

Metatranscriptomic analysis reveals dissimilarity in viral community activity between an ice-free and ice-covered winter in Lake Erie

Elizabeth R. Denison,¹ Brittany N. Zepernick,¹ R. Michael L. McKay,² Steven W. Wilhelm¹

AUTHOR AFFILIATIONS See affiliation list on p. 13.

ABSTRACT Winter is a relatively under-studied season in freshwater ecology. The paucity of wintertime surveys has led to a lack of knowledge regarding microbial community activity during the winter in Lake Erie, a North American Great Lake. Viruses shape microbial communities and regulate biogeochemical cycles by acting as top-down controls, yet very few efforts have been made to examine active virus populations during the winter in Lake Erie. Furthermore, climate change-driven declines in seasonal ice cover have been shown to influence microbial community structure, but no studies have compared viral community activity between different ice cover conditions. We surveyed surface water metatranscriptomes for viral hallmark genes as a proxy for active virus populations and compared activity metrics between ice-covered and ice-free conditions from two sampled winters. Transcriptionally active viral communities were detected in both winters, spanning diverse phylogenetic clades of putative bacteriophage (*Caudo-viricetes*), giant viruses (*Nucleocytoviricota*, or NCLDV), and RNA viruses (*Orthornavirae*). However, viral community activity metrics revealed pronounced differences between the ice-covered and ice-free winters. Viral community composition was distinct between winters and viral hallmark gene richness was reduced in the ice-covered relative to the ice-free conditions. In addition, the observed differences in viral communities correlated with microbial community activity metrics. Overall, these findings contribute to our understanding of the viral populations that are active during the winter in Lake Erie and suggest that viral community activity may be associated with ice cover extent.

IMPORTANCE As seasonal ice cover is projected to become increasingly rare on large temperate lakes, there is a need to understand how microbial communities might respond to changing ice conditions. Although it is widely recognized that viruses impact microbial community structure and function, there is little known regarding wintertime viral activity or the relationship between viral activity and ice cover extent. Our metatranscriptomic analyses indicated that viruses were transcriptionally active in the winter surface waters of Lake Erie. These findings also expanded the known diversity of viral lineages in the Great Lakes. Notably, viral community activity metrics were significantly different between the two sampled winters. The pronounced differences we observed in active viral communities between the ice-covered and ice-free samples merit further research regarding how viral communities will function in future, potentially ice-free, freshwater systems.

KEYWORDS viral diversity, winter limnology, metatranscriptomics, climate change, diatom bloom

Decades of research have demonstrated that viruses are key biological components of aquatic ecosystems (1). Through top-down control, viruses influence microbial

Editor Jeffrey Lawrence Blanchard, University of Massachusetts Amherst, Amherst, Massachusetts, USA

Address correspondence to Steven W. Wilhelm, wilhelm@utk.edu.

The authors declare no conflict of interest.

See the funding table on p. 13.

Received 31 May 2024

Accepted 10 June 2024

Published 28 June 2024

Copyright © 2024 Denison et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

community structure and the movement of organic matter *via* the “viral shunt” and “viral shuttle” (2–4). Viruses also modulate the metabolism of infected cells by hijacking host cell physiology and through virus-encoded metabolic genes (5), which can have far-reaching implications for biogeochemical cycles (6, 7). As many of Earth’s aquatic ecosystems are faced with the unprecedented changes in climate (8, 9), there is growing interest in how environmentally relevant viruses are influenced by climate change, such as how climate modulates viral diversity and production rates (10). The winter period in temperate freshwater lakes is particularly vulnerable to climate change, where one apparent response is the decline in ice cover duration and thickness (9, 11). In fact, the frequency of ice-free winters is projected to increase globally in current climate trajectories (12). Seasonally ice-covered lakes are not biologically dormant during the winter (13), and yet there is relatively little documented about viral activity during this period (14).

The Great Lakes contain about 20% of the global freshwater supply (15) and have substantial contributions to regional socioeconomic services (16). Winter in the Laurentian Great Lakes is undergoing pronounced changes due to a warming climate (17). Since the 1970s, Lake Erie has exhibited significant declines in wintertime ice cover extent and duration (18, 19). Furthermore, ice cover extent drastically alters microbial community structure by shifting the winter ecosystem from one characterized by dense accumulations of filamentous diatoms (20–22) to one dominated by picoplankton (23). Putative phage gene fragments (24) and RNA virus sequences (25) have been detected during the mid-winter in Lake Erie, but there has been no comprehensive characterization of winter viral activity in the Great Lakes to date. In addition, the relationship between viral community activity and ice cover extent has not previously been explored. Viral community composition and activity are shaped by the surrounding microbial community (26, 27). Given that ice cover can influence microbial population structure, it is reasonable to expect a relationship between ice cover extent and viral activity. Viral activity could also be directly impacted by other ice cover-dependent factors as well (e.g., the effect of UV exposure on viral decay rates (28, 29)). Characterizing active viral populations during the winter is an important step toward understanding how virus-mediated processes may be impacted by changing ice conditions.

Targeted gene amplification and shotgun metagenomic approaches have been informative in the study of viruses in the Great Lakes (24, 30, 31), but a limitation of DNA-based approaches is that they do not reveal virus activity (i.e., replicating viruses) or capture RNA viruses. Metatranscriptomics can facilitate both the discovery of RNA viruses and estimate *in situ* activity levels of viruses using transcript abundance as a proxy (32). Here, we surveyed metatranscriptomes from filtered (>0.2 μ m) samples collected from Lake Erie surface waters during two contrasting winters: one winter with high-ice cover (2019, 95% mean maximum ice cover) and a subsequent relatively ice-free winter (2020, 15% mean maximum ice cover) (33). Samples from the spring months following the ice-free winter were also collected and served as an outgroup in this study. Using hallmark genes as proxies for putative viral populations, we identified phylogenetically diverse viral communities that were transcriptionally active in the winter surface waters. Viral community activity metrics (alpha and beta diversity indices) revealed significant differences in community composition and viral richness (i.e., the number of different viral hallmark gene variants) between the ice-covered and ice-free samples. In addition, viral activity metrics were correlated with microbial community activity metrics, namely with microbial richness and the proportion of diatom reads within the winter libraries. This suggested that the differences in viral activity between the winters may in part be linked to differences in the surrounding microbial community. Although the cause(s) of the observed differences in viral community activity remain uncertain, they coincided with the substantial variation in ice cover extent between the winters.

MATERIALS AND METHODS

Sample collection

Opportunistic surface water samples were collected by USCGC *Neah Bay* between February and March of two consecutive years (2019 and 2020). Additional spring samples in 2020 were collected by MV *Orange Apex*. Samples were collected from 0.5 meters below the surface as either whole water ($n = 20$) or net concentrated ($n = 57$, 64 μm net opening). Samples intended for RNA extraction were passed through 0.22 μm nominal pore-size filters, flash-frozen, and stored at -80°C until further processing. Details regarding RNA extraction and sequence processing can be found in an associated *Microbiology Resource Announcement* (34). Meteorological and surface water conditions, ice thickness, nutrient concentration, chlorophyll *a* ($>20 \mu\text{m}$ and $>0.22 \mu\text{m}$ size classes), phytoplankton taxonomy, and phytoplankton cell abundance were recorded. Chlorophyll *a* measurements were performed using the method described by Twiss et al. (21). Nutrient measurements were performed by the National Center for Water Quality Research (Heidelberg University, Tiffin, OH, USA). Phytoplankton identification and enumeration were performed as described in Zepernick et al. (35). Sample metadata can be accessed through the Biological and Chemical Oceanography Data Management Office (BCO-DMO) under data set number 809945.

Metatranscriptome assembly and gene prediction

This study used the metatranscriptome co-assembly and gene predictions generated by Zepernick et al. (35). Briefly, in Zepernick et al., quality-filtered reads were co-assembled using MEGAHIT v1.2.9 (<https://github.com/voutcn/megahit>) and the resulting co-assembly was assessed using QUAST v5.0.2 (Table S1) (<https://github.com/ablab/quast>). Gene sequences were predicted from the co-assembly contigs using MetaGeneMark v3.38 (36). Here, to detect and quantify the relative transcript abundance for genes within a sample, we mapped trimmed reads to gene sequences with $\geq 90\%$ identity and $\geq 90\%$ read length fraction thresholds using CoverM v0.6.1 (<https://github.com/wwood/CoverM>). To remove poorly covered genes, at least 50% of the gene must be covered to be considered detected in a metatranscriptome library (37). Read counts for genes with less than 50% coverage were reset to zero on a per-library basis. Read counts were normalized using the transcripts per million (TPM) method (38).

Viral hallmark gene discovery

Conserved viral hallmark genes were identified as proxies for viral populations to estimate viral diversity and activity, similar to as performed previously (32, 39). All hallmark genes were identified through protein sequence similarity searches. Predicted protein sequences from MetaGeneMark were aligned to databases of viral hallmark genes (detailed below) using DIAMOND BLASTp (<https://github.com/bbuchfink/diamond>) (E-value threshold of $1e^{-5}$). Protein sequences that aligned to translated hallmark genes were then aligned to the RefSeq v213 database and sequences with a top hit to a viral protein were retained as putatively viral (32). As a second quality control measure, only viral sequences with an appropriate protein domain identified by a Pfam domain search (database v32) were retained for analysis (i.e., a putative Gp23 sequence retrieved from the BLASTp that had no capsid domain was not included). RdRp-like protein sequences were further filtered using Palmscan (-rdp option) and only high confidence RdRp sequences retained (<https://github.com/rcedar/palmscan>). As an additional measure to remove non-viral sequences, the metatranscriptome contig from which the hallmark gene originated was piped through VirSorter2 v2.2.3 (<https://github.com/jiarong/VirSorter2>) and CheckV v1.0.1 (<https://bitbucket.org/berkeleylab/CheckV>). While the majority of contigs were too short to determine whether they were non-viral, no hallmark gene contigs had any cellular genes called by VirSorter2 or CheckV.

Hallmark gene databases

The terminase large subunit (TerL) was used as a general phage marker (i.e., a marker that is present in all *Caudoviricetes*) (Table S2) (40). Relatively few TerL were detected, most of which resembled T4-like myophage based on sequence homology. Therefore, the T4-like major capsid protein (Gp23), DNA polymerase B (Gp43), and tail sheath domains were used to detect myophage. The Gp23 had the greatest richness compared to the other phage markers and was therefore chosen as the representative hallmark gene for myophage for further analysis (Table S2). Viral integrase, excisionase, CI repressor, and Cro repressor were also screened for as these markers have been used to detect lysogenic phage in metagenomic data (41, 42). All *Caudoviricetes* marker sequences were retrieved from Interpro, where only viral sequences were downloaded.

The major capsid protein (MCP) was used as a proxy for active *Nucleocytoviricota*, or NCLDV, populations. We chose the MCP marker because (i) it is conserved in many NCLDV genomes (43) (all known groups except for *Pandoraviridae* and *Pithoviridae*) and (ii) it is most likely to be detectable since the MCP can be highly expressed relative to other NCLDV core genes (44). We also searched for DNA polymerase B (PolB) as an alternative NCLDV marker, but only nine PolB sequences were detected in the metatranscriptomes *via* read mapping (Table S2). Therefore, we only discuss the MCP as a proxy for NCLDV with the caveat that the MCP may not be the most robust phylogenetic marker for NCLDV (45). NCLDV marker sequences were retrieved from the NCVOG database (<https://ftp.ncbi.nih.gov/pub/wolf/COGs/NCVOG/>).

RNA-dependent RNA polymerase (RdRp) was used as the hallmark gene for RNA viruses (*Orthornavirae*) (46), and the RdRp-Scan database was used for RdRp searches (<https://github.com/JustineCharon/RdRp-scan>). Replicase sequences compiled by Kazlauskas et al. (47) and Moniruzzaman et al. (32) were used to screen for circular replicase-encoding single-stranded (CRESS) DNA viruses (*Cressdnaviricota*).

Phylogenetic analysis

Hallmark protein sequences were aligned to reference alignments using Clustal Omega v1.2.4 (76). Positions with >90% gaps were trimmed from the alignments by trimAl v1.4 (<https://github.com/inab/trimal>). Trees were constructed using FastTree v2.1.11 (-lg option) (<http://www.microbesonline.org/fasttree>) or IQ-TREE v2.2.0.3 (-m TEST option) (<http://www.iqtree.org>) and were visualized in iTOL (80). NCLDV reference genomes were retrieved from Gilbert et al. (81) and the MCP sequences were predicted using ViralRecall v2 (<https://github.com/faylward/viralrecall>). RdRp-Scan alignments were used for the RdRp tree references. References for all other trees were manually curated.

Microbial community characterization

Taxonomic annotation and functional annotation of predicted cellular genes were retrieved from Zepernick et al. (35). Briefly, taxonomy was estimated using EUKulele v1.0.6 (<https://github.com/AlexanderLabWHOI/EUKulele>), where protein sequences were aligned to the PhyloDB database v1.076 by DIAMOND BLASTp. Cellular genes were functionally annotated by eggNOG-mapper v2.1.6 using the eggNOG 5.0 orthology database with a DIAMOND BLASTp cutoff of $1e^{-10}$ (<https://github.com/eggnogdb/egg-nog-mapper>).

Statistical analyses

Community diversity metrics were calculated using the R package vegan (<https://github.com/vegandevs/vegan>) and in PRIMER v7 (48, 49). Normalized transcript abundance (TPM) tables were used as the input for all diversity analyses. For beta diversity statistics, Bray-Curtis dissimilarity matrices were generated from the standardized TPM table (Wisconsin double standardization). Similarity percentage (SIMPER) analyses were used to quantify the contribution of individual marker genes to average dissimilarity between sample groups.

Spearman's correlations were performed using R (`cor.test`) (48). The virus-host hallmark gene interaction network was generated using FlashWeave ($\alpha < 0.01$) (<https://github.com/meringlab/FlashWeave.jl>), where genes annotated as DNA-directed RNA polymerase subunits Rpb1 (KEGG orthology K03006) and RpoB (K03043) were used as cellular hallmark genes (32, 50). A TPM table of all viral and cellular hallmark genes was used as input for the correlation analysis.

RESULTS

Preliminary comparison of viral community activity between whole-water and net-concentrated samples

Beta diversity analysis based on the relative transcript abundance of viral hallmark genes showed that the composition of active viral communities was significantly different between the whole-water (no filtration) and net-concentrated (64 μm opening) samples (ANOSIM $R = 0.91$, $P = 0.001$) (Fig. S1A). The total number of putative viral hallmark gene variants detected within a library was significantly lower in the net-concentrated relative to the whole-water samples (ANOVA, $P < 0.001$) (Fig. S1B). Examination of the relative abundance of viral hallmark genes within the libraries revealed that phage (Gp23), NCLDV (MCP), and RNA virus (RdRp) transcripts were detected in each whole-water sample, but that phage and NCLDV transcripts were relatively absent in the net-concentrated metatranscriptomes (Fig. S1C). Since we observed more hallmark gene variants in the unfiltered samples, both in terms of the number of hallmark genes detected and the different virus types detected (phage, NCLDV, and RNA viruses), we focused on the whole-water samples and the net-concentrated metatranscriptomes are not discussed here further.

Physiochemical parameters and phytoplankton abundance estimates

The whole-water samples spanned temporal, climatic, and spatial gradients. Whole-water samples were collected during February–March of 2019 ($n = 4$), February–March of 2020 ($n = 10$), and May–June of 2020 ($n = 6$) across 12 sites (Fig. S2; Table S3). The winter of 2019 was considered a high-ice year (95% mean maximum ice cover) while the winter of 2020 was largely ice-free (15% mean maximum ice cover) (33). Ice cover ranged from 90% to 100% at the 2019 winter (i.e., February/March) sample sites and was 0% at all 2020 winter sample sites (Table S3). Ice thickness ranged from approximately 2.5 to 15 cm at the ice-covered sites. The concentration of dissolved ammonia, chlorine, silicate, and soluble reactive phosphorus and the silica:nitrate ratio were not significantly different between 2019 (ice-covered) and 2020 (ice-free) winters (two-tailed unpaired t -test, $P > 0.1$) (Fig. S3; Table S4). Dissolved nitrate was on average significantly higher in the ice-covered relative to the ice-free winter ($P = 0.03$) (Fig. S3). Chlorophyll *a* (both $>20 \mu\text{m}$ and $>0.22 \mu\text{m}$ size classes) and diatom cell abundance were on average lower in the ice-free samples, but not significantly so ($P > 0.1$) (Fig. S4; Table S5). An in-depth analysis on phytoplankton dynamics is outside of the scope of this study, but details regarding phytoplankton taxonomy and distribution in these samples can be found in the associated study by Zepernick et al. (35).

Signatures of active viral communities within the winter surface water metatranscriptomes

Viral hallmark genes detected in the metatranscriptomes were interpreted to represent transcriptionally active or cell-associated viral populations. This was presumed because (i) DNA viruses must be replicating intracellularly to synthesize mRNA and (ii) samples were filtered ($>0.2 \mu\text{m}$) prior to nucleic acid extraction at a size cutoff that would enrich for cell-associated RNA viruses. Hereafter we refer to the detected virus populations as “active” with the consideration that this approach may detect cell-associated RNA virus genomes that are not replicating.

Putative phage (Gp23), NCLDV (MCP), and RNA virus (RdRp) hallmark genes were detected in the winter surface water samples (Fig. 1A). Overall, active winter viral communities displayed high hallmark gene richness (i.e., the number of hallmark gene variants detected) and were phylogenetically diverse (Fig. 1A through D). The majority of Gp23 detected in the winter metatranscriptomes *via* read mapping (439 predicted gene sequences) were most closely related to uncultured myophage environmental sequences with uncertain hosts as opposed to the established cultured clades, although some Gp23 were related to isolated cyanophage (two sequences) and *Pelagibacter* phage (six sequences) (Fig. 1B; Table S6). The NCLDV MCP detected in the winter libraries (427 sequences) represented five of the six established NCLDV orders (Fig. 1C) (71), most of which were assigned to the *Imitervirales* (395 sequences) (Table S6). The relatively higher *Imitervirales* MCP richness may reflect the broad host range of this group (92, 93). The phylogenetic diversity of RdRp present in the winter libraries (332 sequences) was similarly high, spanning the known phylum-level diversity of the *Orthornavirae* (Fig. 1D; Table S6). Furthermore, only a single lysogenic marker (integrase) was considered viral based on its top BLASTp hit to uncultured freshwater *Caudoviricetes* (GenBank accession: [CAB5220699.1](https://www.ncbi.nlm.nih.gov/nuccore/CAB5220699.1)) and was only present in one library in the ice-free conditions (MT6, 02/14/2020).

The viral hallmark gene transcript pool was consistently dominated by Gp23 and RdRp transcripts across both winters, although MCP representation increased in the spring, specifically the May 1–22 samples (Fig. S5). Only eight single-stranded DNA virus replicase markers were detected with low representation in the water column libraries (Table S2; Fig. S5) and they were not analyzed further.

Viral community activity metrics differed between the ice-covered and ice-free conditions

Alpha and beta diversity metrics were calculated based on the detected viral hallmark genes to compare viral community activity between the two winters. Community activity was highly dissimilar between the ice-covered and ice-free conditions when comparing the relative abundance (TPM) of viral hallmark genes (73% dissimilar on average, SIMPER analysis). Furthermore, non-metric multidimensional scaling (nMDS) plots revealed that viral community composition clustered by season and that composition was significantly different between the winters based on both hallmark gene relative abundance (ANOSIM $R = 0.79$, $P = 0.003$) and presence-absence matrices (ANOSIM $R = 0.86$, $P = 0.002$) (Fig. 2A). Mean viral community evenness (Pielou's) and diversity (Shannon's H) were not significantly different between seasons (Fig. 2B; Table S7). While community evenness was high in all winter samples (Pielou's $J > 0.6$), evenness was on average lower in the ice-free samples. Mean hallmark richness, however, was significantly lower (Tukey HSD $P < 0.001$) in the ice-covered condition compared to the ice-free (Fig. 2B). In other words, there were significantly fewer viral hallmark genes detected *via* read mapping in the ice-covered winter samples. Viral hallmark richness was still significantly lower when differences in sequencing depth were accounted for (i.e., when hallmark gene counts were normalized to library size) (Tukey HSD $P < 0.001$).

Viral community activity was significantly different between winters when examining virus hallmark types individually. The separate Gp23, MCP, and RdRp communities each clustered by season and were significantly different between the ice-covered and ice-free winters based on relative abundance (Fig. S6). Similarly, the mean hallmark gene richness of all three virus types was significantly lower in the ice-covered winter relative to the ice free (Fig. S7; Table S8). Overall, active viral communities displayed pronounced differences in diversity when compared between the ice-covered and ice-free winter samples and these differences were observed in a broad range of virus types.

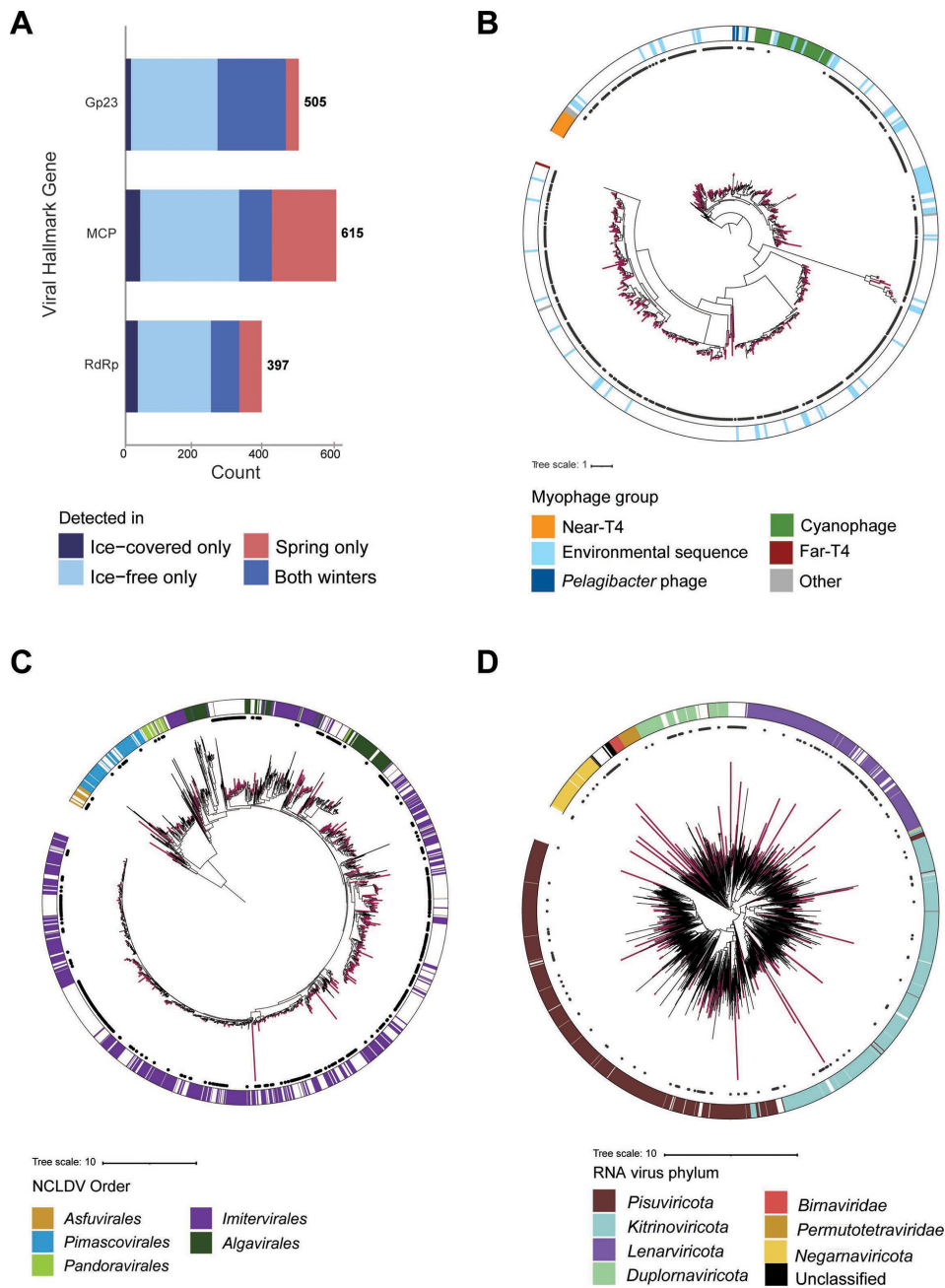


FIG 1 Viral hallmark genes detected in the surface water metatranscriptomes. (A) Total count of genes discovered for Gp23, MCP, and RdRp. Maximum likelihood trees for (B) Gp23, (C) MCP, and (D) RdRp. Branches for Lake Erie sequences are bolded and in color, reference sequence branches are in black. Inner ring circles represent Lake Erie sequences detected in at least one winter library. The outer ring shows the taxonomy of the reference sequence.

Differences in viral community activity were correlated with microbial community activity metrics

Collectively, viral hallmark genes comprised a larger portion of the transcript pool (i.e., total TPM) in the ice-free samples compared to the ice-covered (Fig. 3A). Higher relative viral transcript abundance in the ice-free conditions coincided with a lower abundance of the bloom-forming diatoms (*Bacillariophyta*) in terms of both diatom relative transcript abundance (Fig. 3B) and cell counts (Fig. S4). Conversely, prokaryotic transcripts had higher representation in the ice-free conditions (13%–36% versus 30%–67%,

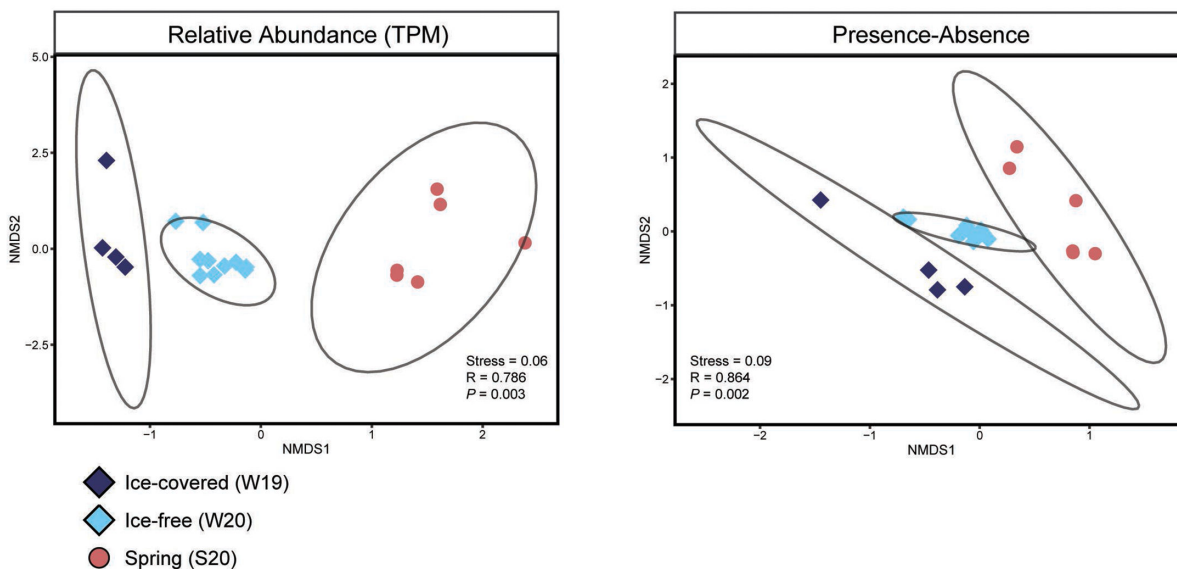
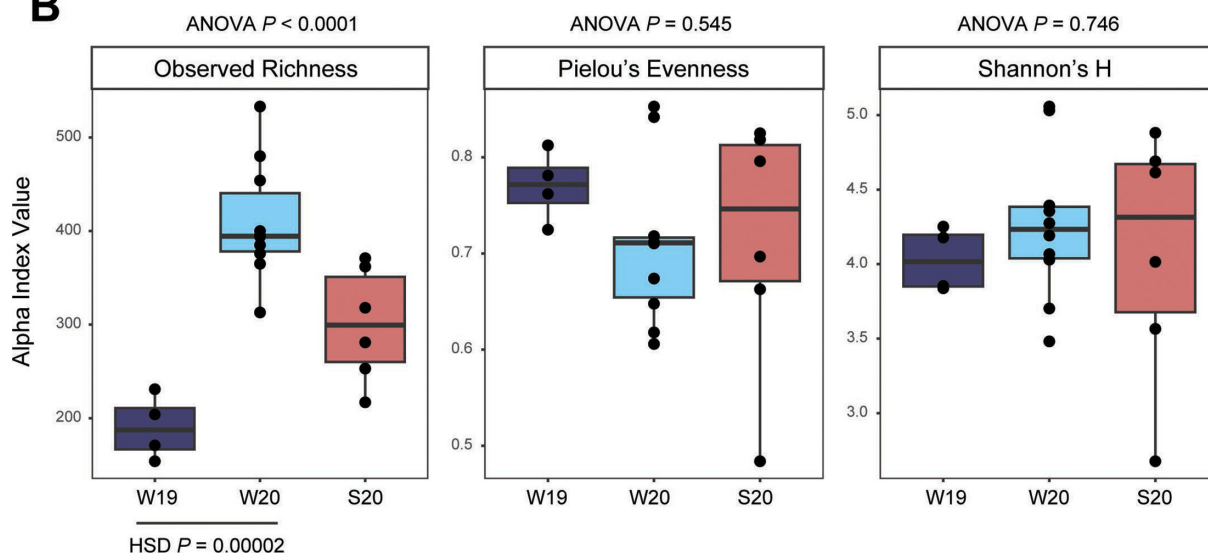
A**B**

FIG 2 Comparison of viral activity metrics between the two winters. Beta diversity non-metric multidimensional scaling (nMDS) plot illustrating the similarity of a sample structure is based on (A) relative transcript abundance and (B) presence-absence transformed abundance of viral markers. ANOSIM results are shown for the pairwise winter comparison. Ellipses represent 95% confidence intervals for the three seasons sampled. (C) Alpha diversity metrics grouped by season. Tukey HSD is shown for the ice-covered winter 2019 (W19) and ice-free winter 2020 (W20) sample comparison when applicable.

respectively) (Fig. 3B). In fact, collective prokaryotic relative abundance was significantly negatively correlated with diatom relative abundance across the winter samples (Spearman's $\rho = -0.83$, $P < 0.001$) (Table S9). Moreover, both prokaryotic (*rpoB*) (Tukey HSD $P < 0.001$) and eukaryotic (*rpb1*) (Tukey HSD $P = 0.02$) hallmark gene richness were lower in the ice-covered winter (Fig. S8).

The shifts in microbial community metrics correlated with the observed differences in viral hallmark gene abundance and richness. Viral hallmark gene relative transcript abundance was significantly negatively correlated with diatom relative transcript abundance ($\rho = -0.82$, $P < 0.001$) (Fig. 3C) and positively so with prokaryotic abundance

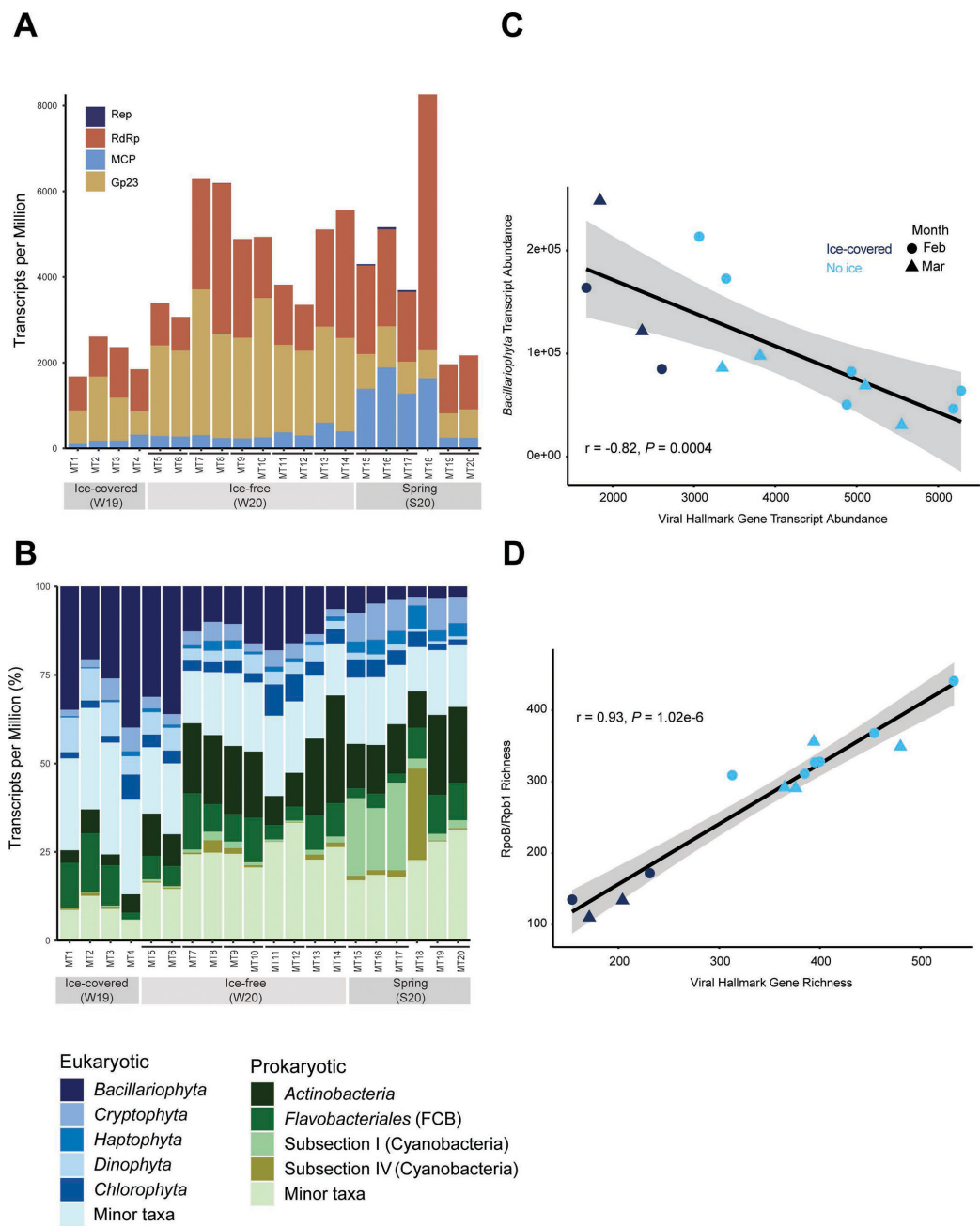


FIG 3 (A) Relative transcript abundance (TPM) of the viral hallmark genes myophage (Gp23), NCLDV (MCP), RNA viruses (RdRp), and CRESS DNA viruses (Rep). (B) Relative transcript abundance of the dominant eukaryotic and prokaryotic taxa. Spearman's correlations between (C) viral hallmark and diatom (*Bacillariophyta*) transcript abundance and (D) viral and cellular hallmark gene richness. Biological replicates are connected by horizontal bars on the bar charts on panels A and B.

(Table S9). Furthermore, viral richness was positively correlated with microbial community richness (RpoB/Rpb1) ($p = 0.93, P = 2.2e-16$) (Fig. 3D). Altogether, lower diatom representation in the ice-free metatranscriptomes was associated with higher microbial community richness and prokaryotic representation, and, in turn with higher viral richness and representation.

Hallmark genes with the largest contributions to the dissimilarity between winter conditions

Although viral richness and representation (collective TPM) were higher in the ice-free conditions, only a few individual viral markers increased in relative abundance and contributed to the dissimilarity between the ice cover conditions. Out of the total 1,525 viral hallmark genes, only 11 had greater than 0.5% contribution to average dissimilarity and amounted to the cumulative 10% average dissimilarity between the ice cover conditions based on the SIMPER analysis (Table S10). While this cut-off is somewhat arbitrary, it captures the viral markers with the largest shifts in relative abundance between the winters, and the remaining markers have low individual contributions. Eight of the eleven markers of interest (four Gp23 and four RdRp) had higher average representation in the ice-free relative to the ice-covered winter where they “spiked” in relative transcript abundance in the ice-free conditions, while the remaining three (one Gp23 and two RdRp) had the opposite trend (Fig. 4). Notably, the three markers that had higher average representation in the ice-covered winter had an overall low relative abundance (Fig. 4), which is in line with the overall low viral TPM in the ice-covered winter samples noted previously.

All five of the top Gp23 contributors were placed within a clade containing an uncultured freshwater Gp23 sequence with an uncertain host range (GenBank accession CAB5226458.1). The RdRp of interest spanned three RNA virus phyla (Fig. 4). Interestingly, three RdRp markers were related to diatom viruses within the *Bacillarnavirus* (order *Picornavirales*) (Fig. S9) and diatom colony-associated viruses within the *Narnaviridae* (phylum *Lenarviricota*) (Fig. S10) and *Totiviridae* (phylum *Duplornaviricota*) (Fig. S11). To further infer virus-host pairs, we assessed correlations in relative transcript abundance between the cellular marker genes (*rpoB/rpb1*) and the virus hallmark genes, but no significant interactions ($\alpha < 0.01$) were found between the diatom Rpb1 and putative diatom virus RdRp markers.

Major capsid protein transcript abundance was consistently low throughout both winters, and subsequently, no single MCP marker had a large contribution to the dissimilarity between winters. However, MCP had higher representation in the spring samples (May 1–22) relative to both winters largely due to the relative increase in two MCPs. Their dominance was also likely responsible for the lower evenness within the MCP communities in the spring samples (Fig. S7). Although host range is uncertain, these MCPs resembled isolated Chloroviruses (order *Algavirales*) that infect green algae and *Mimiviridae* (order *Imitervirales*) isolates that infect amoeba (Fig. 1C). It was noteworthy that the Chlorovirus-like MCP had a significant positive correlation with a Rpb1 annotated as Chlorophyta (*Trebouxiophyceae*) ($\alpha < 0.01$, $\text{cor} > 0.8$).

DISCUSSION

Viral activity during the winter months in the Great Lakes is largely undocumented. Here, viral hallmark genes were detected in metatranscriptomes from two contrasting winters in Lake Erie, indicating that there are viral communities “active” in the surface waters. Our findings expanded the known diversity of viruses that have been detected during the winter in Lake Erie. In one of the few studies to quantify viruses in winter samples collected from the Great Lakes, Matteson et al. enumerated putative cyanophage in mid-winter samples and found them to be a substantial component of the standing stock of viruses (24). However, the qPCR-based detection approach that was used only characterized a narrow group of viruses (cyanomyophage of *Synechococcus* spp.) and could not reveal if those viruses were transcriptionally active. While our metatranscriptomic approach also detected cyanomyophage-like hallmark genes, they represented a small component of the total myophage phylogenetic diversity. Instead, most myophage markers detected here belonged to clades of putative aquatic viruses with unknown hosts. It was previously also hypothesized that phage may “overwinter” as prophage in the Lake Erie water column (24). While we do not dismiss lysogeny as a component of Lake Erie winter microbial communities, we detected little evidence of lysogenic

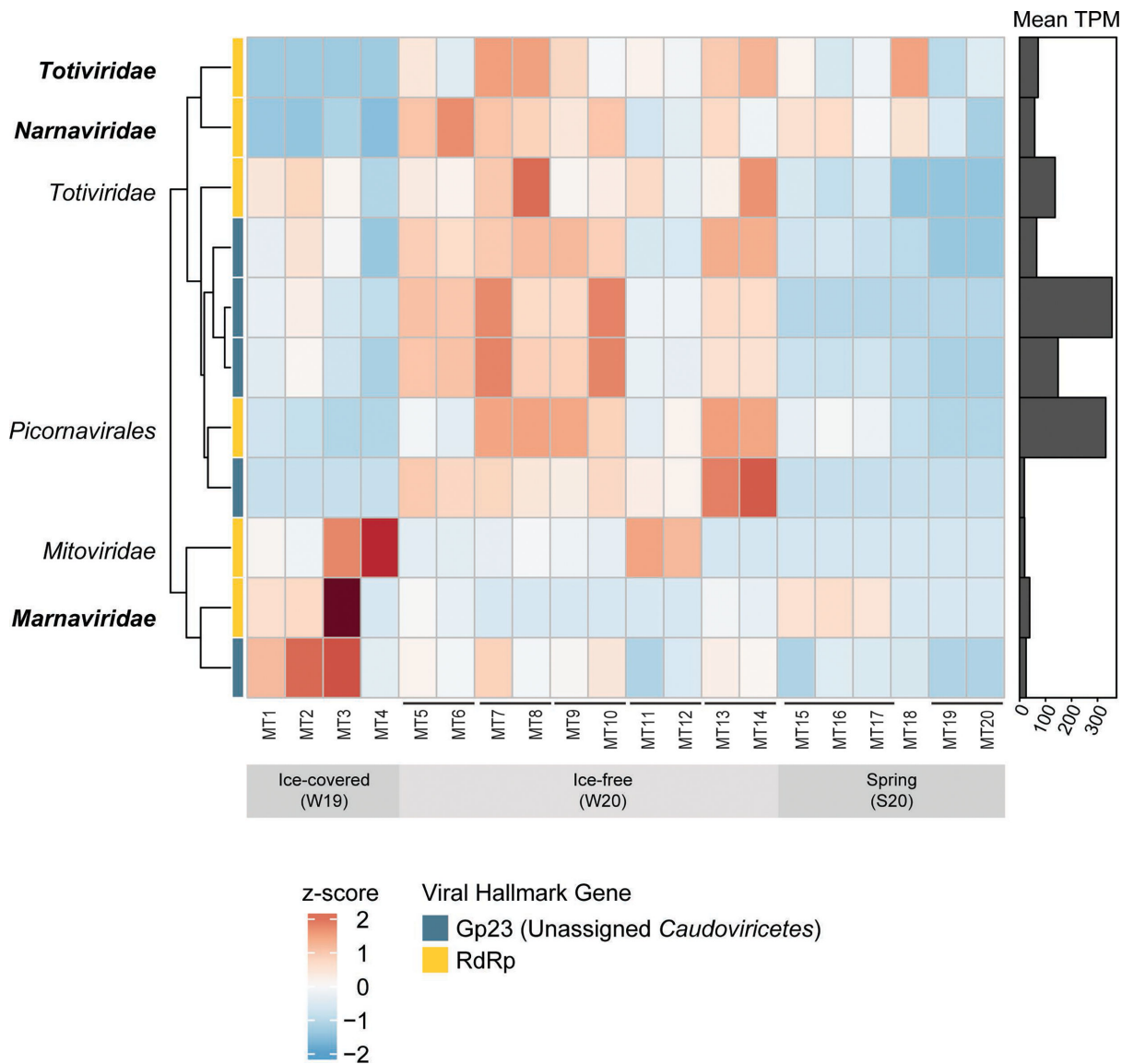


FIG 4 Relative transcript abundance (z-scored TPM) of the 11 top contributors to the dissimilarity between the ice-covered and ice-free samples based on SIMPER analysis. Rows are clustered by abundance pattern. The bar plot shows the mean TPM across all 20 samples. Biological replicates are connected by horizontal bars. Phylogenetic placement for RdRp are indicated by text labels, and diatom-associated RdRp clades are in bold.

lifestyles based on the transcripts assembled here. Metagenomic sequencing may enable the assembly of phage genomes that could help clarify the prevalence of temperate phage in Lake Erie.

Diverse populations of putative eukaryotic viruses within the NCLDV and RNA virus groups were also discovered in the winter samples. Previously, only 11 dsRNA viruses within the *Partitiviridae* (*Pisuviricota*), *Totiviridae* (*Duplornaviricota*), and *Birnaviridae* (*incertae sedis*) families were identified in a single Lake Erie metatranscriptome by Edgar et al. (25), whereas here we detected hundreds of putative RNA viruses belonging to all five currently established phyla. While perhaps a limitation of sequencing depth and sample size, Edgar et al. detected no NCLDV transcripts using a similar hallmark gene approach (25). Our survey, however, revealed putative NCLDV spanning most of the established diversity (five of the six NCLDV orders). Regarding viral discovery, one consideration when using a hallmark gene approach is that viral sequence detection is limited by the diversity of the database (i.e., highly divergent viruses may not be detected by sequence homology searches). In addition, metatranscriptomes represent

“snapshots” of the environment at the time and conditions sampled; therefore, the true diversity of active viruses in the surface waters was likely not captured.

Based on the putative viral sequences we identified, active virus communities were compositionally distinct between winters and viral richness was significantly lower in the ice-covered winter. Our findings showed that the observed differences in active viral communities strongly correlated with the shifts in the microbial community, suggesting a direct relationship between active viral and microbial community composition. Similar to previous observations (23), we saw evidence of a decrease in the magnitude of the winter diatom bloom in the ice-free conditions compared to the ice-covered state determined by both the representation in the metatranscriptomes (diatom relative transcript abundance) and cell count data. Other studies in aquatic environments have attributed lower viral abundance or richness in the winter to lower potential host pool abundance (51, 52), and, in principle, viral diversity and abundance are shaped by the surrounding susceptible host diversity and density (53). The drivers behind the observed differences in viral activity between winters are uncertain, but perhaps the decrease in diatom representation and increase in microbial richness contributed to the greater viral richness in the ice-free winter metatranscriptomes.

The higher levels of viral activity (viral TPM) in the ice-free winter relative to the ice-covered winter were primarily due to increased representation of a few hallmark gene variants. In other words, most viral hallmark genes did not display higher relative transcript abundance in ice-free conditions and were instead consistently low. Interestingly, the subset of viral hallmark genes that did display higher relative transcript abundance contained hallmark genes similar to those of diatom-associated RNA viruses. This included one RdRp related to lytic *Bacillarnavirus* (order *Marnaviridae*) isolates that infect marine diatoms (54), as well as RNA viruses associated with an environmental diatom colony (55). Culture-dependent studies have shown titers of some diatom viruses increased with the proliferation of diatom blooms and have been speculated to control the bloom crash (56), but also coexist during the bloom course (57). None of the putative diatom RNA viruses interacted with diatom (or any) Rpb1 markers based on our abundance-based network analysis; thus, the hosts of these putative RNA viruses remain uncertain.

We note that NCLDV did not have large contributions to the differences in active viral community composition between ice-cover states and that overall NCLDV displayed relatively low transcript abundance during both winters. Low counts (gene copy number) of various algal viruses (*Algavirales*) have been detected during the wintertime in previous studies surveying the Great Lakes (58, 59). Our findings support the idea that NCLDV activity is generally low during winter months in the Great Lakes, but continued winter surveys are needed to resolve this.

In all, we show that diverse viral populations were an active component of the winter surface waters in Lake Erie, emphasizing the need for any future winter surveys to account for viruses when investigating microbial dynamics during the winter. While we documented distinctions in viral activity between the ice-covered and ice-free winter samples, we cannot definitively attribute those differences to ice cover extent. However, we posit that the stark difference in ice cover extent (0% versus 90%–100% cover) is likely a contributing factor to the variation in microbial and viral activity since ice cover is considered a “master variable” due to its influence on biological, biogeochemical, and physical processes in freshwater lakes (17). Aspects such as the reduced active virus richness observed in the ice-covered samples could be characteristic of ice-covered waters in Lake Erie, but ultimately further winter sampling efforts spanning temporal and climatic gradients are needed to resolve trends in viral activity associated with ice cover extent.

ACKNOWLEDGMENTS

We acknowledge the Kenneth & Blaire Mossman Endowment to UTK. The work conducted by the U.S. Department of Energy Joint Genome Institute, a Department

of Energy Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy under contract DE-AC02-05CH11231. We are thankful to the members of the U.S. Coast Guard Cutter Neah Bay and the MV Orange Apex along with Dan Peck and Thijs Frenken for assistance with sample collection and processing. We thank Derek Niles of Orange Force Marine Ltd. for coordinating sample collection during spring 2020 in the face of COVID disruptions that prevented conventional surveillance of the Great Lakes by federal agencies. Finally, we thank George Bullerjahn, Christa Pennachio (DOE JGI), and Justin Chaffin for their collaboration.

This work was supported by the National Institutes of Health, NIEHS grant 1P01ES02328939-01, National Science Foundation grant OCE-1840715 (R.M.L.M. and S.W.W.), NSERC grant RGPIN-2019-03943 (R.M.L.M.), funding from the Simons Foundation (735077), funding from the US Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program Grant to JPG, under Award Number DE-SC0020362, and JGI project 503851 (S.W.W. and R.M.L.M.).

The project was designed by S.W.W. and R.M.L.M. RNA extractions were performed by B.N.Z. B.N.Z. provided the metatranscriptome assembly and gene predictions. Data analysis, statistics, and interpretation of results were carried out by E.R.D. Manuscript drafted by E.R.D. All authors reviewed and edited the final version of the manuscript.

AUTHOR AFFILIATIONS

¹Department of Microbiology, University of Tennessee, Knoxville, Tennessee, USA

²Great Lakes Institute for Environmental Research, University of Windsor, Windsor, Ontario, Canada

AUTHOR ORCIDs

Elizabeth R. Denison  <http://orcid.org/0000-0003-3320-8847>

Brittany N. Zepernick  <http://orcid.org/0000-0003-3121-1225>

R. Michael L. McKay  <http://orcid.org/0000-0003-2723-5371>

Steven W. Wilhelm  <http://orcid.org/0000-0001-6283-8077>

FUNDING

Funder	Grant(s)	Author(s)
HHS NIH National Institute of Environmental Health Sciences (NIEHS)	1P01ES02328939-01	R. Michael L. McKay Steven W. Wilhelm
National Science Foundation (NSF)	OCE-1840715	R. Michael L. McKay Steven W. Wilhelm
Canadian Government Natural Sciences and Engineering Research Council of Canada (NSERC)	RGPIN-2019-03943	R. Michael L. McKay
Simons Foundation (SF)	735077	Steven W. Wilhelm

AUTHOR CONTRIBUTIONS

Elizabeth R. Denison, Data curation, Formal analysis, Funding acquisition, Visualization, Writing – original draft, Writing – review and editing | Brittany N. Zepernick, Data curation, Formal analysis, Writing – review and editing | R. Michael L. McKay, Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Writing – review and editing | Steven W. Wilhelm, Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Writing – review and editing

DATA AVAILABILITY

Raw and quality-filtered reads are publicly available through the JGI Genomes Online Database (GOLD) under GOLD Study ID [Gs0142002](https://gold.jgi.doe.gov/study/Gs0142002).

ADDITIONAL FILES

The following material is available [online](#).

Supplemental Material

Supplemental Materials (mSystems00753-24-S0001.pdf). Supplemental figures and tables.

REFERENCES

- Suttle CA. 2007. Marine viruses - major players in the global ecosystem. *Nat Rev Microbiol* 5:801–812. <https://doi.org/10.1038/nrmicro1750>
- Brussaard CPD. 2004. Viral control of phytoplankton populations - a review. *J Eukaryot Microbiol* 51:125–138. <https://doi.org/10.1111/j.1550-7408.2004.tb00537.x>
- Wilhelm SW, Suttle CA. 1999. Viruses and nutrient cycles in the sea - viruses play critical roles in the structure and function of aquatic food webs. *Bioscience* 49:781–788. <https://doi.org/10.2307/1313569>
- Sullivan MB, Weitz JS, Wilhelm SW. 2017. Viral Ecology comes of age. *Environ Microbiol Rep* 9:33–35. <https://doi.org/10.1111/1758-2229.12504>
- Zimmerman AE, Howard-Varona C, Needham DM, John SG, Worden AZ, Sullivan MB, Waldbauer JR, Coleman ML. 2020. Metabolic and biogeochemical consequences of viral infection in aquatic ecosystems. *Nat Rev Microbiol* 18:21–34. <https://doi.org/10.1038/s41579-019-0270-x>
- Sieradzki ET, Ignacio-Espinoza JC, Needham DM, Fichot EB, Fuhrman JA. 2019. Dynamic marine viral infections and major contribution to photosynthetic processes shown by spatiotemporal picoplankton metatranscriptomes. *Nat Commun* 10:1169. <https://doi.org/10.1038/s41467-019-09106-z>
- Coutinho FH, Cabello-Yeves PJ, Gonzalez-Serrano R, Rosselli R, López-Pérez M, Zemskaia TI, Zakharenko AS, Ivanov VG, Rodriguez-Valera F. 2020. New viral biogeochemical roles revealed through metagenomic analysis of lake baikal. *Microbiome* 8:163. <https://doi.org/10.1186/s40168-020-00936-4>
- Hoegh-Guldberg O, Bruno JF. 2010. The impact of climate change on the world's marine ecosystems. *Science* 328:1523–1528. <https://doi.org/10.1126/science.1189930>
- Woolway RI, Sharma S, Smol JP. 2022. Lakes in hot water: the impacts of a changing climate on aquatic ecosystems. *Bioscience* 72:1050–1061. <https://doi.org/10.1093/biosci/biac052>
- Danovaro R, Corinaldesi C, Dell'anno A, Fuhrman JA, Middelburg JJ, Noble RT, Suttle CA. 2011. Marine viruses and global climate change. *FEMS Microbiol Rev* 35:993–1034. <https://doi.org/10.1111/j.1574-6976.2010.00258.x>
- Li X, Long D, Huang Q, Zhao F. 2022. The state and fate of Lake ice thickness in the northern hemisphere. *Science Bulletin* 67:537–546. <https://doi.org/10.1016/j.scib.2021.10.015>
- Filazzola A, Blagrove K, Imrit MA, Sharma S. 2020. Climate change drives increases in extreme events for lake ice in the northern hemisphere. *Geophys Res Lett* 47. <https://doi.org/10.1029/2020GL089608>
- Hampton SE, Galloway AWE, Powers SM, Ozersky T, Woo KH, Batt RD, Labou SG, O'Reilly CM, Sharma S, Lottig NR, et al. 2017. Ecology under lake ice. *Ecol Lett* 20:98–111. <https://doi.org/10.1111/ele.12699>
- Bertilsson S, Burgin A, Carey CC, Fey SB, Grossart H-P, Grubisic LM, Jones ID, Kirillin G, Lennon JT, Shade A, Smyth RL. 2013. The under-ice microbiome of seasonally frozen lakes. *Limnol Oceanogr* 58:1998–2012. <https://doi.org/10.4319/lo.2013.58.6.1998>
- U.S. EPA. 2023. Great Lakes facts and figures 2023. Available from: <https://www.epa.gov/greatlakes/great-lakes-facts-and-figures>
- Rau E, Riseng C, Vaccaro L, Read JG. 2020. The dynamic Great Lakes economy: employment trends from 2009 to 2018. Available from: <https://www.michiganseagrant.org/topics/resilient-coastal-communities/great-lakes-and-the-economy/2020>
- Ozersky T, Bramburger AJ, Elgin AK, Vanderploeg HA, Wang J, Austin JA, Carrick HJ, Chavarie L, Depew DC, Fisk AT, Hampton SE, Hinchey EK, North RL, Wells MG, Xenopoulos MA, Coleman ML, Duhaime MB, Fujisaki - Manome A, McKay RM, Meadows GA, Rowe MD, Sharma S, Twiss MR, Zastepa A. 2021. The changing face of winter: lessons and questions from the Laurentian Great Lakes. *JGR Biogeosciences* 126:25. <https://doi.org/10.1029/2021JG006247>
- Wang Jia, Kessler J, Bai X, Clites A, Lofgren B, Assuncao A, Bratton J, Chu P, Leshkevich G. 2018. Decadal variability of Great Lakes ice cover in response to AMO and PDO, 1963–2017. *J Clim* 31:7249–7268. <https://doi.org/10.1175/JCLI-D-17-0283.1>
- Wang J, Kessler J, Hang F, Hu H, Clites AH, Chu P. 2017. Great Lakes ice climatology update of winters 2012–2017: seasonal cycle, interannual variability, decadal variability, and trend for the period 1973–2017
- Saxton MA, D'souza NA, Bourbonniere RA, McKay RML, Wilhelm SW. 2012. Seasonal Si:C ratios in Lake Erie diatoms - evidence of an active winter diatom community. *J Gt Lakes Res* 38:206–211. <https://doi.org/10.1016/j.jglr.2012.02.009>
- Twiss MR, McKay RML, Bourbonniere RA, Bullerjahn GS, Carrick HJ, Smith REH, Winter JG, D'souza NA, Furey PC, Lashaway AR, Saxton MA, Wilhelm SW. 2012. Diatoms abound in ice-covered Lake Erie: an investigation of offshore winter limnology in Lake Erie over the period 2007 to 2010. *J Gt Lakes Res* 38:18–30. <https://doi.org/10.1016/j.jglr.2011.12.008>
- Twiss MR, Smith DE, Cafferty EM, Carrick HJ. 2014. Phytoplankton growth dynamics in offshore Lake Erie during mid-winter. *J Gt Lakes Res* 40:449–454. <https://doi.org/10.1016/j.jglr.2014.03.010>
- Beall BFN, Twiss MR, Smith DE, Oyserman BO, Rozmarynowycz MJ, Binding CE, Bourbonniere RA, Bullerjahn GS, Palmer ME, Reavie ED, Waters LMK, Woityra LWC, McKay RML. 2016. Ice cover extent drives phytoplankton and bacterial community structure in a large north-temperate Lake: implications for a warming climate. *Environ Microbiol* 18:1704–1719. <https://doi.org/10.1111/1462-2920.12819>
- Matteson AR, Loar SN, Bourbonniere RA, Wilhelm SW. 2011. Molecular enumeration of an ecologically important cyanophage in a Laurentian Great Lake. *Appl Environ Microbiol* 77:6772–6779. <https://doi.org/10.1128/AEM.05879-11>
- Edgar RE, Morris PF, Rozmarynowycz MJ, D'souza NA, Moniruzzaman M, Bourbonniere RA, Bullerjahn GS, Phuntumart V, Wilhelm SW, McKay RML. 2016. Adaptations to photoautotrophy associated with seasonal ice cover in a large lake revealed by metatranscriptome analysis of a winter diatom bloom. *J Gt Lakes Res* 42:1007–1015. <https://doi.org/10.1016/j.jglr.2016.07.025>
- Bekliz M, Pramateftaki P, Battin TJ, Peter H. 2022. Viral diversity is linked to bacterial community composition in alpine stream biofilms. *ISME Commun* 2:27. <https://doi.org/10.1038/s43705-022-00112-9>
- Coclet C, Sorensen PO, Karaoz U, Wang S, Brodie EL, Eloé-Fadrosch EA, Roux S. 2023. Virus diversity and activity is driven by snowmelt and host dynamics in a high-altitude watershed soil ecosystem. *Microbiome* 11:237. <https://doi.org/10.1186/s40168-023-01666-z>

28. Long AM, Short SM. 2016. Seasonal determinations of algal virus decay rates reveal overwintering in a temperate freshwater pond. *ISME J* 10:1602–1612. <https://doi.org/10.1038/ismej.2015.240>
29. Madan NJ, Marshall WA, Laybourn - Parry J. 2005. Virus and microbial loop dynamics over an annual cycle in three contrasting Antarctic lakes. *Freshw Biol* 50:1291–1300. <https://doi.org/10.1111/j.1365-2427.2005.01399.x>
30. Wilhelm SW, Carberry MJ, Eldridge ML, Poorvin L, Saxton MA, Doblin MA. 2006. Marine and freshwater cyanophages in a Laurentian Great Lake: evidence from infectivity assays and molecular analyses of G20 genes. *Appl Environ Microbiol* 72:4957–4963. <https://doi.org/10.1128/AEM.00349-06>
31. Mohiuddin M, Schellhorn HE. 2015. Spatial and temporal dynamics of virus occurrence in two freshwater lakes captured through metagenomic analysis. *Front Microbiol* 6:960. <https://doi.org/10.3389/fmicb.2015.00960>
32. Moniruzzaman M, Wurch LL, Alexander H, Dyhrman ST, Gobler CJ, Wilhelm SW. 2017. Virus-host relationships of marine single-celled eukaryotes resolved from metatranscriptomics. *Nat Commun* 8:16054. <https://doi.org/10.1038/ncomms16054>
33. NOAA-GLER. 1970 Historical ice cover. Available from: <https://www.glerl.noaa.gov/data/ice/#historical>
34. Zepernick BN, Denison ER, Chaffin JD, Bullerjahn GS, Pennacchio CP, Frenken T, Peck DH, Anderson JT, Niles D, Zastepa A, McKay RML, Wilhelm SW. 2022. Metatranscriptomic sequencing of winter and spring planktonic communities from Lake Erie, a Laurentian Great Lake. *Microbiol Resour Announc* 11:e0035122. <https://doi.org/10.1128/mra.00351-22>
35. Zepernick BN, Chase EE, Denison ER, Gilbert NE, Truchon AR, Frenken T, Cody WR, Martin RM, Chaffin JD, Bullerjahn GS, McKay RML, Wilhelm SW. 2024. Declines in ice cover are accompanied by light limitation responses and community change in freshwater diatoms. *ISME J* 18:wrad015. <https://doi.org/10.1093/ismej/wrad015>
36. Zhu WH, Lomsadze A, Borodovsky M. 2010. Ab Initio gene identification in metagenomic sequences. *Nucleic Acids Res* 38:e132. <https://doi.org/10.1093/nar/gkq275>
37. Franzosa EA, McIver LJ, Rahnavard G, Thompson LR, Schirmer M, Weingart G, Lipsen KS, Knight R, Caporaso JG, Segata N, Huttenhower C. 2018. Species-level functional profiling of metagenomes and metatranscriptomes. *Nat Methods* 15:962–968. <https://doi.org/10.1038/s41592-018-0176-y>
38. Wagner GP, Kin K, Lynch VJ. 2012. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci* 131:281–285. <https://doi.org/10.1007/s12064-012-0162-3>
39. Kaneko H, Blanc-Mathieu R, Endo H, Chaffron S, Delmont TO, Gaia M, Henry N, Hernández-Velázquez R, Nguyen CH, Mamitsuka H, Forterre P, Jaillon O, de Vargas C, Sullivan MB, Suttle CA, Guidi L, Ogata H. 2021. Eukaryotic virus composition can predict the efficiency of carbon export in the global ocean. *iScience* 24:102002. <https://doi.org/10.1016/j.isci.2020.102002>
40. Adriaenssens EM, Cowan DA. 2014. Using signature genes as tools to assess environmental viral ecology and diversity. *Appl Environ Microbiol* 80:4470–4480. <https://doi.org/10.1128/AEM.00878-14>
41. Luo EL, Eppley JM, Romano AE, Mende DR, DeLong EF. 2020. Double-stranded DNA viroplankton dynamics and reproductive strategies in the oligotrophic open ocean water column. *ISME J* 14:1304–1315. <https://doi.org/10.1038/s41396-020-0604-8>
42. Knowles B, Silveira CB, Bailey BA, Barott K, Cantu VA, Cobián-Güemes AG, Coutinho FH, Dinsdale EA, Felts B, Furby KA, et al. 2016. Lytic to temperate switching of viral communities. *Nature* 531:466–470. <https://doi.org/10.1038/nature17193>
43. Koonin EV, Yutin N. 2018. Multiple evolutionary origins of giant viruses. *F1000Res* 7:F1000 Faculty Rev-1840. <https://doi.org/10.12688/f1000research.16248.1>
44. Legendre M, Audic S, Poirot O, Hingamp P, Seltzer V, Byrne D, Lartigue A, Lescot M, Bernadac A, Poulain J, Abergel C, Claverie J-M. 2010. mRNA deep sequencing reveals 75 new genes and a complex transcriptional landscape in mimivirus. *Genome Res* 20:664–674. <https://doi.org/10.1101/gr.102582.109>
45. Aylward FO, Moniruzzaman M, Ha AD, Koonin EV. 2021. A phylogenomic framework for charting the diversity and evolution of giant viruses. *PLoS Biol* 19:e3001430. <https://doi.org/10.1371/journal.pbio.3001430>
46. Koonin EV, Dolja VV, Krupovic M. 2015. Origins and evolution of viruses of eukaryotes: the ultimate modularity. *Virology* 479–480:2–25. <https://doi.org/10.1016/j.virol.2015.02.039>
47. Kazlauskas D, Varsani A, Krupovic M. 2018. Pervasive chimerism in the replication-associated proteins of uncultured single-stranded DNA viruses. *Viruses* 10:187. <https://doi.org/10.3390/v10040187>
48. R Core Team. 2022. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
49. Clarke KR, Gorley RN. 2015. PRIMER V7: user manual/tutorial. PRIMER-E2015
50. Ogier J-C, Pagès S, Galan M, Barret M, Gaudriault S. 2019. rpoB, a promising marker for analyzing the diversity of bacterial communities by amplicon sequencing. *BMC Microbiol* 19:171. <https://doi.org/10.1186/s12866-019-1546-z>
51. Liu YM, Zhang QY, Yuan XP, Li ZQ, Gui JF. 2006. Seasonal variation of viroplankton in a eutrophic shallow Lake. *Hydrobiologia* 560:323–334. <https://doi.org/10.1007/s10750-005-1280-4>
52. Sandaa RA, Larsen A. 2006. Seasonal variations in virus-host populations in norwegian coastal waters: focusing on the cyanophage community infecting marine *Synechococcus* spp. *Appl Environ Microbiol* 72:4610–4618. <https://doi.org/10.1128/AEM.00168-06>
53. Jacquet S, Miki T, Noble R, Peduzzi P, Wilhelm SW. 2010. Viruses in aquatic ecosystems: important advancements of the last 20 years and prospects for the future in the field of microbial oceanography and limnology. *Adv Ocean Limnol* 1:97. <https://doi.org/10.4081/aol.2010.5297>
54. Tomaru Y, Shirai Y, Toyoda K, Nagasaki K. 2011. Isolation and characterization of a single-stranded DNA virus infecting the marine planktonic diatom *Chaetoceros tenuissimus*. *Aquat Microb Ecol* 64:175–184. <https://doi.org/10.3354/ame01517>
55. Urayama S-I, Takaki Y, Nunoura T. 2016. FLDS: a comprehensive dsRNA sequencing method for intracellular RNA virus surveillance. *Microbes Environ* 31:33–40. <https://doi.org/10.1264/jsme2.ME15171>
56. Bettarel Y, Kan J, Wang K, Williamson K, Cooney S, Ribblett S, Chen F, Wommack K, Coats D. 2005. Isolation and preliminary characterisation of a small nuclear inclusion virus infecting the diatom *Chaetoceros cf. gracilis*. *Aquat Microb Ecol* 40:103–114. <https://doi.org/10.3354/ame040103>
57. Tomaru Y, Fujii N, Oda S, Toyoda K, Nagasaki K. 2011. Dynamics of diatom viruses on the western coast of Japan. *Aquat Microb Ecol* 63:223–230. <https://doi.org/10.3354/ame01496>
58. Short SM, Short CM. 2009. Quantitative PCR reveals transient and persistent algal viruses in Lake Ontario, Canada. *Environ Microbiol* 11:2639–2648. <https://doi.org/10.1111/j.1462-2920.2009.01988.x>
59. Short CM, Rusanova O, Short SM. 2011. Quantification of virus genes provides evidence for seed-bank populations of phycodnaviruses in Lake Ontario, Canada. *ISME J* 5:810–821. <https://doi.org/10.1038/ismej.2010.183>