AEM Accepts, published online ahead of print on 6 July 2012
Appl. Environ. Microbiol. doi:10.1128/AEM.01705-12

Temporal dynamics and decay of putatively allochthonous and autochthonous viral genotypes in contrasting freshwater lakes

Ian Hewson*, Jorge G. Barbosa, Julia M. Brown, Ryan P. Donelan, James Eaglesham, Erin M. Eggleston, Brenna A. LaBarre

Department of Microbiology, Cornell University, Ithaca, USA

Running Head: Lake Viral Diversity and Dynamics

*Corresponding Author

Ian Hewson
Department of Microbiology
Cornell University
Wing Hall 403
Ithaca NY 14853
USA
Tel: +1 607 255 0151
Fax: +1 607 255 3904
hewson@cornell.edu

ABSTRACT

Aquatic viruses play important roles in the biogeochemistry and ecology of lacustrine ecosystems; however their composition, dynamics, and interactions with viruses of terrestrial origin are less extensively studied. We used a viral shotgun metagenomics approach to elucidate candidate autochthonous (i.e. produced within the lake) and allochthonous (i.e. those washed in from other habitats) viral genotypes for comparative study of their dynamics in lake waters. Based on shotgun metagenomes prepared from catchment soil and freshwater samples in two contrasting lakes (Cayuga Lake and Fayetteville Green Lake), we selected two putatively autochthonous viral genotypes (Phycodnaviruses likely infecting algae and Cyanomyoviruses likely infecting picocyanobacteria) and two putatively allochthonous viral genotypes (Geminiviruses likely infecting terrestrial plants and Circoviruses who infect unknown hosts but were common in soil libraries) for analysis using genotype-specific quantitative PCR (TaqMan©) applied to DNA from the 0.2 µm < virus < 0.02 µm size fraction of lake plankton. The abundance of autochthonous genotypes largely reflected expected host abundance, while the abundance of allochthonous genotypes corresponded with rainfall and storm events in the respective catchments, suggesting they may have been transported to the lake in runoff. The decay rates of allochthonous and autochthonous genotypes, assessed in incubations where all potential hosts were killed, were generally slower (0.13 – 1.50 % h^{-1}) than reported for marine virioplankton but similar to freshwater virioplankton. Both allochthonous and autochthonous viral genotypes were detected in greater concentrations in subsurface sediments compared to the water-sediment interface. Taken together, our data indicate that putatively allochthonous viruses are present in lake plankton and sediments, where their temporal dynamics reflect active transport to the lake during hydrological events and decay once there.

Keywords: Virus, Allochthonous, Green Lake, Decay, Cayuga Lake, Transport

INTRODUCTION

Viruses play crucial roles in the ecology and biogeochemistry of aquatic ecosystems by causing mortality of bacteria, archaea, and eukaryotes (23, 64). Viruses maintain large abundances in both fresh and salt waters (4, 46), typically exceeding bacterial abundances by 10 to 100 fold. Viral lysis accounts for a significant percentage of bacterial biomass daily in marine and freshwater ecosystems (57, 64, 66), releasing particulate organic matter within bacterioplankton to the dissolved organic matter pool. Since bacteria are the most abundant hosts, most virioplankton are believed to infect co-occurring bacterioplankton (45). However, in coastal waters, and in particular freshwater habitats, viruses may also be produced by allochthonous (terrestrial) hosts and viruses transported to aquatic habitats through rainfall events and runoff (38), groundwater discharge (22), sewage outfall (32, 33), or aerosols (1). Because the abundance of viruses in plankton does not typically vary strongly over time (24, 43), the production and decay of viruses is balanced in coastal ocean (28) and lake (59) habitats, however may be unbalanced in eutrophic marine waters (8). While there have been several studies examining that fate and persistence of human pathogenic viruses in aquatic habitats (7, 52), few address viruses of other terrestrial hosts.

In contrast to aquatic ecosystems, there have been fewer studies of the diversity and distribution of viruses in soils. Williamson et al (67) optimized extraction protocols for viruses across several soil types, and found that viral abundance ranged from $10^8 – 10^9$ VLP (virus like particle) g^{-1} dry weight with highest abundances in forest soils and lower abundance in agricultural soils. Interestingly, viral abundance was correlated with soil moisture and land use type, but was not related to soil texture (68, 69). Relatively few studies have investigated the diversity of viruses in soils using metagenomic approaches, and only one (21) that targeted viruses during sequencing. A comparison of rainforest, desert and prairie soils demonstrated that viral assemblages are unique with the highest diversity occurring in rainforest soils (~ 10^6 viral genotypes) (21). Viral assemblages in Delaware and Wisconsin soils compared to aquatic viruses reflected host composition (*Actinobateria* in soils, and cyanobacteria in aquatic habitats), with a lack of homology to known viruses and environmental sequence databases (53). Similarly, analyses of soil and aquatic viral communities demonstrated that these harbor mostly genetically distinct assemblages (53).

Viral decay in aquatic ecosystems is the result of UV light exposure (58), attachment to particles and sinking (29), heat-labile organic matter (notably nucleases) (42), and consumption by heterotrophic nanoflagellates (27). The magnitude of these factors in aquatic habitats has been subject to significant study in the past two decades, with most studies reporting rates of 2 – 4 % of viruses h^{-1} (64). VLP in sediment and soil habitats may have slower decay rates due to the chelating properties of organic matter that inhibit extracellular nucleases (35). For example, viable cyanophage have been recovered from deep sediments in the Saanich Inlet (56). The decay rates of viruses of eukaryotic hosts has not been extensively studied in natural aquatic settings, with the exception of pathogenic human viruses in groundwater (6) and nucleopolyhedroviruses of insects (31).

The aim of this study was to examine the dynamics of allochthonous and autochthonous viruses in two contrasting freshwater habitats, Cayuga Lake (CL) and Fayetteville Green Lake (FGL), and their respective catchments. First, we elucidated putatively autochthonous and allochthonous viral genotypes in the epilimnion of these lakes by shotgun metagenomic sequencing which provided targets for quantitative approaches to study the dynamics of specific viral genotypes. Second, we analyzed the abundance of several viral genotypes presumably infecting allochthonous and autochthonous hosts in surface waters over a 6 week period in summer to capture dynamics during rainfall events. We also examined the presence of viral genotypes within sediments as a possible environmental reservoir. Finally, we estimated decay rates of viral genotypes in incubations to compare the persistence of allochthonous with autochthonous within the lake habitat.

METHODS

Lake settings and sampling locations: Water column and soil samples were collected from two locations –Fayetteville Green Lake (FGL) and Cayuga Lake (CL; Fig. 1). FGL is a small (0.258 km^2), deep (mean depth = 52 m) meromictic (permanently stratified) marl lake in Onondaga County, New York. The epilimnion (surface waters) of FGL is unusually oligtrophic due to permanent chemical stratification and small watershed area (4.33 km^2) which is mainly state

parkland (15). The lake has been subject of several paleolimnological studies in the past 5 decades since the monimolimnion (sub-thermocline waters) of the lake is believed to represent an early earth analogue (72). The lake also contains littoral laminated $CaCO_3$ structures (bioherms) (14) which are believed to be caused by precipitation from photosynthesizing benthic microalgae. Dense blooms of coccoid cyanobacteria also cause $CaCO_3$ precipitation during whiting events (60). One hundred L of epilimnion water was collected from a concrete pier located on the northeast arm of FGL on 1 October 2010. Water column samples were collected using a sample-rinsed plastic bucket and transferred to 5, 20L HDPE cubatainers, which were placed into cooler for transport to Cornell University (Ithaca, NY). Soil samples were collected on the same date from agricultural cropland in the Northeast of the catchment (Fig. 1B). Five soil cores (3 cm^3) were collected using syringe corers from surface soil at randomly selected locations within a 10 m radius. Care was taken to avoid live vegetation (roots and leaves) within the cores, although soil cores contained leaf litter and humus from plants. Vegetation was primarily grasses (*Poaceae*). The soil cores were immediately placed into sterile 15 ml centrifuge tubes, which were subsequently frozen in liquid N_2 before transport to the laboratory.

CL is the largest of the New York finger lakes at 172 km^2 and the second deepest at 133 m (mean depth 55 m) and is a glacially-formed warm polymictic (mixing multiple times per year) lake (Fig. 1C). The lake experiences moderate productivity from the fall through spring during the ice-free months (47). The lake receives substantial inorganic nutrient inputs from surrounding catchment (~ 1150 km^2), including point-source inputs from teritiary-treated sewage from the city of Ithaca in the southern basin, and non-point source inputs from several small towns that line its shores (25). Water quality has, however, improved over the past two decades as a consequence of upgraded sewage treatment plants and invasion by zebra mussels (36). The lake is generally P-limited (25). Epilimnion water column samples (40L) were collected from a pier in Stewart Park, Ithaca (approximately 3 m from shore) using a sample-rinsed bucket on 10 October 2010 (Fig. 1C). Five soil cores were collected from pasture and croplands of the Estuarine Research Park, which is within the catchment of Cascadilla Creek, on the same date. The location of soil cores were chosen to reflect agricultural land use (*Sorghum* cropland and soils beneath a livestock pen). Soil cores were collected by syringe corers, transferred into sterile 15ml tubes, and frozen at -80^oC in the laboratory.

147 *Water Column Viral Preparation:* Water from both FGL and CL were sequentially filtered
148 through 142mm diameter 10 µm Nuclepore and 0.2 µm Durapore filters using positive air
149 pressure filtration. The filtrate was subject to tangential flow ultrafiltration using a recirculating
150 PREP/SCALE tangential flow ultrafilter (Millipore) with a 30 kDa molecular weight cut off. The
151 concentrate volume was 175 ml for both lakes. The presence of virus particles (and absence of
152 bacterial cells) was confirmed by SYBR Green I staining and epifluorescence microscopy (41,
153 44).

154 *Soil Viral Preparation:* One cm^3 from each of 5 soil cores were pooled for each metavirome,
155 which represented 2.0 g and 4.9 g dry weight of soil for CL and FGL, respectively. Soil was
156 homogenized in 35 mL 0.02µ-filtered PBS using a sterilized mortar and pestle for 2 min,
157 following the protocols of Thurber et al.(63) after ref (70). After homogenization, the sample
158 was briefly centrifuged at 3,000 x *g* for 5 min, after which the supernatant was syringe filtered
159 through 0.2 µm PES filters (VWR) to remove larger particles. The soil extracts were checked for
160 the presence of viruses by SYBR Green I epifluorescence microscopy.

161 *Metavirome Preparation:* Samples were prepared for sequencing following the protocols of
162 Thurber et al. (63). Water column concentrates were amended with 1 M NaCl before further
163 processing. Both soil and water column concentrates were precipitated by the addition of 10%
164 PEG-8000 (polyethyl glycol) at 4^{o}C overnight in sterilized round-bottom tubes (Oak Ridge).
165 After precipitation, samples were centrifuged at 13,000 x *g* for 30 mins in a fixed angle rotor
166 (Sorvall). The supernatant was decanted, and viral pellets resuspended in 3 mL 0.02 µm-filtered
167 PBS. Resuspended viruses were then filtered through 0.2 µm PES syringe filters (Acrodisc). The
168 filtrate was subject to density gradient ultracentrifugation, according to protocols in Thurber *et*
169 *al*., 2009, using density steps at 1.3, 1.5 and 1.7 g mL^{-1} and performed in a swing-bucket rotor at
170 60,000 x *g* for 2 h. Viruses in the 1.5 – 1.7 g mL^{-1} fraction (~ 1 mL) were removed using a sterile
171 syringe, and re-filtered through a 0.2 µm PES filter. The concentrate was treated with nucleases
172 (2.5U of DNAse and 0.25U RNAse) for 2 h at 37^{o}C to eliminate nucleic acids not incorporated
173 into viral capsids (40). DNA was extracted from 500 µl of each using the CTAB/EDTA
174 extraction procedure outlined in (63). Extracted DNA was tested for the presence of cellular
175 DNA by amplification of both 16S rRNA and 18S rRNA genes following protocols of (63). In

all samples, there were no amplicons generated after 30 cycles of PCR, indicating cell lysates were clear of contaminating cellular DNA.

DNA metaviromes were prepared for each sample using Genomiphi ©, where 1 µl of purified viral DNA was used as template material. Five Genomiphi reactions were pooled and purified using the DNEasy Tissue kit (Qiagen). Cleaned Genomiphi reactions were quantified by Pico Green fluorescence, and submitted for sequencing at EnGenCore (University of South Carolina), where each sample was run on $1/8^{th}$ of a picotiter plate using Titanium pyrosequencing chemistry.

Bioinformatic Analyses:- Reads were initially trimmed to remove low quality sequence reads and sequencing and amplification adapters (~ 2% of sequence reads in each library) using the CLC Genomics Workbench 4.0. In addition, reads < 100 bp were discarded. False duplicate reads (26), which represented only 0.2 – 1.0 % of viral reads were not removed from our dataset. Analysis of cleaned up sequence libraries followed two approaches. First, sequence reads were analyzed using the VIROME pipeline (http://virome.diagcomputing.org/) which compares read ORFs by BLASTp ($e < 0.001$) against several databases of reference peptides (Uniprot, SEED, ACLAME,COG, GO, KEGG). Secondly, libraries were assembled into contiguous sequences (contigs) using the CLC Genomics Workbench 4.0 to increase confidence in annotation of short reads. Reads were assembled using stringent assembly parameters: minimum overlap of 0.2, minimum similarity of 0.95, penalty for mismatch of 2, penalty for insertion of 3, and penalty for deletion of 3. We used an approach similar to the VIROME pipeline to assign contig phylogeny. Open Reading Frames (ORFs) were extracted from assembled contigs using the algorithm GetORF (http://emboss.sourceforge.net/apps/cvs/emboss/apps/getorf.html), and were subsequently compared by BLASTx against the non-redundant (nr) protein database at NCBI using an e-value cut-off of 10^{-3}. Contigs containing at least 1 viral gene ORF (referred to as "Viral Contigs" throughout this study) were subjected to further tBLASTx comparison against viral genomes at the CAMERA website. Phylogeny in all cases was taken as the BLASTx or tBLASTx hit (with lowest e-value) to the respective protein or genome database. Metagenomic libraries and contigs have been deposited at CAMERA under accession ·····.

Selection of targets for qPCR and primer design: Candidate 'allochthonous' viral genotypes were identified in soil metaviromes and 'autochthonous' viral genotypes identified within water

column metaviromes. Candidate contigs were identified based on observations of sequence read phylogeny variation between soil and water column metaviromes, confidence of contig annotation (i.e. lowest E-value), putative host organism, and contig sequence coverage. Selected candidate allochthonous and autochthonous contigs were used to design quantitative PCR primers and TaqMan© probes using the Primer3 program (Table 2 (50)). Oligonucleotide standards were used for generation of a qPCR standard curve for quantification of abundance (Supplemental Table 1). The linearity of standards was established for each primer/probe pair prior to downstream analyses of absolute viral abundance.

Quantitative PCR: Quantitative PCR (qPCR) reactions were carried out in duplicate, with duplicate standards and at least 2 negative controls per run, in an ABI 7300 Real-Time PCR machine. qPCR reactions (25 µl) comprised 1 X TaqMan Master Mix (Applied Biosystems International), 10 pmol forward and reverse primers and probes (Table 1), and 2 µl template DNA. qPCR reactions were subjected to an initial heating step at 50°C for 10min, followed by a hot start at 95°C for 5 min. The reactions were then thermally cycled at 95°C for 30s, followed by 1 minute at 60°C for a total of 60 cycles. The cycle threshold for calculation of gene abundance was calculated automatically by the ABI 7300 software. The R^2 value of standards was > 0.97. The number of genome copies per reaction was determined by comparison of cycle threshold crossing based on 8 standards from 10^8 to 10^1 copies reaction^{-1}. Data on the number of genotype copies per milliliter was converted to absolute abundance per mL lake water by dividing genome copies per reaction by 2, then multiplying them by total DNA extract volume, then dividing that number by volume in mL. qPCR results were only recorded if duplicate reactions both demonstrated amplification.

Survey of viral genotype abundance: Viral nucleic acids were collected every 7 d for 35 d total at the same location as used for metavirome preparation in FGL, and from LC3, a site offshore from the metavirome collection site in CL. Samples were collected using a sample-rinsed bucket from surface waters. Sixty ml of surface water was then syringe filtered through 0.2 µm PES and 0.02 µm Anotop filters, which were frozen in Whirlpak Bags before transport to Cornell. Corresponding meteorological data for FGL (Fayetteville, NY) and CL (Ithaca, NY) were obtained from the National Climatic Data Service (http://www.ncdc.noaa.gov/).

Viral Decay Experiments: To examine the ambient decay rates of viral genotypes within each lake environment, lake surface water was collected using a sample-rinsed bucket and dispensed into two acid-washed and sample-rinsed polycarbonate bottles (2 L). Bottles were kept cool for transport to Cornell, where they were immediately treated with 2 % NaN_3 to inhibit aerobic respiration (29). The bottles were incubated at ~ 25% attenuated surface irradiance on an array within Beebe Lake (an impoundment of Cascadilla Creek on the Cornell Campus), which has light and temperature conditions similar to the two lakes from which water was collected. Samples (50ml) were collected daily for the determination of viral genotype abundance, following the approach used in the survey of viral genotype abundance. The decay rates of selected viruses were determined by first calculating the proportion of viruses relative to initial abundance at each time point, then conducting linear regression of proportions against time using the XLStat plugin (Addinsoft SARL) in Microsoft Excel.

Survey of environmental reservoirs: We examined the abundance of viruses in benthos which may serve as potential reservoirs of settled viruses. Five sediment cores were collected from waters adjacent to the sampling location in FGL. Syringe corers were used to collect sediments from surface (0 to ~ 2 cm) and deep (2 to ~ 5 cm) sediments, which were immediately placed into sterile 15 ml centrifuge tubes and frozen on liquid nitrogen. DNA from the sediments was extracted using the Zymo Soil DNA kit (Zymo Research), which was applied to 200 mg of starting material.

RESULTS

Soil and plankton metaviromic libraries: A total of 35,997 sequence reads across all 4 libraries were obtained, representing 11.8 Mbp of genomic information (Table 2). Between 35 – 67% of reads assembled into contiguous sequences, resulting in 174 – 2,103 contigs per library, where the mean contig length was 400 – 518 bp, representing 0.07 – 1.08 MBp per library (Supplemental Table 2).

The majority of viral reads in DNA libraries did not match genomes or proteins of any organisms (Fig. 2A). Amongst protein hits, viruses comprised 21 – 29 % of total annotations, which were less numerous than to bacteria but greater than eukaryotes. Viral contigs (i.e. those harboring at least 1 ORF which matched viral proteins at $e < 0.001$) represented 2 – 10 % of total contigs. The lower percentage of viral contigs compared to reads suggests that there was greater coverage of viral contigs than contigs from other domains. The phylogenetic affiliation of viral contigs and sequence reads were similar, and were dominated by bacteriophage (73 - 84 % of viral reads and 50 – 89% of viral contigs). Amongst viral annotations, Siphoviruses comprised the greatest proportion of hits (33.7 ± 6.1 % mean ± S.E. of viral reads), followed by Myoviruses (in the water column these were mostly Cyanomyoviruses; 28.2 ± 5.4 % of viral reads) , Podoviruses (9.3 ± 2.3 % of viral reads) and Microviruses (7.0 ± 5.6 % of viral reads). Eukaryotic viruses were also detected and included Phycodnaviruses (1.3 ± 0.2% of viral reads) and Circoviruses (0.1 ± 0.1 % of viral reads) (Fig. 2B). Sequence reads matching most closely Geminiviruses were weakly detected only in the CL Soil library (5.3% of viral reads), however comprised a larger proportion of contigs (34 – 50% of viral contigs). Most Geminiviruses are very similar by nucleotide identity (49), while most Geminivirus-annotated reads in this study shared 26 - 49 % (mean = 38 ± 1 %) amino acid identity to the closest Geminivirus genome. Hence, the Geminiviruses observed in this study may represent distant relatives of known Geminiviruses, or representatives of closely related families (i.e. Cycloviruses and Circoviruses) (48). An estimate of viral diversity in each habitat was obtained by PHACCS (2). Generally, soil communities were more even than those in the water column, where predicted richness of communities was lower for CL soil and water column libraries than for the FGL libraries (Table 3).

Quantitative PCR primer design: Based on our analysis of metaviromic read and contig libraries, we selected candidate viral genotypes that were more strongly represented in terrestrial libraries than in water column libraries (i.e. allochthonous viruses) and vice-versa (autochthonous viruses; Supplemental Fig. 1). In both lakes, Geminivirus and Circovirus contigs were targeted as allochthonous viral genotypes, and Phycodnavirus and Cyanomyovirus contigs were targeted as autochthonous genotypes. The Geminivirus contigs selected most closely matched Bean Yellow Dwarf Virus and Tomato Mottle Virus in FGL and CL, respectively, while the Circovirus contigs most closely resembled two uncultivated circovirus genomes retrieved from the Chesapeake Bay (CB-A) and reclaimed water (RW-E; Table 4). The Cyanomyovirus contigs targeted were most

similar to *Prochlorococcus* phage P-SSM4, and Phycodnaviruses most similar to *Acanthocystis turfacea* Chlorella virus 1 (Table 4). The genes on contigs targeted by quantitative PCR were primarily replication-association proteins, however a predicted protein (gp7) was targeted for the FGL Cyanomyovirus genotype, a cyanophage-encoded phosphorus uptake gene (phoH) was targeted for the CL Cyanomyovirus genotype, and a hypothetical protein was used for both FGL and CL Phycodnavirus genotypes.

Time series analysis of viral genotype abundance: All 4 viral genotypes from FGL and 3 viral genotypes from CL were detected in extracted viral size fraction DNA, however the Geminivirus genotype was not detected in CL (Fig. 3). The greatest genotype abundance was for Cyanomyoviruses ($4.1 \pm 0.9 \times 10^5$ copies mL^{-1} in FGL and $1.5 \pm 0.9 \times 10^3$ copies mL^{-1} in CL, representing ~ 6% and 0.001% of total virioplankton in the lakes based on SYBR Green I counts in 2010 from a nearby location) and least for the Phycodnavirus genotype in FGL ($4.0 \pm 0.9 \times 10^1$ copies mL^{-1} or 0.00006% of total virioplankton abundance) and Circovirus genotype in CL (6.2 ± 3.4 copies mL^{-1} or 0.0000001% of virioplankton abundance).

The abundance of viral genotypes in the lakes changed over time (Fig. 3B and 3C). In FGL, the two allochthonous genotypes were below detection thresholds on the first sampling date, and after day 7 gradually decreased over the remainder of the sampling period. The Cyanomyovirus genotype maintained abundance until mid-August when abundances fell precipitously. Phycodnaviruses declined throughout the sampling period. In CL, the Circovirus genotype was detected in very low abundances only over the first 3 sampling dates then fell below the detection threshold of 1 copy mL^{-1}. The Geminivirus genotype was never detected. The Cyanomyovirus genotype decreased in abundance over time, while the Phycodnavirus genotype increased through the sampling period. The abundance of all viral genotypes in CL was much lower than in FGL.

Total precipitation in upstate NY (http://www.ncdc.noaa.gov/) was episodic during the sampling period, and preceded by a dry period in both catchments (Fig. 3A). Rainfall events were more frequent towards the end of the survey period. Significant rainfall events occurred after the first week's sampling when > 8 cm of rain fell in the FGL catchment, but was not measured in the CL catchment. Less than trace amounts of precipitation (i.e. < 0.25cm in Fig. 3A) were recorded

over the course of the sampling period at the Ithaca weather station, though the region experienced frequent storm events.

Environmental Reservoirs: All four viral genotypes probed in FGL were detected in sediment samples (Fig. 4), however their distribution in the sediment column varied with depth. The abundance of all four viral genotypes was greater in deeper sediments (2 – 5 cm) than in surface sediments (0 – 2 cm). The Phycodnavirus genotype were the only genotype that were in higher abundance in the sediments (200mg sediment~0.5 ml Lake Water) compared to lake water.

Decay Experiment:. Phycodnavirus (autochthonous) and Geminivirus (allochthonous) genotypes were detected in decay experiments using FGL water, while all four genotypes were detected in the CL decay incubations (Supplemental Fig. 2). Phycodnavirus, Cyanomyovirus, and Circovirus genotypes had large increases in abundance at the beginning or middle of the experiment. The CL Cyanomyovirus genotype demonstrated very rapid increase in viral abundance within the first 5 d of incubation, followed by a decline over the remainder of the experiment (Supplemetal Fig. 2). Because this likely represented release of progeny virus from dead or dying cells over the initial part of the experiment, we calculated viral genotype decay rates for these genotypes from 5 to 20 d. There was significant linear regression ($p < 0.05$) in only 2 of the 6 detected genotypes (Table 5). The fastest decay rate was for the CL Circovirus genotype (5.87% h^{-1}), however the regression was not significant. The CL Cyanomyovirus genotype had the second highest decay rate (1.50% h^{-1}), while FGL Phycodnavirus and both CL and FGL Geminivirus genotypes had similar decay rates (0.13 – 0.17 % h^{-1}).

DISCUSSION

Viruses play crucial roles in the ecology and biogeochemistry of aquatic ecosystems, hence the need for understanding their diversity, origins and fates. Our the data illustrate that virioplankton in freshwater ecosystems may represent a convergence of aquatic and terrestrial production, and their dynamics represent a confluence of terrestrial, limnological, and meteorological phenomena.

Diversity of lake viruses and comparison to soil viral communities: Amongst reads and contigs matching viral genomes, those that were well represented across all libraries were consistent with previous studies of phage diversity in seawater, sediments, and elsewhere (3, 10, 12, 13, 18, 21, 37). The large proportion of reads in FGL and CL water columns matching Cyanomyoviruses may reflect strong homology amongst cyanophage which infect picocyanobacteria (54, 55), or poor representation of myovirus genomes from other hosts in the databases used for comparison. The Φ29 amplification used in this and previous studies strongly selects for single stranded, circular genomes (48). This bias may have caused our observation of Circoviruses, Microviruses, and Geminiviruses in metaviromic libraries. The large proportion of unidentifiable reads based on alignment to genomes of sequenced viruses, bacteria or eukaryotes agrees with previous viral metagenomic studies (3, 10-12, 20, 37, 62). The unannotated sequence space may represent novel diversity, which in turn reflects the lack of representative genomes of viruses from lake and soil habitats (19, 53).

Geminiviruses are single stranded circular viruses that infect a wide suite of plants, including grasses (*Poaceae*) and nightshades (*Solanaceae*). They are transmitted primarily by insect vectors, including whiteflies and leafhoppers. A recent survey of whitefly – associated viruses revealed substantial diversity of Geminiviruses across several sampling locations (39). Their presence in soil may reflect accumulation of infected plant material in humus, or insect-derived matter in litter. Soils may form a reservoir of Geminiviruses which may be transported to plants by vectors or other means. Geminiviruses of plants are typically highly conserved in nucleotide sequences (39). These viral genotypes may represent viruses infecting other terrestrial host, since our annotation was based on translated nucleotide-protein comparison (i.e. BLASTx) where average homology to Geminiviruses was low. Phycodnaviruses are large dsDNA viruses that typically infect eukaryotic algae (17, 51). Circoviruses, which are circular ssDNA viruses (48) were large constituents of soil libraries and less common in freshwater libraries. Most commonly known to infect vertebrate hosts, they have also recently been observed in plankton of the Chesapeake Bay, near Bermuda, and in reclaimed wastewater (48).

Based on metaviromic analyses, we identified groups of viruses that were of putatively allochthonous (Geminivirus and Circovirus) and autochthonous (Phycodnavirus and Cyanomyovirus) origin (Supplemental Fig. 1; Table 4). We chose contiguous sequences of these

viral types that were detected in soil libraries and freshwater libraries. However, there were representatives of each of these groups in contig libraries from both habitat types. The allochthonous viruses and autochthonous viral genotypes differed in nucleic acid structure (both allochthonous viruses were ssDNA, while both autochthonous viruses were dsDNA), and targeted different host kingdoms (both allochthonous viruses infected eukaryotic organisms, while the autocthonous Cyanomyovirus infects cyanobacteria). Therefore, our choice of viruses represents a range of infection dynamics and physical characteristics.

Temporal dynamics of viral genotypes: The abundance of viral genotypes reflected the abundance of putative hosts, with the FGL and CL Cyanomyovirus genotypes having the greatest overall abundance (~5 - 10% of virioplankton abundance by SYBR Green Microscopy – which was 6.3×10^6 VLP mL^{-1} in August 2009; unpub. data). This observation agrees with PHACCS analysis (Table 3) which found that the dominant viral genotype within FGL and CL plankton represented 10 – 15% of total viral abundance. FGL is characterized by very large populations of pelagic cyanobacteria, which cause whiting events in spring and fall due to photosynthesis-induced pH changes and their subsequent impacts on carbonate chemistry (60). Picocyanobacteria comprise ~1 % of total bacterioplankton in FGL (1.7×10^7 cells mL^{-1} in August 2009; unpub. data). The abundance of the Cyanomyovirus genotype in FGL was higher than in CL. The low abundance of the Phycodnavirus genotype in FGL relative to others, including putative allochthonous viruses, was expected. Phytoplankton communities in the lake are dominated by cyanobacteria, while eukaryotic algae, like *Chlamydomonas* and small pennate diatoms, typically comprise a much smaller proportion of lake microbial flora (61).

The temporal dynamics of putatively allochthonous and autochthonous viral genotypes may reflect both rainfall events and the dynamics of phytoplankton over the sampling period. The Circovirus genotypes were observed only after the first week of sampling when a large rainfall event occurred in the FGL catchment (delivering ~ 8cm of rain in 24 h after a dry period of several weeks) and were only detected in CL at the start of the sampling period. The FGL Geminivirus genotype followed a similar trend as the FGL Circovirus genotype, but the CL Geminivirus genotype was not detected. The absence of the CL Gemninvirus may be related to differences in autochthonous viral composition between 2010, when metagenomes were prepared, and 2011, when the sampling was conducted.In contrast, the Cyanomyovirus and

411 Phycodnavirus genotypes were detected on most sampling dates in both lakes. Our consistent
412 autochthonous viral detection, but ephemeral allochthonous viral detection, suggests that the
413 presence of the latter may reflect a balance between the arrival of viruses to the catchment
414 (possibly via rainfall) and their decay or sedimentation within the lake environment.

415

416 *Viral presence in environmental reservoirs:* Viral genotypes of both allochthonous and
417 autochthonous origin were detected within sediments, and interestingly were more abundant (per
418 unit weight of sediment) in deeper sediment horizons. These results suggest that sediments may
419 be a sink for both putatively allochthonous and autochthonous virus particles within lake
420 habitats, as has been observed in coastal marine ecosystems. For example, an investigation of
421 cyanophage within sediments off the Saanich peninsula, which is considered a high deposition
422 area, revealed the presence of infective cyanophage 20 m beneath the sediment surface, which
423 are estimated to be several hundred years old (56). The greater abundance of viral genotypes in
424 deeper sediments relative to shallow ones suggests that they may be entrained and concentrated
425 there, possibly on particle surfaces or colloids. Hence, sediments may form environmental
426 reservoirs of allochthonous and autochthonous viral genotypes. We speculate that these may
427 become resuspended when sediments are disturbed in a manner previously observed in human
428 viruses (9). These results raise interesting questions about the physical and infectivity decay of
429 viruses in sediment habitats.

430

431 *Decay of viral genotypes:* Viral decay in aquatic ecosystems may occur via several mechanisms
432 (42). By stopping virus production and the production of new ectoenzymes by heterotrophic
433 bacteria, then monitoring the decrease in viral abundance, it is possible to estimate viral decay
434 rates (28). We used a similar approach to estimate decay of allochthonous and autochthonous
435 viral genotypes in incubations treated with NaN_3 (c.f. KCN used in previous studies, not used in
436 our study due to environmental concern). To the best of our knowledge, this investigation
437 represents the first genotype-specific investigation of decay using this approach. Previous studies
438 have observed the abundance of decay-resistant polyhedra of nucleopolyhedroviruses in water
439 troughs over time, and detected their presence for several weeks (31). However, investigation of

native phage and viruses, or ssDNA viruses via our approach, has not been previously conducted.

The large increase in the cyanophage, algal viruses, and Circovirus genotypes over the first 5 d of sampling suggests that they may have been released by decaying organisms within the incubations. To account for this, decay rates were calculated for the latter part of the experiments using a previously established approach (30). The rates of virus production across 5 genotypes detected over the course of the experiment were generally lower than reported for entire virioplankton communities in most marine waters (5, 28, 42) and were in line with decay rates estimated for Lake Bourget (59). The CL Circovirus genotype decayed more rapidly than marine and lake virioplankton, however our estimates are in line with a previous study of cyanophage decay in freshwater lakes (34). Previous study of individual viruses have reported a range of decay rates, which depend on the physicochemical conditions of waters in which the phage inhabit (16, 65).The regression used to calculate decay rate was not significant for the CL Circovirus genotype, nor for the FGL Phycodnavirus or CL Geminivirus genotypes. Of the two significant regressions, the decay rate of the cyanophage genotype was almost an order of magnitude higher than for the FGL Geminivirus genotype. The high variability in genotype decay rates suggest that there was no difference between the decay rates of ssDNA and dsDNA eukaryotic virus genotypes.

The stability and persistence of viruses of allochthonous origin has implications for lake water use, as potentially pathogenic viruses of agriculturally important species may be present in lake virioplankton for several days after introduction. For example, the use of lake water containing viruses of agriculturally important crops may give rise to new terrestrial infections when used for irrigation (73). It is important to note, however, that our data applies to the detection of viral nucleic acids only. Infectivity of virus particles may be lost well before physical loss of the viruses (42, 58, 71). The extent to which the re-emergence of allochthonous aquatic viruses in terrestrial habitats will occur remains unknown.

Conclusions: To the best of our knowledge, this study is the first to synchronously investigate the dynamics of allochthonous and autochthonous viruses in aquatic habitats through a combined shotgun metagenomics and quantitative PCR approach. Our data suggest that putatively allochthonous viruses comprise a component of virioplankton in lakes, and that their abundance

may be linked to local factors influencing transport from the catchment. The decay rates of the allochthonous viral genotypes examined in this study were the same as for the autochthonous genotypes, which were generally lower than reported for seawater but similar to those in lakes. The detection of genotypes within sediments of the lake suggests that this compartment may be a reservoir of allochthonous and autochthonous viruses. This study raises interesting questions about the possibility of transport of allochthonous viruses back into the terrestrial environment from aquatic habitats perhaps by anthropogenic processes such as irrigation or by vector transport.

ACKNOWLEDGEMENTS

The authors are grateful to: K. E. Wommack, J. Basheel, D. Nasko and S. Polson (U Delaware) for metagenomic annotation using VIROME and helpful comments on bioinformatics analyses; N. Hairston, L. Schaffner (Cornell) and K. Meyer (U. Maryland) for assistance in sample collection in Cayuga Lake; M.B. Wilhelm (Cornell) for assistance in sampling of Fayetteville Green Lake; J. Jones (EnGenCore) for assistance with sequencing; and M. Breitbart (U South Florida) for helpful conversations on data interpretation. This manuscript is dedicated to the memory of J. Van Deusen (New York State Office of Parks Recreation and Historic Restoration). This work was performed under permits 2010-GL-001 and 2011-GL-001. This work was supported by Federal Formula Funds Hatch Grant NYS-189489 awarded to IH and JWC, and NSF grant DEB1028898 awarded to IH and N. Hairston.

REFERENCES

1. **Aller, J. Y., M. R. Kuznetsova, C. J. Jahns, and P. F. Kemp.** 2005. The sea surface microlayer as a source of viral and bacterial enrichment in marine aerosols. J Aerosol Sci **36:**801-812.
2. **Angly, F., B. Rodriguez-Brito, D. Bangor, P. McNairnie, M. Breitbart, P. Salamon, B. Felts, J. Nulton, J. Mahaffy, and F. Rohwer.** 2005. PHACCS, an online tool for estimating the structure and diversity of uncultured viral communities using metagenomic information. BMC Bioinformatics **6**.
3. **Angly, F. E., B. Felts, M. Breitbart, P. Salamon, R. A. Edwards, C. Carlson, A. M. Chan, M. Haynes, S. Kelley, H. Liu, J. M. Mahaffy, J. E. Mueller, J. Nulton, R. Olson, R. Parsons, S. Rayhawk, C. A. Suttle, and F. Rohwer.** 2006. The marine viromes of four oceanic regions. PLoS Biol **4:**2121-2131.
4. **Bergh, O., K. Y. Borsheim, G. Bratbak, and M. Heldal.** 1989. High abundance of viruses found in aquatic environments. Nature **340:**467-468.
5. **Bettarel, Y., T. Bouvier, and M. Bouvy.** 2009. Viral persistence in water as evaluated from a tropical/temperate cross-incubation. J Plankt Res **31:**909-916.
6. **Bitton, G., S. R. Farrah, R. H. Ruskin, J. Butner, and Y. J. Chou.** 1983. Survival of pathogenic and indicator organisms in groundwater. Ground Water **21:**405-410.

7. **Boehm, A. B., K. M. Yamahara, D. C. Love, B. M. Peterson, K. McNeill, and K. L. Nelson.** 2011. Covariation and photoinactivation of traditional and novel indicator organisms and human viruses at a sewage-impacted marine beach. Environ Sci Technol **43:**8046 - 8052.

8. **Bongiorni, L., M. Magagnini, M. Armeni, R. Noble, and R. Danovaro.** 2005. Viral production, decay rates, and life strategies along a trophic gradient in the north Adriatic sea. Appl Environ Microbiol **71:**6644-6650.

9. **Bosch, A., F. Lucena, R. Girones, and J. Jofre.** 1988. Occurrence of enteroviruses in marine sediment along the coast of Barcelona, Spain. Can J Microbiol **34:**921-924.

10. **Breitbart, M., B. Felts, S. Kelley, J. M. Mahaffy, J. Nulton, P. Salamon, and F. Rohwer.** 2004. Diversity and population structure of a near-shore marine-sediment viral community. Proc Roy Soc London B **271:**565-574.

11. **Breitbart, M., I. Hewson, B. Felts, J. M. Mahaffy, J. Nulton, P. Salamon, and F. Rohwer.** 2003. Metagenomic analyses of an uncultured viral community from human feces. J Bacteriol **185:**6220-6223.

12. **Breitbart, M., A. Hoare, A. Nitti, J. Siefert, M. Haynes, E. Dinsdale, R. Edwards, V. Souza, F. Rohwer, and D. Hollander.** 2009. Metagenomic and stable isotopic analyses of modern freshwater microbialites in Cuatro Cienegas, Mexico. Environ Microbiol **11:**16-34.

13. **Breitbart, M., P. Salamon, B. Andresen, J. Mahaffy, A. Segall, D. Mead, F. Azam, and F. Rohwer.** 2002. Genomic analysis of uncultured marine viral communities. Proc Nat Acad Sci USA **99:**14250-14255.

14. **Brunskill, G. J.** 1969. Fayetteville Green Lake, New York. II. Precipitation and sedimentation of calcite in a meromictic lake with laminated sediments. Limnol Oceanogr **14:**830-847.

15. **Brunskill, G. J., and S. D. Ludlam.** 1969. Fayetteville Green Lake, New York. I. Physical and chemical limnology. Limnol Oceanogr **14:**817-829.

16. **Caballero, S., F. X. Abad, F. Loisy, F. S. Le Guyader, J. Cohen, R. M. Pinto, and A. Bosch.** 2004. Rotavirus virus-like particles as surrogates in environmental persistence and inactivation studies. Appl Environ Microbiol **70:**3904-3909.

17. **Clasen, J. L., and C. A. Suttle.** 2009. Identification of freshwater phycodnaviridae and their potential phytoplankton hosts, using DNA *pol* sequence fragments and a genetic-distance analysis. Appl Environ Microbiol **75:**991-997.

18. **Desnues, C., B. Rodriguez-Brito, S. Rayhawk, S. Kelley, T. Tran, M. Haynes, H. Liu, M. Furlan, L. Wegley, B. Chau, Y. J. Ruan, D. Hall, F. E. Angly, R. A. Edwards, L. L. Li, R. V. Thurber, R. P. Reid, J. Siefert, V. Souza, D. L. Valentine, B. K. Swan, M. Breitbart, and F. Rohwer.** 2008. Biodiversity and biogeography of phages in modern stromatolites and thrombolites. Nature **452:**340-U345.

19. **Djikeng, A., R. Kuzmickas, N. G. Anderson, and D. J. Spiro.** 2009. Metagenomic analysis of RNA viruses in a fresh water lake. PLoS One **4:**e7264.

20. **Edwards, R. A., B. Rodriguez-Brito, L. Wegley, M. Haynes, M. Breitbart, D. M. Peterson, M. O. Saar, S. Alexander, E. C. Alexander, and F. Rohwer.** 2006. Using pyrosequencing to shed light on deep mine microbial ecology. BMC Genomics **7**

21. **Fierer, N., M. Breitbart, J. Nulton, P. Salamon, C. Lozupone, R. Jones, M. Robeson, R. A. Edwards, B. Felts, S. Rayhawk, R. Knight, F. Rohwer, and R. B. Jackson.** 2007. Metagenomic and small-subunit rRNA analyses reveal the genetic diversity of bacteria, archaea, fungi, and viruses in soil. Appl Environ Microbiol **73:**7059-7066.

22. **Fong, T. T., L. S. Mansfield, D. L. Wilson, D. J. Schwab, S. L. Molloy, and J. B. Rose.** 2007. Massive microbiological groundwater contamination associated with a waterborne outbreak in Lake Erie, South Bass Island, Ohio. Environ Health Perspectives **115:**856-864.

23. **Fuhrman, J. A.** 1999. Marine viruses and their biogeochemical and ecological effects. Nature **399:**541-548.

24. **Fuhrman, J. A., and R. T. Noble.** 1995. Viruses and protists cause similar bacterial mortality in coastal seawater. Limnol Oceanogr **40:**1236-1242.

25. **Godfrey, P. J.** 1982. The eutrophication of Cayuga Lake - a historical analysis of the phytoplanktons response to phosphate detergents. Freshwater Biol **12:**149-166.

26. **Gomez-Alvarez, V., T. K. Teal, and T. M. Schmidt.** 2009. Systematic artifacts in metagenomes from complex microbial communities. ISME J **3:**1314-1317.

27. **Gonzalez, J. M., and C. A. Suttle.** 1993. Grazing by marine nanoflagellates on viruses and virus-sized particles - ingestion and digestion. Mar Ecol Progr Ser **94:**1-10.

28. **Heldal, M., and G. Bratbak.** 1991. Production and decay of viruses in aquatic environments. Mar Ecol Progr Ser **72:**205-212.

29. **Hewson, I., and J. A. Fuhrman.** 2003. Viriobenthos production and virioplankton sorptive scavenging by suspended sediment particles in coastal and pelagic waters. Microb Ecol **46:**337-347.

30. **Hewson, I., S. R. Govil, D. G. Capone, E. J. Carpenter, and J. A. Fuhrman.** 2004. Evidence of Trichodesmium viral lysis and potential significance for biogeochemical cycling in the oligotrophic ocean. Aquat Microb Ecol **36:**1-8.

31. **Holmes, S. B., W. E. Fick, D. P. Kreutzweiser, P. M. Ebling, L. S. England, and J. T. Trevors.** 2008. Persistence of naturally occurring and genetically modified *Choristoneura fumiferana* nucleopolyhedroviruses in outdoor aquatic microcosms. Pest Manage Sci **64:**1015-1023.

32. **Katayama, H., K. Okuma, H. Furumai, and S. Ohgaki.** 2004. Series of surveys for enteric viruses and indicator organisms in Tokyo Bay after an event of combined sewer overflow. Water Sci Technol **50:**259-262.

33. **Krikelis, V., P. Markoulatos, and N. Spyrou.** 1986. Viral pollution of coastal waters resulting from the disposal of untreated sewage effluents. Water Sci Technol **18:**43-48.

34. **Liu, W., K. Cheng, Y. J. Zhao, Y. J. Wu, C. F. Cai, S. Y. Lu, and X. C. Jin.** 2011. Cyanophage decay and its causes in eutrophic freshwater. J Food Agric Environ **9:**963-966.

35. **Lorenz, M. G., B. W. Aardema, and W. E. Krumbein.** 1981. Interaction of marine sediment with DNA and DNA availability to nucleases. Mar Biol **64:**225-230.

36. **Matthews, D. A., S. V. Stehman, and S. W. Effler.** 2002. Limnological and statistical issues for monitoring the impact of the Lake Source Cooling Facility: Cayuga Lake, NY. Lake and Reservoir Management **18:**239-256.

37. **McDaniel, L., M. Breitbart, J. Mobberley, A. Long, M. Haynes, F. Rohwer, and J. H. Paul.** 2008. Metagenomic analysis of lysogeny in Tampa Bay: Implications for prophage gene expression. PLoS One **3:**e3363.

38. **McLeod, M., J. Aislabie, J. Ryburn, A. McGilla, and M. Taylor.** 2003. Microbial and chemical tracer movement through two Southland soils, New Zealand. Aust J Soil Res **41:**1163-1169.

39. **Ng, T. F. F., S. Duffy, J. E. Polston, E. Bixby, G. E. Vallad, and M. Breitbart.** 2011. Exploring the diversity of plant DNA viruses and their satellites using vector-enabled metagenomics on whiteflies. PLoS One **6**.

40. **Ng, T. F. F., C. Manire, K. Borrowman, T. Langer, L. Ehrhart, and M. Breitbart.** 2009. Discovery of a novel single-stranded DNA virus from a sea turtle fibropapilloma by using viral metagenomics. J Virol **83:**2500-2509.

41. **Noble, R. T., and J. A. Fuhrman.** 1998. Use of SYBR Green I rapid epifluoresence counts of marine viruses and bacteria. Aquat Microb Ecol **14:**113-118.

42. **Noble, R. T., and J. A. Fuhrman.** 1997. Virus decay and its causes in coastal waters. Appl Environ Microbiol **63:**77-83.

43. **Parsons, R. J., M. Breitbart, M. W. Lomas, and C. A. Carlson.** 2012. Ocean time-series reveals recurring seasonal patterns of virioplankton dynamics in the northwestern Sargasso Sea. ISME J **6:**273-284.

44. **Patel, A., R. Noble, J. Steele, M. Schwalbach, I. Hewson, and J. Fuhrman.** 2007. Virus and prokaryote enumeration from planktonic aquatic environments by epifluorescence microscopy with SYBR Green I. Nature Protocols **2:**269-276.

45. **Proctor, L. M., and J. A. Fuhrman.** 1990. Viral mortality of marine bacteria and cyanobacteria. Nature **343:**60-61.

46. **Proctor, L. M., J. A. Fuhrman, and M. C. Ledbetter.** 1988. Marine bacteriophages and bacterial mortality. Eos **69:**1111-1112.

47. **R.T., O.** 1978. The limnology of Cayuga Lake, p. 2-121. *In* B. A. (ed.), Lakes of New York State. I, Ecology of the Finger Lakes. Academic Press, New York.

48. **Rosario, K., S. Duffy, and M. Breitbart.** 2009. Diverse circovirus-like genome architectures revealed by environmental metagenomics. J Gen Virol **90:**2418-2424.

49. **Rosario, K., M. Marinov, D. Stainton, S. Kraberger, E. J. Wiltshire, D. A. Collings, M. Walters, D. P. Martin, M. Breitbart, and A. Varsani.** 2011. Dragonfly cyclovirus, a novel single-stranded DNA virus discovered in dragonflies (Odonata: Anisoptera). J Gen Virol **92:**1302-1308.

50. **Rozen, S., and H. Skaletsky.** 2000. Primer3 on the WWW for general users and for biologist programmers, p. 365-386. *In* S. Krawertz and S. Misener (ed.), Bioinformatics Methods and Protocols. Humana Press, Totowa, NJ.

632 51. **Short, S. M., O. Rusanova, and M. A. Staniewski.** 2011. Novel phycodnavirus genes
633 amplified from Canadian freshwater environments. Aquat Microb Ecol **63:**61-67.
634 52. **Skraber, S., J. Schijven, R. Italiaander, and A. M. D. Husman.** 2009. Accumulation
635 of enteric bacteriophage in fresh water sediments. J Water Health **7:**372-379.
636 53. **Srinivasiah, S., J. Bhavsar, K. Thapar, M. Liles, T. Schoenfeld, and K. E.**
637 **Wommack.** 2008. Phages across the biosphere: contrasts of viruses in soil and aquatic
638 environments. Res Microbiol **159:**349-357.
639 54. **Sullivan, M. B., M. L. Coleman, P. Weigele, F. Rohwer, and S. W. Chisholm.** 2005.
640 Three Prochlorococcus cyanophage genomes: Signature features and ecological
641 interpretations. PLoS Biol **3:**790-806.
642 55. **Sullivan, M. B., J. B. Waterbury, and S. W. Chisholm.** 2003. Cyanophages infecting
643 the oceanic cyanobacterium *Prochlorococcus*. Nature **424:**1047-1051.
644 56. **Suttle, C. A.** 2000. Cyanophages and their role in the ecology of cyanobacteria, p. 563-
645 589. *In* A. Whitton and M. Potts (ed.), The Ecology of Cyanobacteria, vol. 1. Kluwer
646 Academic, Amsterdam.
647 57. **Suttle, C. A.** 2007. Marine viruses - major players in the global ecosystem. Nature Rev
648 Microbiol **5:**801-812.
649 58. **Suttle, C. A., and C. Feng.** 1992. Mechanisms and rates of decay of marine viruses in
650 seawater. Appl Environ Microbiol **58:**3721-3729.
651 59. **Thomas, R., L. Berdjeb, T. Sime-Ngando, and S. Jacquet.** 2011. Viral abundance,
652 production, decay rates and life strategies (lysogeny versus lysis) in Lake Bourget
653 (France). Environ Microbiol **13:**616-630.
654 60. **Thompson, J. B., F. G. Ferris, and D. A. Smith.** 1990. Geomicrobiology and
655 sedimentology of the mixolimnion and chemocline in Fayetteville Green Lake, New
656 York. Palaios **5:**52-75.
657 61. **Thompson, J. B., S. SchultzeLam, T. J. Beveridge, and D. J. DesMarais.** 1997.
658 Whiting events: Biogenic origin due to the photosynthetic activity of cyanobacterial
659 picoplankton. Limnol Oceanogr **42:**133-141.
660 62. **Thurber, R. L. V., K. L. Barott, D. Hall, H. Liu, B. Rodriguez-Mueller, C. Desnues,**
661 **R. A. Edwards, M. Haynes, F. E. Angly, L. Wegley, and F. L. Rohwer.** 2008.

662 Metagenomic analysis indicates that stressors induce production of herpes-like viruses in
663 the coral Porites compressa. Proc Nat Acad Sci USA **105:**18413-18418.
664 63. **Thurber, R. V., M. Haynes, M. Breitbart, L. Wegley, and F. Rohwer.** 2009.
665 Laboratory procedures to generate viral metagenomes. Nature Protocols **4:**470-483.
666 64. **Weinbauer, M. G.** 2004. Ecology of prokaryotic viruses. FEMS Microbiol Rev **28:**127-
667 181.
668 65. **Wells, L. E., and J. W. Deming.** 2006. Effects of temperature, salinity and clay particles
669 on inactivation and decay of cold-active marine Bacteriophage 9A. Aquat Microb Ecol
670 **45:**31-39.
671 66. **Wilhelm, S. W., and C. A. Suttle.** 1999. Viruses and nutrient cycles in the sea.
672 Bioscience **49:**781-788.
673 67. **Williamson, K., M. Radosevich, and K. Wommack.** 2005. Abundance and diversity of
674 viruses in six Delaware soils. Appl Environ Microbiol **71:**3119-3125.
675 68. **Williamson, K. E., M. Radosevich, D. W. Smith, and K. E. Wommack.** 2007.
676 Incidence of lysogeny within temperate and extreme soil environments. Environ
677 Microbiol **9:**2563-2574.
678 69. **Williamson, K. E., J. B. Schnitker, M. Radosevich, D. W. Smith, and K. E.**
679 **Wommack.** 2008. Cultivation-based assessment of lysogeny among soil bacteria. Microb
680 Ecol **56:**437-447.
681 70. **Williamson, K. E., K. E. Wommack, and M. Radosevich.** 2003. Sampling natural viral
682 communities from soil for culture-independent analyses. Appl Environ Microbiol
683 **69:**6628-6633.
684 71. **Wommack, K. E., R. T. Hill, T. A. Muller, and R. R. Colwell.** 1996. Effects of
685 sunlight on bacteriophage viability and structure. Appl Environ Microbiol **62:**1336-1341.
686 72. **Zerkle, A. L., A. Kamyshny, L. R. Kump, A. L. Riccardi, M. A. Arthur, and J.**
687 **Farquhar.** 2009. Biogeochemical sulfur cycling in meromictic Fayetteville Green Lake,
688 NY. Geochimica Et Cosmochimica Acta **73:**A1501-A1501.
689 73. **Zhang, T., M. Breitbart, W. H. Lee, J. Q. Run, C. L. Wei, S. W. L. Soh, M. L.**
690 **Hibberd, E. T. Liu, F. Rohwer, and Y. J. Ruan.** 2006. RNA viral community in human
691 feces: Prevalence of plant pathogenic viruses. PLoS Biol **4:**108-118.

692

693

694 **Table 1:** Sequences for quantitative PCR primers and probes used in this study. The sequence for oligonucleotide standards is
695 presented in Supplemental Table 1. Primers were designed using Primer3 based on assembled contiguous sequences from the water
696 column and soil of Fayetteville Green Lake (FGL) and Cayuga Lake (CL).

Genotype	Forward Primer (5'-3')	Reverse Primer (5'-3')	Probe Sequence (5' – 3')
FGL Phycodnavirus	TTTCATTTTTGCCGATGGAT	TTGTGCGTACAATTCGTCGT	TCGCCAAGCCCCATATCAGGA
FGL Cyanomyovirus	GCACAGATCAGCACCAGTGT	GGATTAGCAGGCAGACGAAG	CACTGGCGCTACATCTGGATCGA
FGL Circovirus	CCATCCCACCATTTATTTGC	GGGTCCATCTGGAACTGGTA	GGCATTGGGAAAAAGCTCTCTTGC
FGL Geminivirus	TCCGAGGAGCAGAGTATCGT	ATGCTAATATCGGGCGAGTG	TTCACCGTCCTTGCGGGCAT
CL Phycodnavirus	GCAGGCCGAACAGAAGATAC	AAGGCACTGCGACAGGTTAT	GGCGCTTCTCCAGCATACAGCA
CL Cyanomyovirus	ACGGTATCAAGGCCAATGAG	CGACCACCGAAGTAGAAGGA	TGTCCAACTGTTAGGTCAGTGGGGT
CL Circovirus	GGAAGTCAAGGGTTCGTCAA	TACCATTCTCGGGGATCAAG	GCCGAGGTTATCTGGATCACCAGC
CL Geminivirus	GGAATGCACCTCCGATAAGA	AATGTCGTACCGTTGGAAGC	GCCTGTGTCTTCGTACGTAAGCTTCC

697

698

699 **Table 2:** Viral metagenomic characteristics of Cayuga (CL) and Fayetteville Green Lake (FGL)
700 soil and freshwater libraries.

Location	Habitat	Raw Reads	Reads after Trimming	Avg Length (bp)	Avg G+C %
CL	Soil	1,714	1,673	320	48
CL	Water	11,901	11,671	349	44
FGL	Soil	6,406	6,288	321	48
FGL	Water	16,648	16,365	317	43

701

Table 3: Predicted diversity and richness of metaviromic libraries assessed using the phage communities from contig spectra (PHACCS) algorithm (2). The analysis was based on contigs assembled from sequence reads using the CLC Genomics Workbench 4.0 using a minimum overlap of 0.8 and similarity of 95%.

Lake	Habitat	Richness	Evenness	Most Abundant (%)	Shannon-Wiener Index
FGL	Water	10,000	0.8786	10.8	8.0922
FGL	Soil	20,000	0.2692	44.2	2.6660
CL	Water	13	0.9759	15.5	2.5032
CL	Soil	1,452	0.2844	67.1	2.0707

708 **Table 4:** Putatively autochthonous and allochthonous viral contig closest matches in the non-redundant (nr) database at Genbank by
709 BLASTx using an expect score cut off of $10^{-3.}$ Avg Cov. = Average coverage.

Genotype	Contig Length (bp)	Avg Cov.	Closest Match	Protein	GenBank ID	Amino Acid ID (%)
FGL Phycodnavirus	1,558	9.4	*Acanthocystis turfacea* Chlorella virus 1	Hypothetical prot. ATCV1_Z838L	ABT16972	54
FGL Cyanomyovirus	1,825	7.0	*Prochlorococcus* phage P-SSM4	gp7	AAX46881	32
FGL Circovirus	364	4.0	Circovirus-like genome CB-A	Replicase	ACQ78166	47
FGL Geminivirus	2,180	26.4	Bean yellow dwarf virus	Replication-associated protein	CAA71908	28
CL Phycodnavirus	471	9.6	*Acanthocystis turfacea* Chlorella virus 1	Hypothetical prot. ATCV1_Z838L	ABT16972	72
CL Cyanomyovirus	3,269	8.0	*Prochlorococcus* phage P-SSM4	phoH	AAX46998	44
CL Circovirus	1,302	2.7	Circovirus-like genome RW-E	Replicase	ACQ78164	59
CL Geminivirus	1,194	3.5	Tomato mottle virus	Replicative protein	AAC32414	33

710

711

712 **Table 5:** Regression statistics for CL and FGL viral genotypes in decay incubations. *r* =
713 regression coefficient; *p* = statistical significance of regression; * = $p < 0.05$

Lake	Genotype	Slope	Intercept	*r*	Decay Rate (% h^{-1})	*p*
FGL	Geminivirus	-0.04	0.84	0.60	0.17	*
FGL	Phycodnavirus	-0.03	0.43	0.40	0.13	ns
CL	Circovirus	-1.41	24.44	0.54	5.87	ns
CL	Geminivirus	-0.03	1.02	0.22	0.13	ns
CL	Cyanomyovirus	-0.36	6.63	0.59	1.50	*

714

715

716 **FIGURE LEGENDS**

717

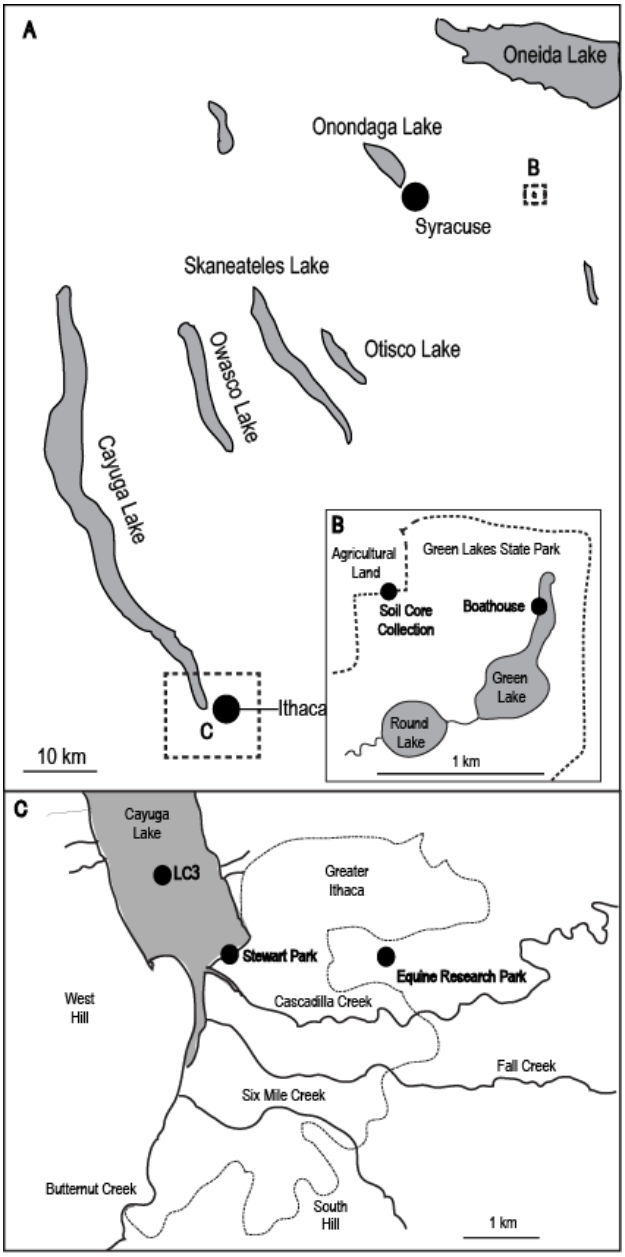
718 **Fig. 1:** Map of sampling locations in the Finger Lakes region of upstate New York, USA (A).
719 Samples were collected in two contrasting lakes: (B) Fayetteville Green Lake with in the Green
720 Lakes State Park at the Boathouse for plankton and sediments, and on Agricultural Land nearby
721 for soil; and (C) Cayuga Lake in Stewart Park and LC3 for plankton, and at the Equine Research
722 Park for soil.

723 **Fig. 2:** Phylogenetic annotation of metaviromes of sequence reads (A,B) and contiguous
724 sequences (C). The read annotations were performed using the VIROME pipeline
725 (http://virome.diagcomputing.org/) , while the contig annotation was performed based on
726 BLASTx of contig ORFs against viral, bacterial + archaeal, and eukaryotic databases at
727 CAMERA (http://camera.calit2.net) using an e-value cutoff of 10^{-3}. A– affiliation of reads by
728 kingdom. B – Viral read annotation by family. C – Viral contig affiliation by family.

729 **Fig. 3:** Catchment precipitation (A) and viral genotype abundance in Fayetteville Green Lake
730 (FGL; panel B) and Cayuga Lake (CL; Panel C) during summer 2011. Error bars = SE between
731 duplicate samples. Autochthonous viral genotypes are indicated by open symbols, while
732 allochthonous viral genotypes are indicated by closed symbols. Precipitation data was obtained
733 for Ithaca and Syracuse weather stations from the National Climatic Data Service. Missing data
734 at each time point indicates that abundances were below detection threshold.

735 **Fig. 4:** Abundance of viral genotypes in surface and deep sediment and within net plankton (> 64
736 µm) in Fayetteville Green Lake. Error Bars = SE of duplicate samples. ND = Not Detected.

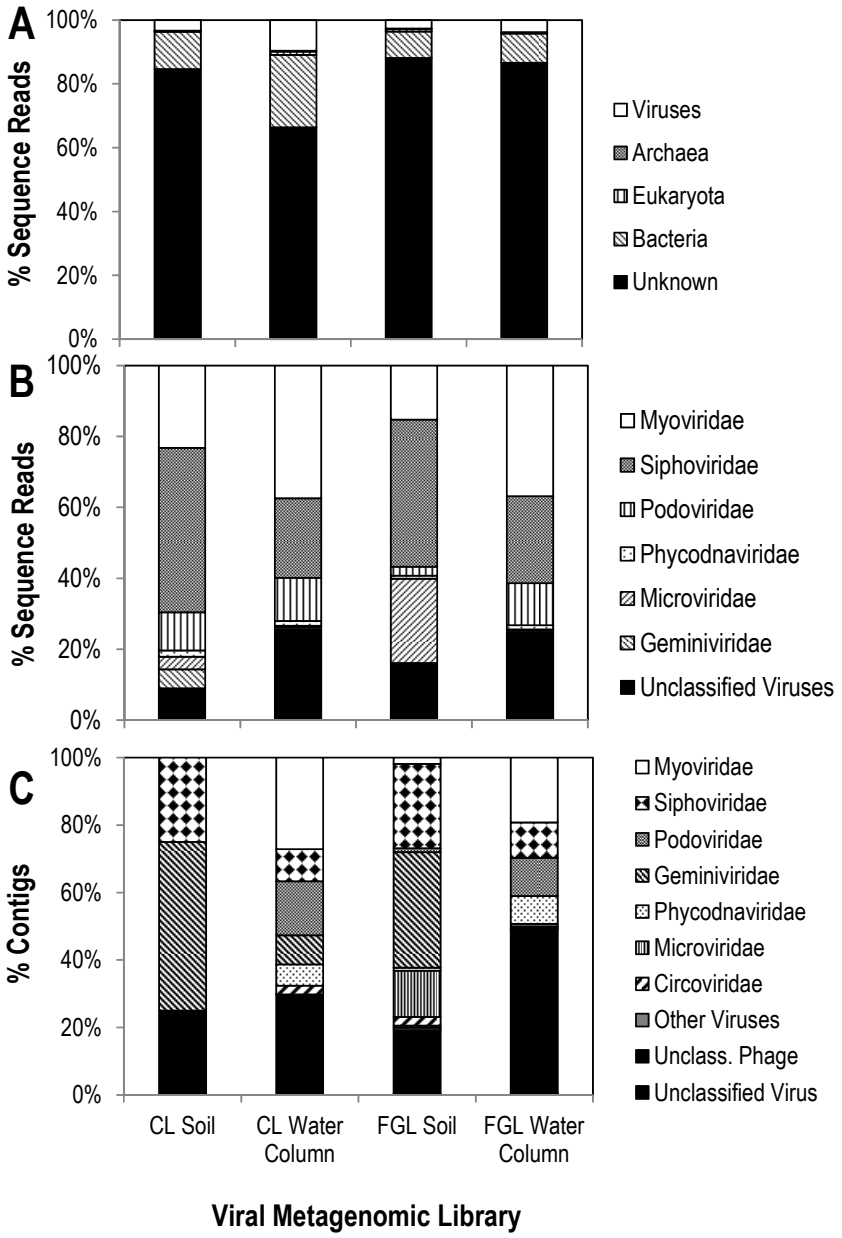
737



738

739 **Figure 1 Hewson et al.**

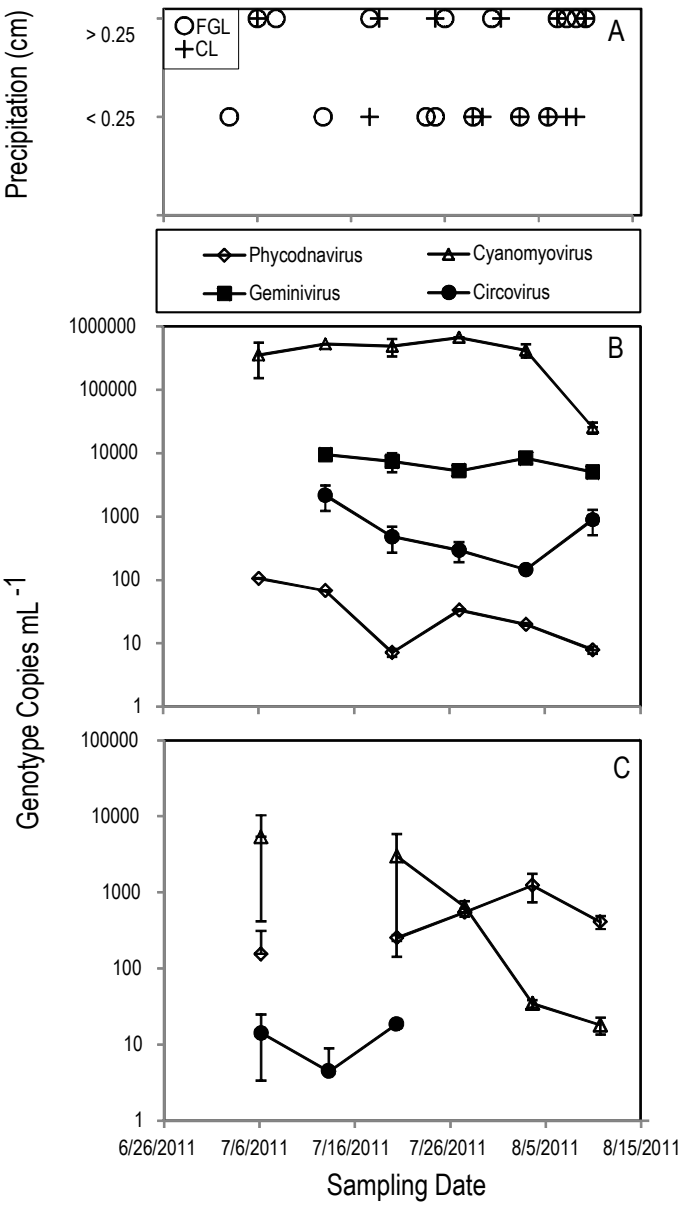
740



741

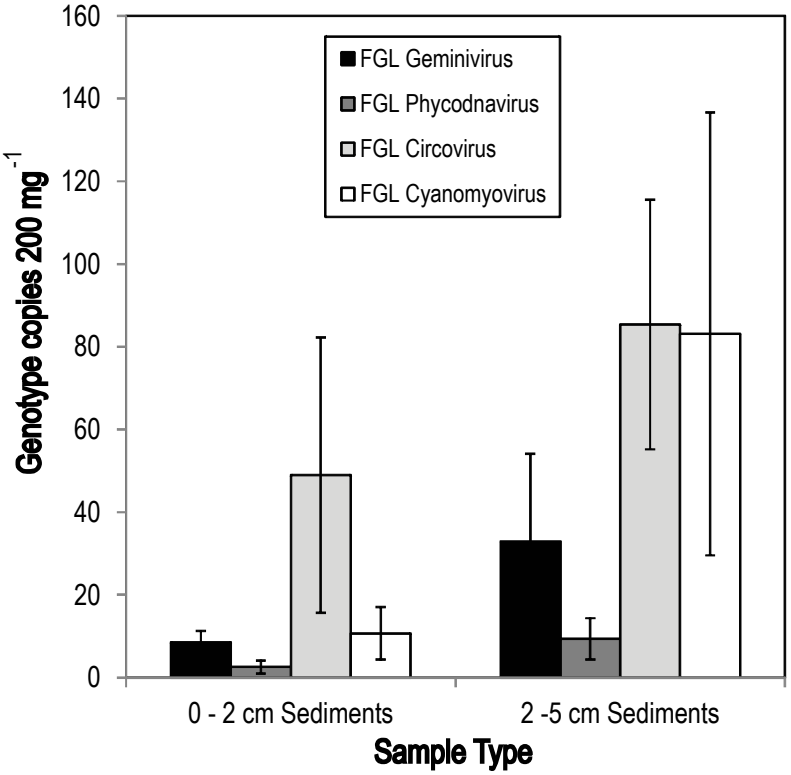
742 **Figure 2 Hewson et al.**

743



744

745 **Figure 3 Hewson et al.**



746

747 **Figure 4 Hewson et al.**

748

749