, *Siphoviridae*, and unclassified *Caudovirales*) with the exception of the IB virome (Lake Huron) with a low relative abundance of 19.3% (Figure 1D). These dsDNA phages were associated with 62 different bacterial hosts, with the majority being *Cellulophaga* (average 14.1%) followed by *Synechococcus* (9.1%) and *Pelagibacter* (6.9%; Table S4). Along with *Synechococcus*, a number of contigs was found to be associated with phages whose hosts belong to cyanobacteria such as *Prochlorococcus* (3.3%). Moreover, contigs most similar to those infecting bacteria in genera containing human pathogens, including *Burkholderia* (1.3%), *Klebsiella* (0.3%), *Pseudomonas* (3.6%), *Salmonella* (2.1%), and *Vibrio* (6.0%), were detected.

Vertebrate (including those that could infect humans) and invertebrate viruses were present in all samples (average 12.7%). Two ballast water viromes, BB (Lake Ontario; 28.0%) and IB (Lake Huron; 31.8%), had higher relative abundances of vertebrate and invertebrate viruses due to significantly higher abundances of *Parvoviridae*, which is capable of infecting either vertebrates or invertebrates (14.8% and 19.8% for the BB and IB viromes, respectively). The contigs related to *Alphatetraviridae* (insect viruses), *Iflaviridae* (insect viruses), and unassigned *Picornavirales* (vertebrate/invertebrate viruses) were present in at least one of the ballast water viromes but not in any of the Lake Superior harbor water viromes (Table S5). In contrast, contigs related to *Reoviridae* (vertebrate/invertebrate viruses) were detected only in one of the harbor water viromes, BH (Lake Superior).

Contigs belonging to algal viruses, *Phycodnaviridae*, were present in all viromes (average 2.9%), with higher relative abundances in the IH (Lake Superior; 9.7%) and MB (Lake Michigan; 7.0%) viromes. Viruses infecting plants were also present in all viromes (average 0.17%). The contigs related to *Benyviridae*, *Nanoviridae*, *Sobemovirus*, *Tymovirales*, and *Virgaviridae* were present in at least one of the ballast water viromes but not in any of the Lake Superior harbor water viromes (Table S5). Viruses infecting fungi/protozoa, *Totiviridae*, were only present in one of the Lake Superior harbor water viromes, BH (0.12%). Contigs belonging to amoeba viruses, *Mimiviridae*, were present in all viromes, while *Marseilleviridae* were present only in one of the harbor water viromes, IH (Lake Superior), and in the MB (Lake Michigan) virome.

**Viral Pathogens of Fish and Shrimp in the Great Lakes Ballast and Harbor Waters.** Our metagenomic data allowed for an in-depth examination of the types of viruses that might be considered to be key pathogens of fish and shrimp in the Great Lakes. Of the assigned contigs in the ballast and harbor water viromes, 75 contigs were identified as viral pathogens of fish or shrimp (Table 2). The identified contigs had lower amino acid similarity (23−44%) to known viruses in the RefSeq database except one contig in the MH virome (Lake Superior) identified as the infectious spleen and kidney necrosis virus (ISKNV) with 72% amino acid similarity. Mapping of reads to complete reference genomes of the identified viral pathogens of fish and shrimp showed low mapping rates (data not shown) except for the koi herpesvirus (KHV). Read mapping to five different KHV genomes exhibited slightly higher mapping rates with an average coverage of 7.6% (Table S6). Overall, the low mapping rate of reads together with low amino acid similarity of contigs indicated that these viral pathogens are potentially novel or genetically diverse.

**Comparison among the Great Lakes Ballast and Harbor Water Viromes.** Comparisons of viromes of ballast waters from Lakes Erie, Huron, Michigan, and Ontario and harbor waters from Lake Superior were conducted to examine the similarities between lakes and to inspect whether the harbor waters near vessels were reflective of the respective ballast waters. For virome comparison, annotation-independent approaches were used as they use all contigs present in the virome data sets while annotation-dependent approaches use a low proportion of the contigs with similarity to known sequences in the existing database (average 22.8% of the contigs from Figure 1A). It should be noted that two annotation-independent approaches used in this study have different ways to analyze shared homologues of contigs in each virome. TBLASTX-based comparison uses a shared number of contigs between viromes while the QUAST tool uses a shared number of aligned bases in contigs.

The difference in contig profiles analyzed by two methods were consistent with only slight variations as observed in Figure 2A and B. PCoA analyses between viromes suggested that the



DOI: 10.1021/acs.est.5b01633
Environ. Sci. Technol. 2015, 49, 8396−8407

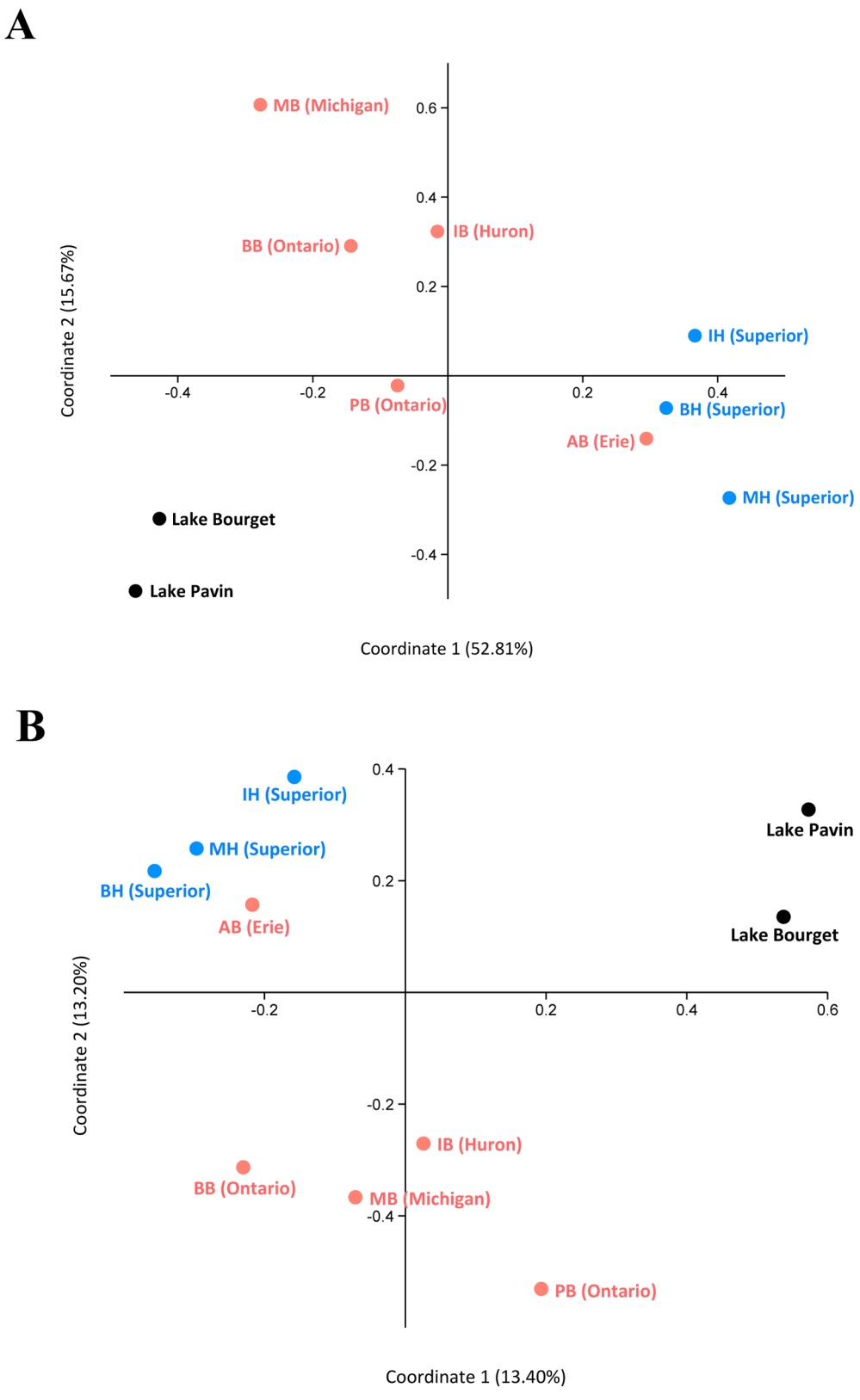



**Figure 2.** PCoA of virome distances based on contig profiles using TBLASTX (A) and QUAST (B). Ballast and harbor water viromes were represented in red and blue circles, respectively. An outlier group (viromes of two temperate freshwater lakes in France) was represented in black circles.

three Lake Superior harbor water viromes (BH, IH, and MH) grouped together and were distinct from the respective ballast water viromes (BB, IB, and MB from Lakes Ontario, Huron, and Michigan, respectively). The ballast water virome, AB, from Lake Erie showed a similar contig profile with the Lake Superior harbor waters. In addition, the two ballast water viromes from Lake Ontario, BB (open lake) and PB (Port of Hamilton), showed different contig profiles to each other



DOI: 10.1021/acs.est.5b01633
*Environ. Sci. Technol.* 2015, 49, 8396−8407

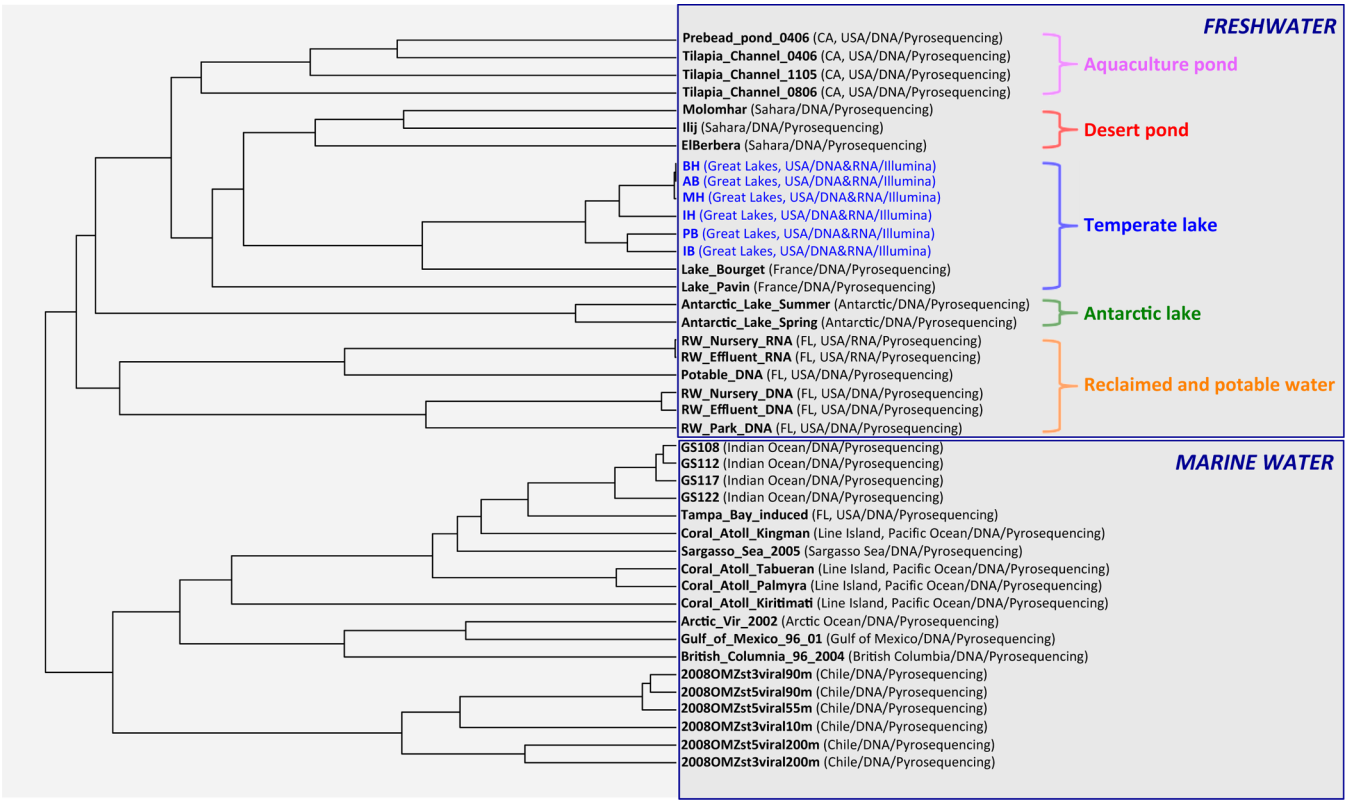



**Figure 3.** Comparison of the ballast and harbor water viromes with published freshwater viromes based on sequence similarity with a cross-TBLASTX search. The six viromes (AB, BH, IB, IH, MH, and PB) from this study are highlighted in blue. TBLASTX-based comparison was not available for the other two viromes (BB and MB), due to insufficient numbers of input contigs. Virome names were retrieved from the MetaVir. Sampling site, virome fraction, and sequencing platform of each virome were shown in the parentheses.

(Table S7 and S8). The similarity in the contig profiles between IB (Lake Huron) and MB (Lake Michigan) was somewhat expected because Lakes Huron and Michigan are two halves of one lake and considered to be hydrologically the most connected.

**Comparison of the Great Lakes Ballast and Harbor Water Viromes with Other Aquatic Viromes.** The Great Lakes ballast and harbor water viromes were also compared to previously published viromes in freshwater and marine water. The hierarchical clustering tree analysis revealed that aquatic viromes could be classified into two representative groups, freshwater virome and marine water virome (Figure 3), highlighting that these different environments contain unique virome signatures. Among the freshwater virome, the Great Lakes ballast and harbor water viromes were closely related to each other and clustered with two viromes from the oligomesotrophic Lake Pavin and the mesotrophic Lake Bourget in France. Viromes from desert ponds and an aquaculture pond generated a subgroup with the temperate lake viromes. Viromes from an Antarctic lake and reclaimed and potable water were aggregated individually and distant from the other freshwater viromes.

## DISCUSSION

This study investigated viromes in ballast and harbor waters originating from the Great Lakes including Lakes Erie, Huron, Michigan, Ontario, and Superior (Figure S1). The Great Lakes, located in northeastern North America, form the largest group of freshwater lakes on Earth.[42] The sampling location at the Port of Duluth—Superior, located at the western part of Lake Superior (Figure S1), is the busiest and largest port on the Great Lakes and receives the largest volume (more than 18 million gallons a day) of ballast water within the Great Lakes.[42,43]

**Diversity of Viruses in Ballast Water.** In contrast to what is known about the diversity of metazoans transported by ships' ballast water, little is known about the diversity of viruses in ballast water. Moreover, most studies have only focused on the abundance of VLPs in ballast water using microscopic approaches.[9,11,44,45] Our study gives us the first insight into the diversity of viruses and their associated host populations in ballast waters across the Great Lakes. Understanding this diversity including types of viruses that are being discharged will assist in defining ballast water treatment and potential ecological risk.

Viruses in the ballast waters from Lakes Erie, Huron, Michigan, and Ontario were characterized and compared with those of Lake Superior harbor waters where they were being discharged. Among 34 viral families identified, 12 viral families were shared by all ballast and harbor waters (highlighted in bold in Table S4), representing 62.2% of the total abundance of phylogenetic groups in the viromes. This suggested that the majority of viromes among the different lakes were similar in the phylogenetic types. However, the presence of six viral families (*Alphatetraviridae*, *Beniviridae*, *Iflaviridae*, *Nanoviridae*, *Sobemovirus*, and *Virgaviridae*) detected in at least one of the ballast waters but not in any of the Lake Superior harbor waters suggested potential opportunities for non-native viruses to be discharged with ballast water into the Port of Duluth—Superior.



DOI: 10.1021/acs.est.5b01633
*Environ. Sci. Technol.* 2015, 49, 8396−8407

In this study, annotation-independent approaches enabled a more comprehensive comparison of viromes from the various lakes.[37,40] The Great Lakes is a single interconnected hydrologic system. However, each lake is unique due to size, topography, and land use impacts. In addition, nearshore and harbor waters are different from open waters in these lakes (as shown by different contig profiles of the two Lake Ontario viromes, BB and PB). It was somewhat unexpected that ballast water, AB, taken from the Port of Toledo located on Lake Erie showed an indistinguishable virome signature from the Lake Superior harbor waters (Figure 2A and B). Port of Toledo has been characterized as "the port of the greatest concern" for ballast water mediated invasions throughout the Great Lakes by U.S. EPA.[46] The Port of Toledo received the second-largest amount of ballast water following the Port of Duluth—Superior from vessels whose sources were outside the Great Lakes.[46] We hypothesize that this may have contributed to the indistinguishing profile of the AB virome from the Lake Superior harbor water viromes. Further water sampling and investigation of hydrographic characteristics are needed to get better insight into similarities between these geographically distinct samples.

This study added new virome data sets associated with freshwater environments to the currently limited virome database. To date, many virome studies have focused on marine environments, and others have investigated viromes in human designed/managed freshwater environments.[37,47−54] Virome studies of natural freshwater environments can be found in two extreme environments, an Antarctic lake and desert ponds, as well as in temperate freshwater lakes in France.[40,55,56] These previous studies examined either DNA or RNA viruses (but not both) and were limited to the lower sequencing yield of earlier technologies such as Roche 454 pyrosequencing.

This new work was the first comprehensive study of both DNA and RNA viruses originated from temperate freshwater lakes in the Great Lakes system. Comparison of the Great Lakes ballast and harbor water viromes with previously characterized aquatic viromes is important to determine how inclusion of an RNA virus fraction affects virome clustering. The comparison demonstrated that freshwater viromes are distinct from marine water viromes, providing evidence of hierarchical clustering according to salinity levels despite vast geographic distances. The difference between freshwater and marine water viromes presented in this study is consistent with previous studies.[40,57] This is due to the dominance of dsDNA phages in our viromes, and thus inclusion of an RNA virus fraction did not significantly affect virome clustering. It is noteworthy that different approaches used to generate and analyze viromes (e.g., sample preparation, sequencing platform, bioinformatics workflow) have the potential to undermine this comparison. Currently, mid-ocean exchange of ballast water is widely used to comply with the IMO ballast water discharge guidelines.[58] The underlying principle of this practice is to replace coastal water in ballast tanks with oceanic water. This can, however, introduce viruses associated with the marine water environment to freshwater environments such as the Great Lake basin. The impact of the transport of viruses between biomes on host populations should be further investigated.

**Implications and Control of Viruses in Ballast Water.** With the elevated public attention on the introduction and spread of invasive species, for example VHSV, which caused extensive losses of wild fish in the Great Lakes, this study examined major emerging viral pathogens of fish and shrimp using metagenomics approaches.[59] Homology searches against existing databases tentatively identified three and four groups of viruses causing diseases in fish and shrimp, respectively. These viruses are listed by the World Organization for Animal Health (OIE) as causing notifiable diseases of fish and shrimp.[60] Notably, KHV (formally classified as the *Cyprinid herpesvirus* 3) has appeared within the Great Lakes basin with multiple mortality events since 2004.[61−63] It causes diseases and mass mortality in common and koi carp (*Cyprinus carpio* and *C. carpo koi*, respectively) and has become the most dramatic example of an emerging disease of fish.[64] The identification of these potential viral pathogens in ballast water is important in improving our understanding of what ballast water treatment would be needed to inactivate viruses in the future. In addition, the presence of viral pathogens with lower amino acid similarity to known viruses means that these viruses are potentially novel or genetically diverse. Further investigations (e.g., phylogenetic approach, gene-specific PCR) are required to confirm the identification of these viral pathogens.

The diverse viral populations in ballast water and the movement of these viruses around the world have potential impacts on phytoplankton, animal health, and even human health.[9,11,65] This reinforces the need for ballast water treatment for controlling potential viral invasion. Mid-ocean exchange of ballast water is an interim solution to control the introduction of aquatic invasive species. A few studies have reported that this practice is not effective in reducing the total number of bacteria and VLPs.[45,66] Over the past few years, special efforts have been made among the scientific and industrial communities to develop technologies for ballast water treatment because of the need for vessels to establish ballast water treatment systems onboard by 2016 according to the U.S.C.G.[7] Several ballast water treatment systems such as filtration, deoxygenation, biocides, and ultraviolet treatment have been developed.[5,67] Unfortunately, these techniques have only been tested with marine water focusing on reducing the abundance of phytoplankton and bacteria. Thus, the effectiveness of ballast water treatment technologies in removing viruses in freshwater environments is currently unknown.

Metagenomics approaches have the potential to overcome the limitations of traditional methods for the detection and characterization of viruses such as cell culture and gene-specific PCR and, thus, provide the opportunity to explore composition and taxonomic diversity of uncultured viruses. However, the metagenomic workflow from sample collection to bioinformatics analysis is experimentally and computationally challenging at each step, and potential biases may be introduced in estimating viral diversity.[68] On sampling, approximately 60 L of ballast water collected from a ballast tank from each vessel in this study may not be representative of the 5 million gallons of ballast water typically carried by a vessel.[3] A more intensive sampling design may be needed to aid in resolving the viral diversity associated with ballast waters. Additionally, the bioinformatics analysis is limited by a lack of viral reference genomes and the need to assemble the short reads for appropriate virome comparison and annotation. The paucity of viral reference databases affects the ability to identify viral pathogens and to do comparisons of the functional capacity of the viromes, when traits cannot be identified.[69−71] The assembly process also skews the analysis by including primarily the most abundant organisms. With large data sets, sequences that are rare do not assemble into contigs and are therefore not


DOI: 10.1021/acs.est.5b01633
Environ. Sci. Technol. 2015, 49, 8396−8407

included in the contig analysis.[72] However, efforts to sequence more deeply than typical, as is done in this study, will help to address these issues.

The findings of our study have several important implications. First, ballast and harbor waters originating from the Great Lakes harbored diverse viruses including viral pathogens associated with fish and shrimp (with low amino acid similarity), emphasizing the need for implementing ballast water discharge limits for viruses and treatment. Second, viromes were distinct among the Great Lakes and formed a specific group of temperate freshwater viromes but separated from viromes associated with marine environments and engineered freshwater systems, suggesting the potential transfer and introduction of viruses between biomes and to the Great Lakes through ballast water discharge. Looking forward, the results of this study will assist in identifying potential viral invasions via ships' ballast water and evaluating ballast water quality and standards to protect public and ecosystem health from invasive species.

## ■ ASSOCIATED CONTENT

### Supporting Information

Additional materials and methods and Figure S1−S2 and Table S1−S8. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.est.5b01633.

## ■ AUTHOR INFORMATION

### Corresponding Author

*Phone: (517) 432-8185. Fax: (517) 432-1699. E-mail: kimyise2@msu.edu.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This study was supported by the National Science Foundation, Partnerships for International Research and Education (OISE-0530174). We would like to thank Cordell H. Manz in WDNR and Andrew J. Reed at University of Minnesota Duluth for help in sampling ballast water, Randall E. Hicks at University of Minnesota Duluth for providing a lab space for the sample processing, Qingpeng Zhang at MSU for fruitful discussion, and the High Performance Computing Center at MSU for bioinformatics support.

## ■ REFERENCES

(1) National Research Council (U.S.). Committee on Ships' Ballast Operations, *Stemming the tide: controlling introductions of nonindigenous species by ships' ballast water*; National Academy Press: Washington, DC, 1996.

(2) Fredricks, R. Aquatic nuisance species, manditory ballast water management, and alternative ballast water treatment standards to protect the marine environment. Submitted to U.S. commission on ocean policy, Boston, Massachusetts; 2002. http://govinfo.library.unt.edu/oceancommission/publicomment/northeastcomments/fredricks_comment.pdf.

(3) Carlton, J. T.; Reid, D. M.; Van Leeuwen, H. *Shipping study the role of shipping in the introduction of nonindigenous aquatic organisms to the coastal waters of the United States (other than the Great Lakes) and an analysis of control options*; U.S. Coast Guard, Office of Engineering, Logistics, and Development; National Technical Information Service Distributor: Washington, DC, 1995.

(4) Clark, B. Will Viral Hemorrhagic Septicemia (VHS) Be the Straw That Breaks the Camel's Back? The Balkanization of Great Lakes Ballast Water Law. *Minn. J. Int. Law* **2009**, *18* (1), 227−264.

(5) Tsolaki, E.; Diamadopoulos, E. Technologies for ballast water treatment: a review. *J. Chem. Technol. Biotechnol.* **2010**, *85* (1), 19−32.

(6) Ferrate Treatment Technologies Website. http://www.ferratetreatment.com/ballastwater.htm.

(7) David, M.; Gollasch, S. Global Maritime Transport and Ballast Water Management Issues and Solutions. In *Invading Nature - Springer Series in Invasion Ecology 8*; Springer: Dordrecht, The Netherlands, 2015.

(8) ABS Website. http://ww2.eagle.org/content/dam/eagle/publications/2014/BWTAdvisory14312rev3.pdf.

(9) Ruiz, G.; Rawlings, T.; Dobbs, F.; Drake, L.; Mullady, T.; Huq, A.; Colwell, R. Global spread of microorganisms by ships - Ballast water discharged from vessels harbours a cocktail of potential pathogens. *Nature* **2000**, *408* (6808), 49−50.

(10) Drake, L. A.; Choi, K. H.; Ruiz, G. M.; Dobbs, F. C. Global redistribution of bacterioplankton and virioplankton communities. *Biol. Invasions* **2001**, *3* (2), 193−199.

(11) Drake, L.; Doblin, M.; Dobbs, F. Potential microbial bioinvasions via ships' ballast water, sediment, and biofilm. *Mar. Pollut. Bull.* **2007**, *55* (7−9), 333−341.

(12) Sun, B.; Mouland, R.; Way, C.; Rivkin, R. B. Redistribution of heterotrophic prokaryotes through ballast water: A case study from the west coast of Canada. *Aquat. Invasions* **2010**, *5* (1), 5−11.

(13) Seiden, J. M.; Way, C. J.; Rivkin, R. B. Bacterial dynamics in ballast water during trans-oceanic voyages of bulk carriers: environmental controls. *Mar. Ecol.: Prog. Ser.* **2011**, *436*, 145−159.

(14) Bergh, O.; Borsheim, K.; Bratbak, G.; Heldal, M. High abundance of viruses found in aquatic environments. *Nature* **1989**, *340* (6233), 467−468.

(15) Mokili, J.; Rohwer, F.; Dutilh, B. Metagenomics and future perspectives in virus discovery. *Curr. Opin. Virol.* **2012**, *2* (1), 63−77.

(16) Wommack, K. E.; Colwell, R. R. Virioplankton: viruses in aquatic ecosystems. *Microbiol. Mol. Biol. Rev.* **2000**, *64*, 69−114.

(17) Wilhelm, S. W.; Carberry, M. J.; Eldridge, M. L.; Poorvin, L.; Saxton, M. A.; Doblin, M. A. Marine and freshwater cyanophages in a Laurentian Great Lake: Evidence from infectivity assays and molecular analyses of g20 genes. *Appl. Environ. Microbiol.* **2006**, *72* (7), 4957−4963.

(18) Lumsden, J. S.; Morrison, B.; Yason, C.; Russell, S.; Young, K.; Yazdanpanah, A.; Huber, P.; Al-Hussinee, L.; Stone, D.; Way, K. Mortality event in freshwater drum Aplodinotus grunniens from Lake Ontario, Canada, associated with viral haemorrhagic septicemia virus, Type IV. *Dis. Aquat. Org.* **2007**, *76* (2), 99−111.

(19) Elsayed, E.; Faisal, M.; Thomas, M.; Whelan, G.; Batts, W.; Winton, J. Isolation of viral haemorrhagic septicaemia virus from muskellunge, Esox masquinongy (Mitchill), in Lake St Clair, Michigan, USA reveals a new sublineage of the North American genotype. *J. Fish Dis.* **2006**, *29* (10), 611−619.

(20) European Food Safety Authority.. Scientific Opinion of the Panel on Animal Health and Welfare on a request from the European Commission on possible vector species and live stages of susceptible species not transmitting disease as regards certain fish diseases. *EFSA J.* **2007**, *584*, 1−163.

(21) Willner, D.; Hugenholtz, P. From deep sequencing to viral tagging: Recent advances in viral metagenomics. *BioEssays* **2013**, *35* (5), 436−442.

(22) Rohwer, F.; Edwards, R. The Phage Proteomic Tree: a genome-based taxonomy for phage. *J. Bacteriol.* **2002**, *184* (16), 4529−4535.

(23) Fischer, M. G.; Allen, M. J.; Wilson, W. H.; Suttle, C. A. Giant virus with a remarkable complement of genes infects marine zooplankton. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107* (45), 19508−19513.

(24) Nissimov, J. I.; Worthy, C. A.; Rooks, P.; Napier, J. A.; Kimmance, S. A.; Henn, M. R.; Ogata, H.; Allen, M. J. Draft genome sequence of the coccolithovirus EhV-84. *Stand. Genomic Sci.* **2011**, *5* (1), 1−11.



DOI: 10.1021/acs.est.5b01633
*Environ. Sci. Technol.* 2015, 49, 8396−8407

(25) Pagnucco, K. S.; Maynard, G. A.; Fera, S. A.; Yan, N. D.; Nalepa, T. F.; Ricciardi, A. The future of species invasions in the Great Lakes-St. Lawrence River basin. *J. Great Lakes Res.* **2015**, *41*, 139−149.
(26) Hill, V.; Polaczyk, A.; Hahn, D.; Narayanan, J.; Cromeans, T.; Roberts, J.; Amburgey, J. Development of a rapid method for simultaneous recovery of diverse microbes in drinking water by ultrafiltration with sodium polyphosphate and surfactants. *Appl. Environ. Microbiol.* **2005**, *71* (11), 6878−6884.
(27) Jaykus, L.; DeLeon, R.; Sobsey, M. A virion concentration method for detection of human enteric viruses in oysters by PCR and oligoprobe hybridization. *Appl. Environ. Microbiol.* **1996**, *62* (6), 2074−2080.
(28) Lane, D. J. 16S/23S rRNA sequencing. In *Nucleic Acid Techniques in Bacterial Systematics*; Stackebrandt, E., Goodfellow, M., Eds.; John Wiley and Sons: New York, 1991.
(29) Wang, D.; Coscoy, L.; Zylberberg, M.; Avila, P.; Boushey, H.; Ganem, D.; DeRisi, J. Microarray-based detection and genotyping of viral pathogens. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (24), 15687−15692.
(30) Bibby, K.; Peccia, J. Identification of viral pathogen diversity in sewage sludge by metagenome analysis. *Environ. Sci. Technol.* **2013**, *47* (4), 1945−51.
(31) Heinz, S.; Benner, C.; Spann, N.; Bertolino, E.; Lin, Y. C.; Laslo, P.; Cheng, J. X.; Murre, C.; Singh, H.; Glass, C. K. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol. Cell* **2010**, *38* (4), 576−589.
(32) Goecks, J.; Nekrutenko, A.; Taylor, J.; Team, G. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.* **2010**, *11* (8), R86.
(33) Peng, Y.; Leung, H.; Yiu, S.; Chin, F. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **2012**, *28* (11), 1420−1428.
(34) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403−410.
(35) Huson, D.; Auch, A.; Qi, J.; Schuster, S. MEGAN analysis of metagenomic data. *Genome Res.* **2007**, *17* (3), 377−386.
(36) Langmead, B.; Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9* (4), 357−359.
(37) Rodriguez-Brito, B.; Li, L.; Wegley, L.; Furlan, M.; Angly, F.; Breitbart, M.; Buchanan, J.; Desnues, C.; Dinsdale, E.; Edwards, R.; Felts, B.; Haynes, M.; Liu, H.; Lipson, D.; Mahaffy, J.; Martin-Cuadrado, A.; Mira, A.; Nulton, J.; Pasic, L.; Rayhawk, S.; Rodriguez-Mueller, J.; Rodriguez-Valera, F.; Salamon, P.; Srinagesh, S.; Thingstad, T.; Tran, T.; Thurber, R.; Willner, D.; Youle, M.; Rohwer, F. Viral and microbial community dynamics in four aquatic environments. *ISME J.* **2010**, *4* (6), 739−751.
(38) Gurevich, A.; Saveliev, V.; Vyahhi, N.; Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **2013**, *29* (8), 1072−1075.
(39) Hammer, O.; Harper, D. A. T.; Ryan, P. D. PAST: paleontological statistics software package for education and data analysis. *Palaeontol. Electron.* **2001**, *4* (1), 1−9.
(40) Roux, S.; Enault, F.; Robin, A.; Ravet, V.; Personnic, S.; Theil, S.; Colombet, J.; Sime-Ngando, T.; Debroas, D. Assessing the Diversity and Specificity of Two Freshwater Viral Communities through Metagenomics. *PLoS One* **2012**, *7* (3), e33641.
(41) Roux, S.; Tournayre, J.; Mahul, A.; Debroas, D.; Enault, F. Metavir 2: new tools for viral metagenome comparison and assembled virome analysis. *BMC Bioinf.* **2014**, *15* (1), 76.
(42) U.S. Environmental Protection Agency Website. http://www.epa.gov/glnpo/basicinfo.html.
(43) Minnesota Pollution Control Agency. Findings of fact-in the matter of a request for issuance of the SDS general permit MNG300000 for ballast water discharges from vessels transiting Minnesota state waters of Lake Superior, 2008. www.pca.state.mn.us/index.php/view-document.html?gid=10739.
(44) Drake, L. A.; Ruiz, G. M.; Galil, B. S.; Mullady, T. L.; Friedmann, D. O.; Dobbs, F. C. Microbial ecology of ballast water during a transoceanic voyage and the effects of open ocean exchange. *Mar. Ecol.: Prog. Ser.* **2002**, *233*, 13−20.
(45) Leichsenring, J.; Lawrence, J. Effect of mid-oceanic ballast water exchange on virus-like particle abundance during two trans-Pacific voyages. *Mar. Pollut. Bull.* **2011**, *62* (5), 1103−1108.
(46) U.S. Environmental Protection Agency. *Predicting future introductions of nonindigenous species to the Great Lakes*; National Center for Environmental Assessment: Washington, DC, 2008. http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=190305.
(47) Breitbart, M.; Salamon, P.; Andresen, B.; Mahaffy, J.; Segall, A.; Mead, D.; Azam, F.; Rohwer, F. Genomic analysis of uncultured marine viral communities. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (22), 14250−14255.
(48) Angly, F.; Felts, B.; Breitbart, M.; Salamon, P.; Edwards, R.; Carlson, C.; Chan, A.; Haynes, M.; Kelley, S.; Liu, H.; Mahaffy, J.; Mueller, J.; Nulton, J.; Olson, R.; Parsons, R.; Rayhawk, S.; Suttle, C.; Rohwer, F. The marine viromes of four oceanic regions. *PLoS Biol.* **2006**, *4* (11), 2121−2131.
(49) Bench, S. R.; Hanson, T. E.; Williamson, K. E.; Ghosh, D.; Radosovich, M.; Wang, K.; Wommack, K. E. Metagenomic characterization of Chesapeake bay virioplankton. *Appl. Environ. Microbiol.* **2007**, *73* (23), 7629−7641.
(50) Rosario, K.; Nilsson, C.; Lim, Y.; Ruan, Y.; Breitbart, M. Metagenomic analysis of viruses in reclaimed water. *Environ. Microbiol.* **2009**, *11* (11), 2806−2820.
(51) Williamson, S.; Allen, L.; Lorenzi, H.; Fadrosh, D.; Brami, D.; Thiagarajan, M.; McCrow, J.; Tovchigrechko, A.; Yooseph, S.; Venter, J. Metagenomic Exploration of Viruses throughout the Indian Ocean. *PLoS One* **2012**, *7* (10), e42047.
(52) Dinsdale, E. A.; Edwards, R. A.; Hall, D.; Angly, F.; Breitbart, M.; Brulc, J. M.; Furlan, M.; Desnues, C.; Haynes, M.; Li, L.; McDaniel, L.; Moran, M. A.; Nelson, K. E.; Nilsson, C.; Olson, R.; Paul, J.; Brito, B. R.; Ruan, Y.; Swan, B. K.; Stevens, R.; Valentine, D. L.; Thurber, R. V.; Wegley, L.; White, B. A.; Rohwer, F. Functional metagenomic profiling of nine biomes. *Nature* **2008**, *452* (7187), 629−32.
(53) Rooks, D. J.; Smith, D. L.; McDonald, J. E.; Woodward, M. J.; McCarthy, A. J.; Allison, H. E. 454-pyrosequencing: a molecular battiscope for freshwater viral ecology. *Genes* **2010**, *1* (2), 210−26.
(54) Abbai, N. S.; Govender, A.; Shaik, R.; Pillay, B. Pyrosequence analysis of unamplified and whole genome amplified DNA from hydrocarbon-contaminated groundwater. *Mol. Biotechnol.* **2012**, *50* (1), 39−48.
(55) Lopez-Bueno, A.; Tamames, J.; Velazquez, D.; Moya, A.; Quesada, A.; Alcami, A. High Diversity of the Viral Community from an Antarctic Lake. *Science* **2009**, *326* (5954), 858−861.
(56) Fancello, L.; Trape, S.; Robert, C.; Boyer, M.; Popgeorgiev, N.; Raoult, D.; Desnues, C. Viruses in the desert: a metagenomic survey of viral communities in four perennial ponds of the Mauritanian Sahara. *ISME J.* **2013**, *7* (2), 359−369.
(57) Logares, R.; Bråte, J.; Bertilsson, S.; Clasen, J. L.; Shalchian-Tabrizi, K.; Rengefors, K. Infrequent marine-freshwater transitions in the microbial world. *Trends Microbiol.* **2009**, *17* (9), 414−22.
(58) International Marine Organization Website. http://globallast.imo.org.
(59) Bain, M. B.; Cornwell, E. R.; Hope, K. M.; Eckerlin, G. E.; Casey, R. N.; Groocock, G. H.; Getchell, R. G.; Bowser, P. R.; Winton, J. R.; Batts, W. N.; Cangelosi, A.; Casey, J. W. Distribution of an Invasive Aquatic Pathogen (Viral Hemorrhagic Septicemia Virus) in the Great Lakes and Its Relationship to Shipping. *PLoS One* **2010**, *5* (4), e10156.
(60) Walker, P. J.; Winton, J. R. Emerging viral diseases of fish and shrimp. *Vet. Res.* **2010**, *41* (6), 51−75.
(61) Grimmett, S. G.; Warg, J. V.; Getchell, R. G.; Johnson, D. J.; Bowser, P. R. An unusual koi herpesvirus associated with a mortality event of common carp Cyprinus carpio in New York State, USA. *J. Wildl. Dis.* **2006**, *42* (3), 658−62.


DOI: 10.1021/acs.est.5b01633
*Environ. Sci. Technol.* 2015, 49, 8396−8407

(62) Garver, K. A.; Al-Hussinee, L.; Hawley, L. M.; Schroeder, T.; Edes, S.; LePage, V.; Contador, E.; Russell, S.; Lord, S.; Stevenson, R. M.; Souter, B.; Wright, E.; Lumsden, J. S. Mass mortality associated with koi herpesvirus in wild common carp in Canada. *J. Wildl. Dis.* **2010**, *46* (4), 1242−51.
(63) Cornwell, E. R.; Anderson, G. B.; Coleman, D.; Getchell, R. G.; Groocock, G. H.; Warg, J. V.; Cruz, A. M.; Casey, J. W.; Bain, M. B.; Bowser, P. R. Applying multi-scale occupancy models to infer host and site occupancy of an emerging viral fish pathogen in the Great Lakes. *J. Great Lakes Res.* **2015**, *41* (2), 520−9.
(64) Hedrick, R. P.; Gilad, O.; Yun, S.; Spangenberg, J. V.; Marty, G. D.; Nordhausen, R. W.; Kebus, M. J.; Bercovier, H.; Eldar, A. A herpesvirus associated with mass mortality of juvenile and adult koi, a strain of common carp. *J. Aquat. Anim. Health* **2000**, *12* (1), 44−57.
(65) Altug, G.; Gurun, S.; Cardak, M.; Ciftci, P. S.; Kalkan, S. The occurrence of pathogenic bacteria in some ships' ballast water incoming from various marine regions to the Sea of Marmara, Turkey. *Mar. Environ. Res.* **2012**, *81*, 35−42.
(66) Ducklow, H. W. Bacterial production and biomass in the oceans. In *Microbial Ecology of the Oceans*; Kirchman, D. L., Ed.; Wiley-Liss: New York, 2000.
(67) Lloyd's Register Website. http://www.lr.org/Images/BWTT_June%202011_tcm155-222616.pdf.
(68) Beerenwinkel, N.; Günthard, H. F.; Roth, V.; Metzner, K. J. Challenges and opportunities in estimating viral genetic diversity from next-generation sequencing data. *Front. Microbiol.* **2012**, *3*, 329.
(69) Delwart, E. Viral metagenomics. *Rev. Med. Virol.* **2007**, *17* (2), 115−131.
(70) Rosario, K.; Breitbart, M. Exploring the viral world through metagenomics. *Curr. Opin. Virol.* **2011**, *1* (4), 289−97.
(71) Bibby, K. Metagenomic identification of viral pathogens. *Trends Biotechnol.* **2013**, *31* (5), 275−9.
(72) Thomas, T.; Gilbert, J.; Meyer, F. Metagenomics - a guide from sampling to data analysis. *Microb. Inf. Exp.* **2012**, *2* (1), 3.


DOI: 10.1021/acs.est.5b01633
*Environ. Sci. Technol.* 2015, 49, 8396−8407