Extreme Viral Partitioning in a Marine-Derived High Arctic Lake

Myriam Labbé,[a,b,c,d] Catherine Girard,[a,b,c,d] Warwick F. Vincent,[b,c,d,e] Alexander I. Culley[a,b,c,d]

aDépartement de biochimie, de microbiologie et de bio-informatique, Université Laval, Quebec City, Quebec, Canada
bCenter for Northern Studies (CEN), Université Laval, Quebec City, Quebec, Canada
cTakuvik Joint International Laboratory, Université Laval, Quebec City, Quebec, Canada
dInstitut de biologie intégrative et des systèmes (IBIS), Université Laval, Quebec City, Quebec, Canada
eDépartement de biologie, Université Laval, Quebec City, Quebec, Canada

Myriam Labbé and Catherine Girard contributed equally to this article. Author order was determined in order of increasing seniority.

ABSTRACT High-latitude, perennially stratified (meromictic) lakes are likely to be especially vulnerable to climate warming because of the importance of ice in maintaining their water column structure and associated distribution of microbial communities. This study aimed to characterize viral abundance, diversity, and distribution in a meromictic lake of marine origin on the far northern coast of Ellesmere Island, in the Canadian High Arctic. We collected triplicate samples for double-stranded DNA (dsDNA) viromics from five depths that encompassed the major features of the lake, as determined by limnological profiling of the water column. Viral abundance and virus-to-prokaryote ratios were highest at greater depths, while bacterial and cyanobacterial counts were greatest in the surface waters. The viral communities from each zone of the lake defined by salinity, temperature, and dissolved oxygen concentrations were markedly distinct, suggesting that there was little exchange of viral types among lake strata. Ten viral assembled genomes were obtained from our libraries, and these also segregated with depth. This well-defined structure of viral communities was consistent with that of potential hosts. Viruses from the monimolimnion, a deep layer of ancient Arctic Ocean seawater, were more diverse and relatively abundant, with few similarities to available viral sequences. The Lake A viral communities also differed from published records from the Arctic Ocean and meromictic Ace Lake in Antarctica. This first characterization of viral diversity from this sentinel environment underscores the microbial richness and complexity of an ecosystem type that is increasingly exposed to major perturbations in the fast-changing Arctic.

IMPORTANCE The Arctic is warming at an accelerating pace, and the rise in temperature has increasing impacts on the Arctic biome. Lakes are integrators of their surroundings and thus excellent sentinels of environmental change. Despite their importance in the regulation of key microbial processes, viruses remain largely uncharacterized in Arctic lacustrine environments. We sampled a highly stratified meromictic lake near the northern limit of the Canadian High Arctic, a region in rapid transition due to climate change. We found that the different layers of the lake harbored viral communities that were strikingly dissimilar and highly divergent from known viruses. Viruses were more abundant in the deepest part of the lake containing ancient Arctic Ocean seawater that was trapped during glacial retreat and were genomically unlike any viruses previously described. This research demonstrates the complexity and novelty of viral communities in an environment that is vulnerable to ongoing perturbation.

KEYWORDS aquatic viral ecology, limnology, polar science, viromics

The Arctic is experiencing a disproportionate and accelerated rate of warming relative to the rest of the world (1), and this is resulting in profound changes throughout the North Polar Region (2). Among the ecosystems particularly sensitive to climate change are polar lakes because of the integral role that their ice cover plays in their structure and functioning (3). This is especially evident in the Canadian High Arctic, where lakes that were perennially covered in ice in the past are now ice free and in that state for increasing periods (4). One type of lacustrine environment that is considered a model ecosystem to study the impact of these changes is represented by meromictic lakes (3). These ecosystems are found at both poles and are characterized by stable, permanently stratified water columns. This persistent layering results in sharp biogeochemical gradients, which in turn create a wide range of microbial niches that are particularly sensitive to perturbation (5).

Lake A, on the far northern coast of Ellesmere Island in the Canadian High Arctic, is a meromictic lake that has received much limnological attention since its discovery in the 1960s (Fig. 1). This lake formed approximately 3,000 years ago when Arctic seawater
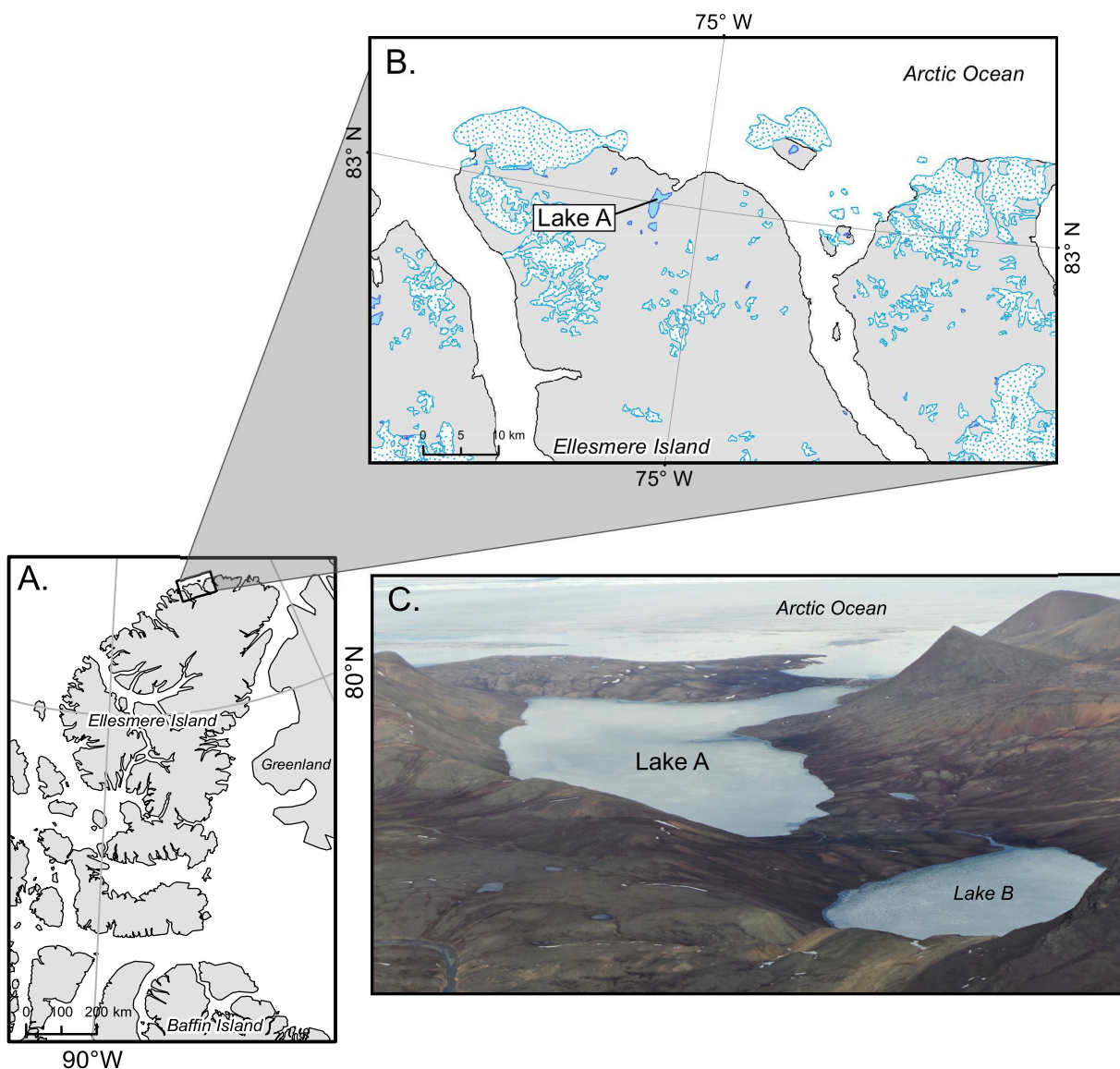

FIG 1 Map of the sampling site showing (A) Ellesmere Island in the Canadian High Arctic (Nunavut), (B) the location of Lake A on the northern coast of the island, and (C) a photograph of Lake A and part of its water catchment (photograph from Alexander Culley; reproduced with permission). Map credit: Michel Paquette, Département de géographie, U. de Montréal.

was trapped by isostatic rebound and then overlaid with freshwater derived from melting snow and ice (6). The lake appears to have been perennially covered with ice throughout much of its history, with a temporary period of ice-out (the disappearance of ice from the surface of a body of water [such as a lake] as a result of thawing) in the 1940s (7). Over the past 2 decades, ice-free incidents have occurred with increasing frequency (3). The 128-m water column of Lake A is highly stratified and is comprised of a high-light, low-conductivity, oxygenated surface layer derived from snowmelt (the mixolimnion), a transition zone (the metalimnion), and a low-light, high-conductivity, anoxic stratum of bottom water that originates from the ancient Arctic Ocean (the monimolimnion) (8).

The pronounced gradients in light, temperature, conductivity, and oxygen in Lake A result in favorable habitats for diverse communities of microbes. In general, the prokaryote and microbial eukaryote communities in the mixolimnion differ entirely from those of the monimolimnion. The most important primary producers in the mixolimnion are chrysophytes and chlorophytes (9) and picocyanobacteria in the genus *Synechococcus* (10, 11). On the basis of amplicon sequencing, Comeau et al. (12) reported that *Proteobacteria*, *Actinobacteria*, and *Cyanobacteria* represent the abundant prokaryotes in surface waters. Pouliot et al. (13) found that the dominant archaeal representatives are affiliated with the *Euryarchaeota* except for the *Nitrosopumilus* (*Thaumarcheota*), a potentially important contributor to nitrification. In contrast, the largest contributors to primary production in the monimolimnion are green sulfur bacteria (GSB) (10). Other abundant bacteria in the monimolimnion include *deltaproteobacteria* and bacteria related to the marine SAR406 clade (12), and the dominant archaea are members of the phylum *Euryarchaeota* (13). Although the diversity of microbial communities of Lake A is strongly shaped by environmental conditions, those communities are also subject to predation by grazers and parasitism by fungi (14) and viruses. With the exception of the characterization of a cyanophage that was isolated from a pooled sample of High Arctic lakes that included Lake A (15), nothing is known about the Lake A viral assemblages and their impact on the microbial communities that they infect.

Viruses are pervasive, prolific, and active components of aquatic environments (16–18). They regulate the abundance and diversity of the organisms that they infect (19), influence the functional potential and evolution of their hosts by facilitating the transfer of genetic material (20), and ultimately may influence the availability of carbon and other critical nutrients (21). Yet viruses in the polar regions have received little attention. Previous studies mostly focused on the Arctic Ocean and Antarctica, and few have used metagenomics to investigate viral diversity (22), with a recent notable exception (23). A study of six lakes and ponds in the island of Spitsbergen in the Svalbard region of Norway used metagenomics to reveal that most of the viral sequences examined had no similarity to those in public databases (24). Among the sequences that were identified, most were related to DNA viruses with small, single-stranded genomes (24). However, these libraries were produced with an amplification method that preferentially amplifies single-stranded DNA (ssDNA), producing a biased portrait of viral diversity (25). This finding is consistent with results from Lake Torneträsk, a subarctic lake in northern Sweden (26). A comparison of the Svalbard lake viromes with viromes from several subarctic and temperate environments showed that Arctic sequences tended to cluster together, suggesting a certain degree of endemism in lacustrine Arctic virus communities (26).

Viruses have been detected in a variety of lakes in the south polar region, including freshwater Limnopolar Lake in the Antarctic Peninsula region (27). The most relevant data for the present study are from Ace Lake, a meromictic lake in the Vestfold Hills of East Antarctica. Like its northern counterpart Lake A, Ace Lake is a well-characterized polar meromictic lake that has been studied over the past 40 years. Ace Lake, like Lake A, is typically covered by thick ice and has a permanently stratified water column with steep physical and chemical gradients (5). The community structure of Ace Lake also resembles that of Lake A, with autotrophic protists, aerobic heterotrophic bacteria, and
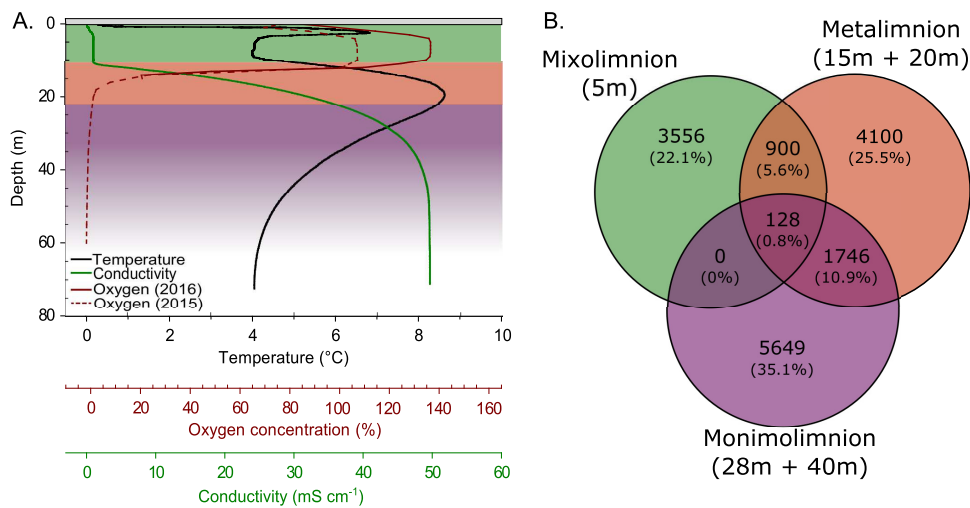
FIG 2 Chemical and biological stratification of the Lake A water column. (A) Profiles of temperature (black), specific conductivity (green), and oxygen concentrations (red). The oxygen profile represents a combination of measurements below 22 m from 2015 (dotted line) and 2016 (unbroken line). Bottom waters (monimolimnion) were anoxic, with an oxygen concentration of 0% at 22 m. The light gray rectangle represents the 92 cm of lake ice cover. (B) Venn diagram showing the proportions of vOTUs shared among the lake strata. The percentage of shared vOTUs for each section is shown in parentheses. In both figures, the three strata of the lake are indicated by color as follows: green, mixolimnion; red, metalimnion; purple, monimolimnion.

abundant picocyanobacteria in its mixolimnion and anaerobic prokaryotes in its monimolimnion, including photosynthetic GSB, sulfate-reducing bacteria, fermentative bacteria, and methanogenic archaea (previously reviewed by Laybourn-Parry and Bell [5]). Viral abundance in the surface waters of Ace Lake remained relatively stable throughout all seasons, ranging from 1×10^7 to 6×10^7 virions per ml (28). Abundances declined markedly in the oxycline and rose to reach peak concentrations in the monimolimnion (29). A comprehensive study of Ace Lake microbial interactions based on metagenomics and metaproteomics found a partitioning of viral taxa between the mixolimnion and monimolimnion, with high relative abundances of protist-infecting phycodnaviruses and their virophages (30) above the oxycline and a high relative abundance of temperate bacteriophages belonging to the order *Caudovirales* below (29). Nevertheless, the points of similarity of Ace Lake's viral ecology to that of Lake A remain unknown. As a sentinel of accelerating climate change in the Arctic, the Lake A microbial ecosystem needs to be fully defined, and establishing its viral diversity and distribution is important.

Our aim in the present study was to characterize the extracellular double-stranded DNA viral community of Lake A. We hypothesized that the distinct physicochemical properties and host communities in the three strata of the lake would result in pronounced differences in viral diversity among these layers. We also aimed to identify the viral assemblages associated with particular host groups and environmental conditions and to evaluate the similarity of Lake A viruses to viral communities elsewhere. We sampled the under-ice water column in midsummer and used a minimally invasive filter-based approach to collect viral particles from replicate samples at multiple depths. These samples were then extracted and sequenced to produce viromic data for analysis.

RESULTS

Physical, biological, and chemical properties of the water column. The surface waters (mixolimnion) of Lake A were fresh and highly oxygenated, with low specific conductivities (Fig. 2A; see also Table 1). There was a sharp transition zone (metalimnion) between 11 and 22 m, with an increase in conductivity and a pronounced maximum in temperature of 8.5°C at 19 m (Fig. 2A; see also Table 1). Oxygen concentrations dropped rapidly with depth across the metalimnion, and data from the

TABLE 1 Physicochemical characteristics of the three strata in Lake A

Stratum	Depth interval (m)	Temp (°C)	Dissolved O$_2$ (%)	Conductivity (mS cm^{-1})
Mixolimnion	0–11	0.4–6.8	84–137	0.1–1.4
Metalimnion	11–22	5.3–8.5	124–0.1[a]	1.4–36
Monimolimnion	22–bottom	4.0–8.5	0.00[a]	36–50

[a]Values based on 2015 data.

previous year showed that the concentration drops to zero at approximately 22 m. The bottom stratum (monimolimnion) exhibited seawater conductivity values (>40 mS cm^{-1}) (Fig. 2A; see also Table 1).

Dissolved organic carbon concentrations (DOC) ranged from 1.8 to 5.0 mg liter^{-1}, with the highest values in the 20-m sample. Both total nitrogen (TN) and total phosphorus (TP) increased with depth. TN reached a maximum concentration of 8.67 mg liter^{-1} in the monimolimnion. TP was below detection limits for depths 15, 20, and 28 m but reached 743 μg liter^{-1} at 40 m (Table 2). Pearson correlation analysis was used to identify closely related environmental variables that could confound further analysis of the data (see Fig. S1 in the supplemental material). The results indicated that TP was closely correlated with TN, and TP was therefore removed from further analyses, as were dissolved inorganic carbon, pH, and fucoxanthin concentration. Chlorophyll-*a* (chl-*a*) was present throughout the water column, while bacteriochlorophyll-*e* (bchl-*e*) and bchl-*e*-like pigments, indicative of GSB, were detected only below 20 m. The highest concentration of the carotenoid zeaxanthin, a proxy for cyanobacteria, was at the surface, with lower concentrations at greater depths (Table 2).

Heterotrophic bacteria were present throughout the water column and were most abundant at the surface, where they reached concentrations of approximately 10^6 cells ml^{-1}. Picocyanobacteria were, in general, an order of magnitude less abundant than other bacterial species, peaking at 15 m at 7.1×10^4 cells ml^{-1}. Autotrophic eukaryotes were present at 5 m and 15 m but were nearly absent in deeper samples. The opposite trend was apparent for the GSB. GSB were not detected in surface samples but were abundant at depth, with a peak level of 1.8×10^5 at 28 m (Table 2). Viral particle concentrations ranged from 9×10^5 to 6×10^6 particles ml^{-1} and were lowest at 15 m and highest at 28 m.

Viral diversity and community structure. Processing of the viral metagenomic data yielded a total of 16,080 viral operational taxonomic units (vOTUs), which are considered here as individual sequences representing a group of highly similar viral contigs. Each stratum of the lake had a distinct set of vOTUs, with over 80% of the total vOTUs recruiting reads from only one layer: 22.1% from the mixolimnion, 25.5% from the metalimnion, and 35.1% from the monimolimnion (Fig. 2B). The divergence between lake strata was further indicated by the relatively low proportion of vOTUs

TABLE 2 Limnological characteristics of sampled depths[a]

Depth (m)	Water chemistry			Pigment concn			Microbial abundance					
	DOC (mg liter^1)	TP (μg liter^{-1})	TN (μg liter^{-1})	Chl-*a* (μg liter^{-1})	Bchl-e (μg liter^{-1})	Zeaxanthin (μg liter^{-1})	Photosynthetic eukaryotes (10^3 ml^{-1})	Green sulfur bacteria (10^3 ml^{-1})	Picocyanobacteria (10^3 ml^{-1})	Heterotrophic bacteria (10^3 ml^{-1})	Viruses (10^4 ml^{-1})	VPR
5	1.8	17	113	0.25 (9)	<LOD	0.15 (1.3)	1.62 (39)	NA	43.7 (8)	813 (6)	435 (23)	5.1
15	3.2	<25	172	0.21 (6)	<LOD	0.13 (14)	0.88 (13.4)	NA	71.1 (29)	261 (13)	96 (24)	2.9
20	5.0	<35	303	0.18 (7)	1.92 (53)	0.08 (7)	1	88	47.6 (11)	189 (3)	113 (48)	4.8
28	4.2	<45	4,401	0.24 (34)	20.5 (27)	0.07 (40)	2	183	18.1 (19)	436 (16)	492 (11)	10.8
40	3.9	743	8,668	0.09 (12)	2.21 (9)	0.05 (11)	0.1	4	7.21 (9)	228 (4)	632 (5)	26.9

[a]Pigment concentration data are from HPLC analysis, and microbial abundance data are from flow cytometry. Each value represents the mean of results from experiments performed in triplicate (percent coefficient of variation [CV] data are indicated in parentheses). The analytical limits of detection (LOD) were as follows: 0.05 mg/liter for DOC; 5 μg/liter for TP; 20 μg/liter for TN; 0.004 μg/liter for Chl-*a*; 0.012 μg/liter for bacteriochlorophyll-*e* (Bchl*e*); and 0.047 μg/liter for zeaxanthin. NA, not applicable.
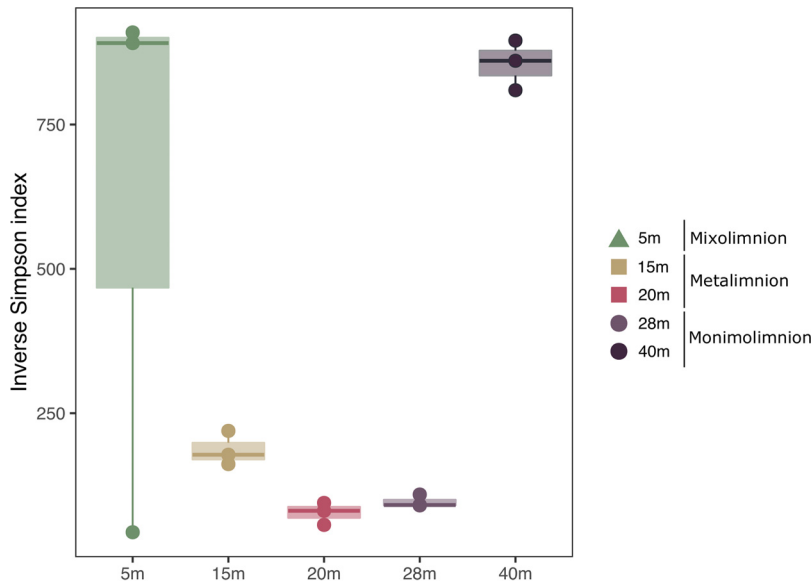
**FIG 3** Within-sample viral diversity by depth assessed from the number of vOTUs in a particular cluster following read mapping. Note that this figure does not reflect the real diversity of viruses in Lake A due to sampling biases and differences in sequencing depths, among other factors. In addition, it is likely that the data indicating viral diversity at 40 m represent underestimates of the levels of diversity relative to the other depths, as the percentage of reads mapped to assembled contigs from this depth was markedly lower (Table S2). No significant differences were observed across depths ($P = 0.122$), likely due to the outlier at 5 m.

shared between layers. While the metalimnion shared a limited number of vOTUs with both the mixolimnion and the monimolimnion, the few vOTUs shared between the mixolimnion and the monimolimnion were also detected in all strata of the lake ($n = 128$, 0.8% of vOTUs). The levels of diversity across depths and layers were not significantly different (Fig. 3, $P > 0.05$).

The viral communities were examined by constrained analysis of principal coordinates (CAP) using Bray-Curtis dissimilarity. The first two axes explained 63% of the variation (CAP1 = 37.9%, CAP2 = 25.1%). Samples from a given depth were highly similar, with nonsignificant dispersal within triplicates ($P > 0.05$, Fig. 4A). However, viral communities from different depths and different lake strata accounted for significant
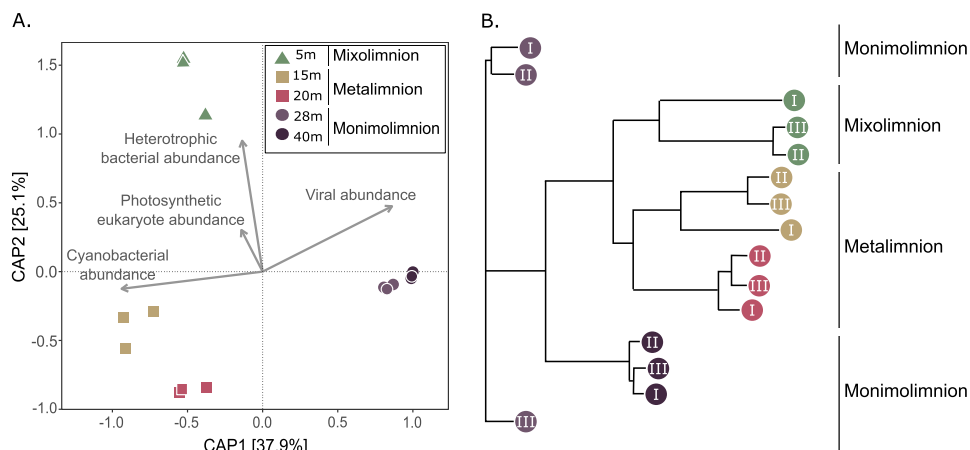
FIG 4 Viral diversity by depth. (A) Constrained analysis of principal coordinates (calculated with Bray-Curtis dissimilarity), including flow cytometry counts of major host groups and viruses. The proximity of a host group (vectors) to a particular viral community (shapes) is an indicator of the strength of the association. (B) Neighbor-joining tree of all depths and replicates color coded by depth.
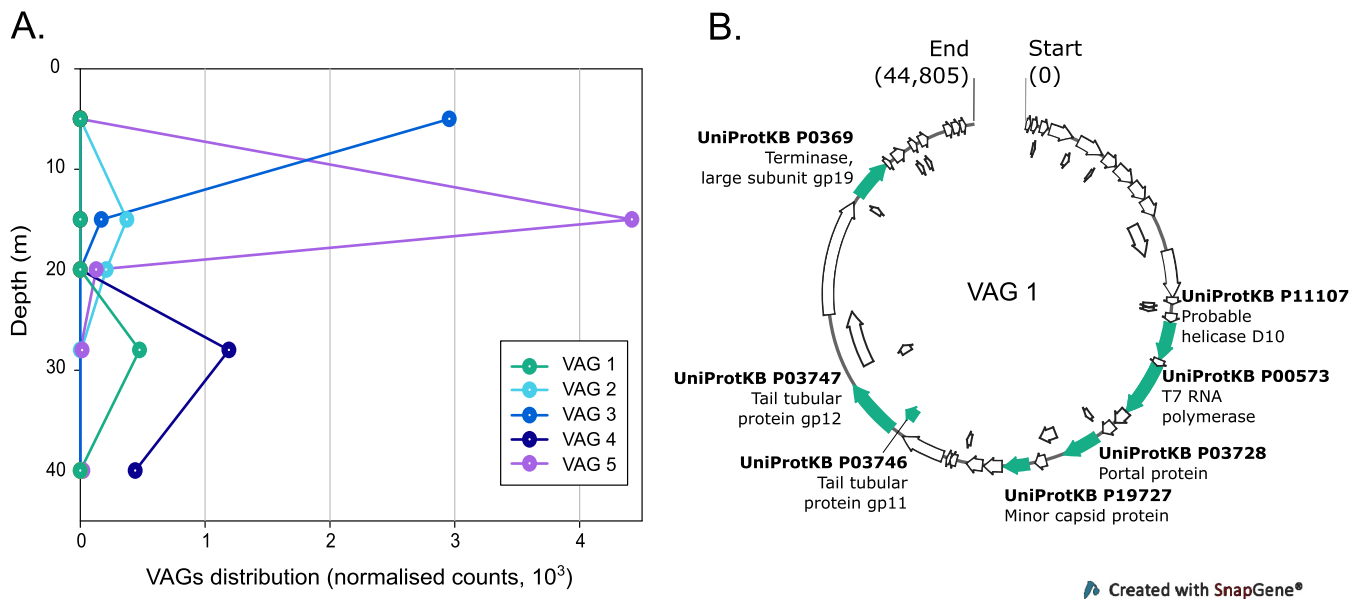
FIG 5 (A) Distribution of 5 ecologically relevant viral assembled genomes (VAGs) in the water column (normalized counts as obtained from Read2RefMapper). (B) Genomic map of VAG 1. Open reading frames (ORFs) corresponding to UniRef90 protein clusters are labeled and shown in turquoise.

amounts of the variations in the data set (87% and 64%, respectively, $P < 0.001$). Among the three strata, the mixolimnion and metalimnion samples grouped more closely together (Fig. 4B). An analysis of CAP space representing vOTU communities and groups of microbes on the basis of flow cytometry counts showed an association between photosynthetic eukaryotes and heterotrophic bacteria and vOTU communities in the mixolimnion, between cyanobacteria and vOTUs from 15 m, and between virus-like particles and vOTUs in the monimolimnion (Fig. 4A).

Weighted gene coexpression network analysis (WGCNA) grouped cooccurring vOTUs into 28 modules that contained from 43 to 3,832 vOTUs per module. Some of these modules were significantly correlated ($P < 0.001$) with several environmental properties, including oxygen concentration (module C), TN concentration (module O), and bchl-e concentration (modules T and V; Fig. S2). Also, each lake stratum was correlated with specific modules (see Table S1 in the supplemental material): modules B and C with the mixolimnion, modules I and J with the metalimnion, and modules O and U with the monimolimnion. We then compared the vOTUs from these modules with viral sequences from the IMG/VR metagenomic database. The percentages of sequences from the modules that were identified ranged from 1% to 37% (see Fig. 6) on the basis of our search criteria (defined below).

Annotation of viral assembled genomes. We identified 504 circular contigs in the database. The circularity of the sequences suggests that they were complete viral genomes, and we refer to these below as VAGs (viral assembled genomes). Among these, we focused on VAGs that were identified as viral in origin with confidence (VirSorter category 1 or 2), were larger than 30 kb, contained hallmark viral genes, and were part of a WGCNA module that was significantly correlated with an important limnological variable such as the concentration of bchl-e (a proxy for GSB) or a particular stratum of the lake, for example. Our analysis of the 10 VAGs that best fit these criteria is summarized in Table S3. Among the 10 VAGs, few to no known genes (less than eight per genome) were identified with Prokka (based on Swiss-Prot, a manually curated database), although those genes that were identified were viral genes from viruses in the order *Caudovirales*. In the case of VAG 1 (Fig. 5B), six genes were found to correspond to *Enterobacteria* phage T7 and one to *Enterobacteria* phage T5, including hallmark structural proteins, a terminase, a helicase, and an RNA polymerase. We used the program ViPTree (31) to further classify the VAGs (Fig. S4). On the basis of
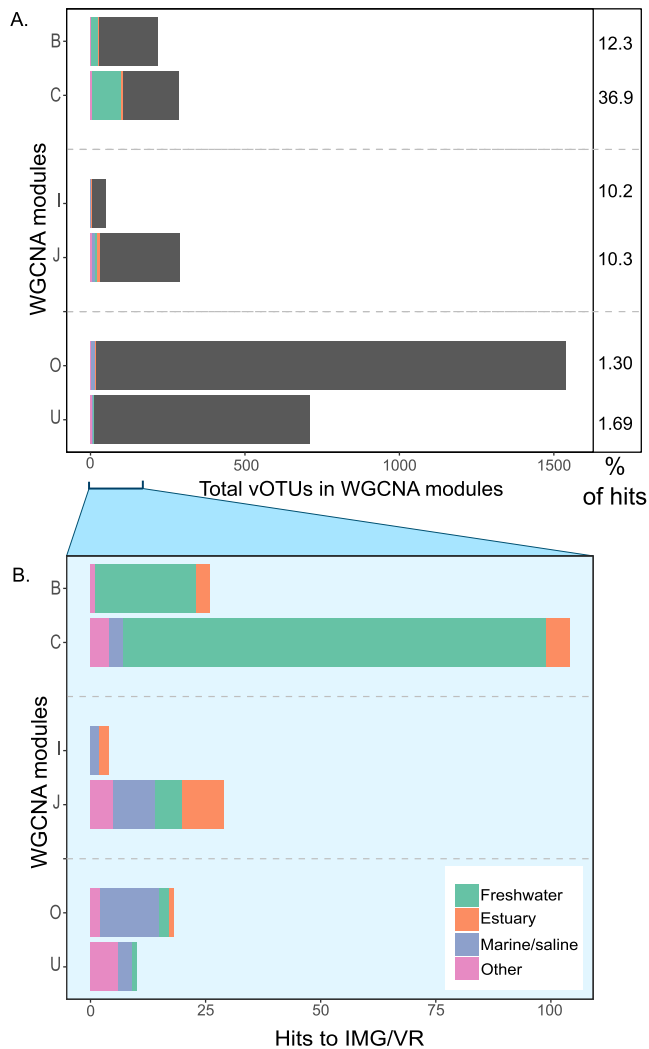
FIG 6 Similarity between WGCNA modules and habitat types. (A) Bar charts showing the habitats of origins of the sequences from the IMG/VR database that matched vOTUs from the WGCNA modules that were highly correlated ($P < 0.001$) to lake strata. The percentage of vOTUs with significant hits to the IMG/VR database is shown on the right for each module. The length of the horizontal column indicates the total number of vOTUs in the module. (B) The expanded part of the graph shows the associated habitat types of only those vOTUs with significant matches.

this analysis, most of the viruses were found to be closely related to viruses in the family *Myoviridae* or the family *Siphoviridae* and to infect hosts in the phylum *Firmicutes*. Overall, however, the VAGs were only distantly related to the nearest taxon. We selected five VAGs based on the depth of annotation, correlation with environmental characteristics (WGCNA analysis, Fig. S2), and abundance in our samples and determined their distribution in the water-column (Fig. 5A; see genomic maps in Fig. S3). The VAGs were generally specific to a lake stratum or depth. For example, VAG 1 was found only at 28 m and VAG 5 was largely restricted to the metalimnion.

Comparative viromic analysis. We compared the most highly correlated modules identified with WGCNA with sequences in the IMG/VR database (32). This analysis revealed that the majority of vOTUs showed little similarity to previously published sequences (Fig. 6). More similarities were found between vOTUs from modules B and C (associated with the mixolimnion) and sequences from freshwater habitats such as lakes, bogs, and rivers, while vOTUs from modules O and U (monimolimnion) were found to be more highly associated with viromes from marine/saline habitats and
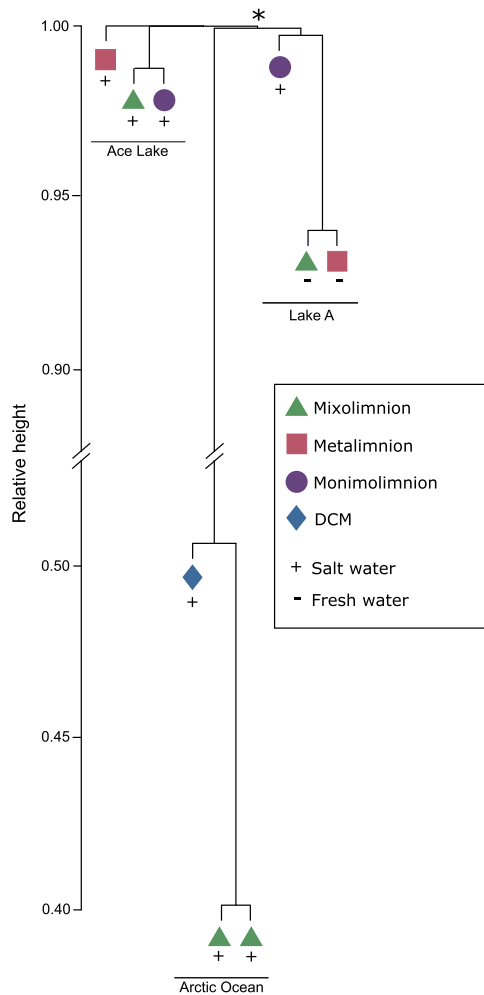
FIG 7 Comparison of Lake A, Ace Lake, and Arctic Ocean viral communities. The dendrogram shows Lake A data for depths 5 m (mixolimnion), 15 m (metalimnion), and 40 m (monimolimnion); Ace Lake data for depths 5 m (mixolimnion), 12.7 m (metalimnion), and 23 m (monimolimnion) (28); and three ocean samples from the Tara Polar expedition (station 196, 5 m; station 194, 5 m and 35 m) (33). The asterisk denotes the node of a clade that includes the Arctic Ocean and Lake A samples.

sediments. The number of hits seen with vOTUs from modules I and J (metalimnion) was equal to the number seen with viromes from marine/saline habitats and estuaries, although module J also had hits to viromes from freshwater and wastewater.

Finally, we compared viral reads from Lake A with viral sequences from two stations in the Arctic Ocean (33) and from Ace Lake, a meromictic lake in Antarctica (28), using Libra (34) (Fig. 7). On the basis of this analysis, the three Lake A strata formed an independent cluster within which the mixolimnion and metalimnion were most similar. Three samples from the Tara Polar project from two stations in the Arctic Ocean (station 196, 5 m; station 194, 5 m and 35 m) were also included (33), and these samples formed a distinct cluster (Fig. 7). The surface samples from the two stations formed a cluster despite the stations being separated by roughly 440 km, with samples from the DCM (deep chlorophyll maximum) zone being less closely related. The Lake A and Arctic Ocean samples formed a cluster with a basal node (identified with an asterisk in Fig. 7). The Ace Lake mixolimnion and monimolimnion grouped closely, while the metalimnion was more divergent.

DISCUSSION

The vertical profiles of physical, biological, and chemical properties that we observed in Lake A (Fig. 2A) were comparable to those observed in studies of the site in

previous years (7, 9, 11, 12, 14, 35), demonstrating that the lake continues to retain a vertical structure that is stable through time. The high concentrations of chl-*a* and zeaxanthin at the surface suggest a high abundance of picocyanobacteria of the genus *Synechococcus* in the mixolimnion, a distribution also observed previously by Antoniades et al. (10). Our cytometry data also confirmed an abundance of GSB at a depth of 28 m as also described previously by Van Hove et al. (11) and Comeau et al. (12), although a coincident peak of chl-*a* and the high counts of photosynthetic eukaryotes suggested that green sulfur bacteria might not be the only autotrophs at this depth. This pattern of abundant picocyanobacteria in the mixolimnion and abundant GSB in the monimolimnion was also observed in Ace Lake, Antarctica (5).

The marked partitioning of viral communities among the three lake strata observed in this study generally fits with previous studies of bacterial (10, 12), archaeal (13), protistan (9), and zooplanktonic (35) communities from Lake A. This segregation was most evident in the viral communities of the mixolimnion and monimolimnion (Fig. 2B) (see also Fig. 3 and 4B), where we found that virtually no virus types were shared (Fig. 2B). Viral community composition is determined by viral production and viral decay, which are determined in turn by abiotic factors (UV-B radiation, concentration of charged particles [36]) and biotic factors (host diversity, host metabolic state, predation, and extracellular enzymatic activity [37]). Consistent with our hypothesis, the distinct abiotic and biotic factors of the three lake strata have resulted in viral communities with striking differences in composition and diversity.

There are few studies that have compared levels of lake viral diversity over depth. A study from Lake Biwa (38) reported that >65% of viral genomes were unique to a particular depth. Similarly to Lake A data, a study of Lake Shunet, a sub-Arctic meromictic lake in Siberia (39), showed that very few viral taxa were present throughout the water column. However, in this study, the highest relative viral diversity was observed in the surface stratum. Unlike Lake A, Lake Shunet is not perennially covered in ice and thus has more exchange with the atmosphere. Seeding by aerosols may account for the relatively high microbial diversity at the surface. As has been found in most freshwater environments, including in this study, a majority (60 to 99%) of viral sequences were found to have no homology with sequences in published databases (40, 41). Although the data from lakes are few, the high relative percentage of unknown viral sequences in the monimolimnion and the extreme partitioning of viral communities between strata (<1% of vOTUs in common) in Lake A appear to be exceptional. It should be taken into consideration, however, that we applied stringent annotation criteria that could result in an overestimation of the number of unclassified sequences.

Studies investigating freshwater viromes have resulted in the observation of sharp differences in viral diversity driven by season (27, 42, 43) or by proximity to the shore or large water inputs or human perturbation (40, 44). Several studies have identified salinity as the primary driver of viral diversity (38, 45), where higher diversity corresponded with higher conductivity (45, 46). Watkins et al. (40) proposed that freshwater viral communities are likely to be more diverse than marine viral communities because freshwater environments are generally more heterogeneous. We did not find this to be the case with Lake A, where the viral assemblage in its unusual saltwater monimolimnion was potentially more diverse than in the less saline strata.

Heterotrophic bacteria and cyanobacteria dominated microbial cell counts at 5 and 15 m (Table 2), suggesting that bacteriophages and cyanophages comprise an important fraction of viral communities at these depths (Fig. 4A). Populations of protists, on the other hand, showed no clear association. This was to be expected since viruses that infect eukaryotes are generally less abundant than prokaryote-infecting viruses, reflecting the much lower concentrations of eukaryotic cells. Moreover, our study was limited to viruses with double-stranded DNA (dsDNA) genomes (there are more RNA viruses that infect eukaryotes than infect bacteria [47]), and our method included a prefiltration step (using a 0.22-μm-pore-size filter) that likely removed viruses of the putative order *Megavirales*, a widespread and abundant group of viruses infecting eukaryotes, with virion diameters that often exceed 220 nm (48).

Our results suggest that sedimentation is not an important driver of viral community composition in Lake A. The higher virus-to-prokaryote ratio (VPR) in the deep samples indicates that virus-host interactions in the mixolimnion are different from those in the monimolimnion. Factors that contribute to a higher VPR in the monimolimnion could include a lower level of exposure to UV radiation and a lower concentration of extracellular enzymes due to a lower abundance of prokaryotes, resulting in lower rates of viral decay and a longer extracellular virus residence time than at the surface (19). It is also possible that there is a higher proportion of lytic viruses (versus temperate viruses) in the monimolimnion than in the mixolimnion (49). A higher relative VPR in deeper waters has also been observed in Antarctic lakes (50) and in the deep ocean (51). In Ace Lake, Lauro et al. (29) observed markedly lower total virus counts and the presence of CRISPR-associated proteins at the depth with the highest abundance of GSB and postulated that GSB may be relatively more resistant to viral infection. However, we did not find the same trend in Lake A, where total virus abundance at the depth of the GSB peak was higher than at all but one other depth.

On the basis of WGCNA, we were able to identify groups of vOTUs (modules) that were associated with particular biogeochemical variables. One module (module C; see Fig. S2 in the supplemental material) was significantly correlated with high oxygen concentrations and may therefore have been enriched with viruses that infect oxygenic phototrophs such as cyanobacteria, although there were no modules that were significantly correlated with zeaxanthin concentrations, a proxy for cyanobacterial abundance. Two modules were significantly associated with the signature pigment of GSB (modules T and V, Fig. S2), which suggests that these modules were enriched with viruses that infect GSB. The WGCNA also identified modules that were significantly correlated with individual lake strata (modules B, C, I, J, O, and U, Fig. S2). We then compared the vOTUs from these modules with viral sequences from the IMG/VR metagenomic database. Consistent with the results of previous analyses of most environmental viromes (52), the vast majority of sequences from the modules were unidentified compared to IMG-VR at an 85% identity threshold. Less-stringent search criteria would likely result in a higher percentage of matches with known sequences, but our results generally reinforce the view that current viral sequence databases represent a small fraction of wild virus diversity. These results also suggest that the vOTUs from the monimolimnion modules, in particular, are highly divergent from those of known viral taxa, reflecting the unusual nature of this habitat.

The scarcity of genes that were identified in our subset of VAGs illustrates the challenges that viral ecologists face with wild viruses from underexplored environments. It should be noted that our annotation was based on matches to a manually curated database. This approach is likely more accurate but may have resulted in a lower percentage of matches than if other databases had been included. Several VAGs with potential ecological relevance (determined by association with environmental parameters) are present in our data set. VAG 1, for example, is found only at 28 m and belongs to a WGCNA module that correlates with the concentration of bchl *e*-like pigments and thus may infect the GSB that were found at high concentrations at this depth (11, 12), although our analysis did not link it with a *Chlorobium* host (Fig. S3). The (albeit limited) annotation of the VAG suggests that it belongs to the group of T-odd phages (Fig. 5B), but further analysis is needed. Several of the VAGs were classified in the *Siphoviridae*, which raises the possibility that they are capable of lysogeny (Fig. S3). However, we were unable to identify any of the genes typically associated with a temperate lifestyle.

We were able to compare the viral community compositions of two meromictic lakes, as well as that of Lake A's monimolimnion (ancient Arctic Ocean seawater), with viral communities from the modern Arctic Ocean (Fig. 7). On the basis of our analysis, there was no overlap between the Arctic and Antarctic lakes. Although Lake A and Ace Lake share several physical and ecological properties, the lack of overlap is unsurprising given the pronounced differences in the histories and locations of the two ecosystems. Little overlap of viral communities from Arctic and Antarctic samples has also been

observed (23, 43) at the species level (22). However, there may be convergence at higher taxonomic levels (22). Another factor to consider is that represented by the differences in methodology between the two studies, including differences in the sequencing platforms (454 pyrosequencing in Ace Lake versus Illumina HiSeq next-generation sequencing in this study) and sequencing depths, which can magnify small-scale variations and exaggerate divergences. This result is consistent with previous comparisons of microbiomes from both poles (23, 52, 53) and also with comparisons of the microbiomes of deep anoxic zones from different lakes (54). This demonstrates that the unique environmental conditions of a particular landscape can be the primary driver of the composition of the microbial communities in these aquatic environments.

Although a basal node is shared (Fig. 7) by the Lake A and Arctic Ocean samples, our analysis does not establish a well-supported link between the Lake A monimolimnion viral community and that in the present-day Arctic Ocean. This suggests that the genomes of viruses in Lake A water that were isolated from the Arctic Ocean several millennia ago are now highly divergent from those of their modern Arctic Ocean counterparts.

The two freshwater samples from the mixolimnion and metalimnion of Lake A formed a tight cluster that was distantly related to samples from brackish and saltwater samples. This result underlines the importance of salinity as a driver of viral diversity in this particular system. The monimolimnion of Lake A appears to contain a high diversity of vOTUs, a majority of which have no equivalent in current databases. This deep-water habitat is also especially vulnerable to climate change. Modeling of warming and ice-cover loss indicated that Lake A will lose its upper water column structure as the lake is increasingly exposed to wind-induced mixing (7). Studies of samples from Lake Romulus, a meromictic lake approximately 350 km south of Lake A, indicated that deep mixing will entrain salt at the surface, where the effects of freeze concentration may lead to even higher conductivity (11). Lake A will persist, but its chemical and physical habitat properties will likely be subjected to major perturbation, with effects on microbial communities, including viral abundance, structure, and diversity. The divergent viral community of the deep stratum may be especially vulnerable to future change.

Conclusions. The three distinct strata of Lake A, including the mixolimnion, metalimnion, and monimolimnion, harbored distinct viral communities and viral assembled genomes that shared few vOTUs, suggesting that viral export was a process of low significance. The highest abundance and highest relative diversity of viruses were in the monimolimnion, a layer of ancient Arctic Ocean seawater at the bottom of the lake. These viruses were mostly novel, indicating the need to better characterize the diversity of wild viruses and their functions, as well as the uniqueness of this habitat. Given that the Arctic is warming at an increasing rate, augmenting the vulnerability of Arctic stratified lacustrine environments such as Lake A, it is crucial to characterize the extant biota, including viruses, in these systems before they are irrevocably perturbed.

MATERIALS AND METHODS

Sample collection and processing. Three holes spaced 10 m apart at a midlake site (lat 82.596679N, long 75.266029W; Fig. 1) were bored through the approximately 92 cm-thick ice cover of Lake A on 16 July 2016. Temperature, conductivity, and oxygen profiling of Lake A was done using a YSI 600QS sonde and an RBR Concerto conductivity, temperature, and depth (CTD) logger. Lake strata were identified based on the physicochemical profiles of the water column (Fig. 2A; see also Table 1). Water samples were then collected at five depths from the three lake strata in each of the three holes with a 7-liter Limnos water sampler as follows: mixolimnion, 5 m; metalimnion, 15 and 20 m; monimolimnion 28 and 40 m. The samples were collected in triplicate, using one replicate per hole. Water was then transferred in the field into Cubicontainers that had been cleaned with 2% (vol/vol) Contrad liquid detergent and 10% (vol/vol) American Chemical Society (ACS)-grade HCl (Sigma-Aldrich) and rinsed with lake water before sampling. Samples were kept cool and in the dark during transportation by helicopter to a field laboratory, where they were processed within 3 h of collection. Subsamples (preserved using a final concentration of 0.5% [vol/vol] glutaraldehyde [grade 1]; Sigma-Aldrich) were taken before filtration for flow cytometry counts and for determination of pigment concentrations (performed using filtration with 25-mm-diameter Whatman GF/F glass-fiber filters [nominal pore size, 0.7 μm] and immediate freezing at

−20°C) and of levels of TP, TN, and DOC (stored at 4°C in the dark) (54). Viral DNA was collected by filtering samples independently through a 0.22-μm-pore-size capsule filter (Millipore Sterivex-GV, PVDF [polyvinylidene difluoride] membrane) to remove most cellular organisms followed by collection of virus-sized particles on 25-mm-diameter 0.02-μm-pore-size Anotop aluminum oxide filters (Whatman). Sample volumes ranged from 375 to 960 ml. Filters were immediately frozen at −20°C in the field and then stored at −80°C at Université Laval until extraction.

Nutrient samples were analyzed as described previously by Laurion et al. (55) at l'Institut National de la Recherche Scientifique, Centre Eau-Terre-Environnement (Québec, Canada), while flow cytometry samples were processed at l'Université du Québec à Rimouski, Institut des sciences de la mer de Rimouski (Rimouski, Canada), as described previously by Brussaard et al. (56). Pigments were analyzed by high-performance liquid chromatography (HPLC) as described previously by Thaler et al. (57). Chlorophyll-a and its allomers were quantified by their absorbance at 450 nm and by fluorescence, zeaxanthin by absorbance at 450 nm, and bacteriochlorophyll-e-like pigments by absorbance at 467 nm. Total nucleic acids were extracted directly from individual Anotop filters using a Complete DNA & RNA purification kit (MasterPure; Epicentre) and the backflushing technique described previously by Mueller et al. (58). Libraries were prepared independently for all 15 samples with 10 ng of sheared DNA (350 to 400 bp) (Covaris M220 Ultrasonicator) using a NEBNext Ultra II library preparation kit (New England Biolabs). Paired-end sequencing (2 × 125 bp) was performed on an Illumina HiSeq 2500 system at the McGill University and Génome Quebec Innovation Centre (Montréal, Canada), yielding 125,403,906 reads (see Table S2 in the supplemental material).

Sequence processing. Initial read quality was assessed with FastQC (v.0.11.2 [59]). Adapter sequences were removed with Trimmommatic (v.0.36 [60]) in "readthrough" mode using a custom adapter file to supplement Illumina's default adapter file. Reads were also trimmed to remove low-quality nucleotides (Phred score, <20) over the first and last 20 bases, and unpaired reads were discarded (Table S2). Error correction of base-calling was performed on paired reads with Karect (v.1.0) with default settings (61). The average length of the processed reads was 124 bp. Processed paired reads obtained independently from each sample were assembled de novo with MetaSPAdes (v.3.10.1) with default settings (62).

Viral sequences in the assembled data set were identified using VirSorter (v.1.0.3) and VirFinder (63, 64). VirSorter confidence categories 1 and 2 were kept for our final viral data set. Contigs that fell into category 3 or were rejected by VirSorter were subsequently analyzed with VirFinder, trained with a model that included eukaryotic viruses downloaded from the VirFinder GitHub site (https://github.com/jessieren/VirFinder). Contigs with a VirFinder confidence score greater or equal to 0.8 were added to the final viral data set. Only viral contigs longer than or equal to 2 kb were kept for further analysis (Table S2).

Viral contigs from the 15 independent samples were clustered together into vOTUs using Cluster-Genomes software (v.1.1.3) on the iVirus.us platform (65). This program groups sequences that share an identity level of 95% over 85% of their length into a sequence cluster; the longest sequence in the cluster is then designated the vOTU. Processed quality reads from each sample were then individually mapped to the vOTUs with Bowtie2 (v.2.3.4.1 [66]) with the "end to end" parameter and SAMtools (v.1.8 [67]). These alignment files were then merged into a single distribution matrix showing the number of reads mapped to each vOTU in each sample. The program Read2Ref Mapper (v.1.1.0 [68]) was used to merge the files and normalize counts by vOTU sequence length (68), ultimately producing a vOTU table.

An unprocessed vOTU table was used for viral diversity and presence-absence analyses, while distance-based (beta diversity, neighbor-joining tree) and correlation analyses were performed on the basis of a vOTU table that was filtered to remove the least abundant 5% of vOTUs and those present in only one sample, resulting in the removal of 1,056 vOTUs. The table was then standardized using the total square root method (Hellinger) [decostand() in vegan{}] (69), and the data were log-transformed in R (R Development Core Team, v. 3.5.1).

Data analysis. All statistical analyses were performed in R, and P values are reported for statistically significant results only (α = 0.05). Limnological parameters and flow cytometry data were compared using the Kruskal-Wallis nonparametric test [kruskal.test() function in stats{}] to compare communities, as the assumptions for parametric tests were not met [tested with shapiro.test() and bartlett.test()]. Dunn's test was used as a post hoc test using multiple-comparison correction (Bonferroni correction) in the FSA{} package (70).

Community-level diversities were compared across depths as well as across lake strata (mixolimnion, metalimnion, and monimolimnion). Metadata collinearity was verified in vegan{} by the use of backward selection and the vif.cca() function. Adjusted R^2 values were calculated in vegan{} with the RsquareAdj() function. Sample-wise community comparisons (beta diversity) were performed using the Bray-Curtis dissimilarity index with phyloseq{}. Ordinations were visualized with principal-coordinate analysis (PCoA) and constrained analysis of principal coordinates (CAP). Differences between centroids of groups were assessed using permutational multivariate analysis of variance (PERMANOVA) with adonis(), and within-group homogeneity of dispersions was verified with betadisper(), both in vegan{}. A sample replicate neighbor-joining tree was created using the Bray-Curtis dissimilarity index with the ape{} package (71) in R. Overlap in vOTUs across lake layers is shown in a Venn diagram computed using Venny 2.0 (72) and was plotted using the VennDiagram{} package in R (73). A correlation matrix of metadata variables was produced with cor() in stats{} and visualized with corrplot{} (74). Boxplot graphs (showing medians and whiskers of the 25% and 75% quartiles) were produced with ggplot2{} (75) on the basis of the number of vOTUs at each depth. As normality and heteroscedasticity were not respected, we used the nonparametric Kruskal-Wallis rank sum test to compare diversities between depths as well as layers (mixolimnion,

metalimnion, and monimolimnion) and used Dunn's test as a *post hoc* test to correct for multiple comparisons.

The WGCNA{} R package (76) was used to determine correlations between vOTUs and environmental parameters, as described previously by Guidi et al. (77). The vOTUs were first clustered into modules based on their cooccurrence within samples using the Pearson correlation. Next, a second Pearson correlation, with *P* values corrected for multiple comparisons (Bonferroni) by p.adjust(), was used to determine how environmental parameters correlated with each module's occurrence (see Fig. S2 in the supplemental material). Correlations with *P* values of ≤0.001 were considered highly significant, and vOTUs belonging to the corresponding modules were used for further analysis. Step-wise positive selection was used to identify covarying variables, which were then removed from the final WGCNA.

The vOTUs from modules significantly correlated with a lake stratum were further investigated using BLASTx (78). The complete IMG/VR protein database (32) was used to explore the similarity between the vOTUs from this study and viral sequences from metagenomes collected throughout the globe. BLASTx results were filtered for top hits with an E value lower than 0.01, an alignment of a minimum of 65 amino acids, and ≥85% identity, and the top hit (*max_target_seqs* option set to 1) was used. The IMG/VR "taxonIDs" metadata were then used to group these results into manually curated habitat types. This threshold was selected on the basis of an iterative process in which we progressively lowered the percent identity value until a threshold (85%) was reached that resulted in at least one match for each WGCNA module. vOTUs that were classified as circular by VirSorter were filtered to select sequences that were >30 kb. Of these, 10 vOTUs were selected based on the following criteria: they were classified in VirSorter category 1 or 2, contained hallmark viral genes, and were part of a WGCNA module that was significantly correlated with an important biological or limnological variable (see Table S3 for a list of characteristics of the VAGs selected). These VAGs were annotated using Prokka (v1.13.7 [79]) with the viral kingdom option. Gene annotations based on Swiss-Prot reference clusters (80) for each VAG can be found in Table S3. Genome maps were made for 5 VAGs with SnapGene software. Classification of the VAGs was determined with ViPTree (v1.9 [31]).

We used Libra (34) to examine the similarities between viral reads of samples collected from the depths of 5, 15, and 40 m of Lake A (replicate III only), those of samples from the Arctic Ocean (33) collected approximately 100 km from Utkiagvik, AK. (station 196, 5 m; ENA accession number ERS1309397), and those of samples from the boundary between the Chukchi Sea and the Arctic Basin (station 194) (5 and 35 m; ENA accession numbers ERS1309308 and ERS1309369, respectively) and of samples from Ace Lake (29), a meromictic lake in the Vestfold Hills, Antarctica (depths of 5, 12.7, and 23 m; MG-Rast accession numbers mgm4443684.3, mgm4443681.3, and mgm4443683.3). For both meromictic lakes, the depths selected were considered representative of the three strata of each lake. Viral reads were identified in the following manner: for each sample, reads were assembled with Megahit (81) (default settings), viral contigs were identified with VirFinder (EPV model; all contigs with a score of ≥0.9 were retained), and raw reads were then mapped to viral contigs using Bowtie2 (66) with the option –al-conc. Viral reads were then compared using Libra (34) with *k*-mers of 20 bp, using the default cosine similarity scoring and logarithmic weighting parameters. The distance matrix produced by Libra with the as.dendrogram() function in the standard R stats{} package was used to create Fig. 7.

Data availability. High-quality processed reads are available in the Sequence Read Archive of the National Center for Biotechnology Information (SRA-NCBI) under BioProject accession number PRJNA545459 or individually as listed in Table S2. The vOTU abundance table and the vOTU sequence files are now available on the ViDEL GitHub site (https://github.com/LabViDEL/Viral-diversity-in-Arctic-meromictic-lake). Data from Ace Lake (28) and the Tara Polar Expedition (33) were downloaded from MG-RAST (www.mg-rast.org [82]) and the European Nucleotide Archive (PRJEB9742), respectively, on 4 April 2019. Environmental metadata for Lake A are archived in Nordicana D (83).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, PDF file, 0.2 MB.
FIG S2, PDF file, 0.2 MB.
FIG S3, PDF file, 0.1 MB.
FIG S4, PDF file, 0.2 MB.
TABLE S1, PDF file, 0.1 MB.
TABLE S2, PDF file, 0.1 MB.
TABLE S3, PDF file, 0.9 MB.

REFERENCES

1. Boy M, Thomson ES, Navarro J-C, Arnalds O, Batchvarova E, Bäck J, Berninger F, Bilde M, Brasseur Z, Dagsson-Waldhauserova P, Castarède D, Dalirian M, de Leeuw G, Dragosics M, Duplissy E-M, Duplissy J, Ekman AM, Fang K, Gallet J-C, Glasius M, Gryning S-E, Grythe H, Hansson H-C, Hansson M, Isaksson E, Iversen T, Jonsdottir I, Kasurinen V, Kirkevåg A, Korhola A, Krejci R, Kristjansson J, Lappalainen HK, Lauri A, Leppäranta M, Lihavainen H, Makkonen R, Massling A, Meinander O, Nilsson DE, Olafsson H, Pettersson JB, Prisle NL, Riipinen I, Roldin P, Ruppel M, Salter M, Sand M, Seland Ø, Seppä H, et al. 2019. Interactions between the atmosphere, cryosphere, and ecosystems at northern high latitudes. Atmos Chem Phys 19:2015–2061. https://doi.org/10.5194/acp-19-2015-2019.

2. Cohen J, Screen JA, Furtado JC, Barlow M, Whittleston D, Coumou D, Francis J, Dethloff K, Entekhabi D, Overland J, Jones J. 2014. Recent Arctic amplification and extreme mid-latitude weather. Nature Geosci 7:627–637. https://doi.org/10.1038/ngeo2234.

3. Mueller DR, Van Hove P, Antoniades D, Jeffries MO, Vincent WF. 2009. High Arctic lakes as sentinel ecosystems: cascading regime shifts in climate, ice cover, and mixing. Limnol Oceanogr 54:2371–2385. https://doi.org/10.4319/lo.2009.54.6_part_2.2371.

4. Sharma S, Blagrave K, Magnuson JJ, O'Reilly CM, Oliver S, Batt RD, Magee MR, Straile D, Weyhenmeyer GA, Winslow L, Woolway IR. 2019. Widespread loss of lake ice around the Northern Hemisphere in a warming world. Nat Clim Chang 9:227–231. https://doi.org/10.1038/s41558-018-0393-5.

5. Laybourn-Parry J, Bell EM. 2014. Ace Lake: three decades of research on a meromictic, Antarctic lake. Polar Biol 37:1685–1699. https://doi.org/10.1007/s00300-014-1553-3.

6. Hattersley-Smith G, Serson H, Keys J, Mielke J. 1970. Density stratified lakes in northern Ellesmere Island. Nature 225:55–56. https://doi.org/10.1038/225055a0.

7. Vincent AC, Mueller DR, Vincent WF. 2008. Simulated heat storage in a perennially ice-covered high Arctic lake: sensitivity to climate change. J Geophys Res Oceans 113: C04036. https://doi.org/10.1029/2007JC004360.

8. Vincent WF, Fortier D, Lévesque E, Boulanger-Lapointe N, Tremblay B, Sarrazin D, Antoniades D, Mueller DR. 2011. Extreme ecosystems and geosystems in the Canadian High Arctic: Ward Hunt Island and vicinity. Ecoscience 18:236–261. https://doi.org/10.2980/18-3-3448.

9. Charvet S, Vincent WF, Comeau A, Lovejoy C. 2012. Pyrosequencing analysis of the protist communities in a High Arctic meromictic lake: DNA preservation and change. Front Microbiol 3:422. https://doi.org/10.3389/fmicb.2012.00422.

10. Antoniades D, Veillette J, Martineau M-J, Belzile C, Tomkins J, Pienitz R, Lamoureux S, Vincent WF. 2009. Bacterial dominance of phototrophic communities in a High Arctic lake and its implications for paleoclimate analysis. Polar Sci 3:147–161. https://doi.org/10.1016/j.polar.2009.05.002.

11. Van Hove P, Belzile C, Gibson JAE, Vincent WF. 2006. Coupled landscape-lake evolution in High Arctic Canada. Can J Earth Sci 43:533–546. https://doi.org/10.1139/e06-003.

12. Comeau AM, Harding T, Galand PE, Vincent WF, Lovejoy C. 2012. Vertical distribution of microbial communities in a perennially stratified Arctic lake with saline, anoxic bottom waters. Sci Rep 2:604. https://doi.org/10.1038/srep00604.

13. Pouliot J, Galand PE, Lovejoy C, Vincent WF. 2009. Vertical structure of archaeal communities and the distribution of ammonia monooxygenase A gene variants in two meromictic High Arctic lakes. Environ Microbiol 11:687–699. https://doi.org/10.1111/j.1462-2920.2008.01846.x.

14. Comeau AM, Vincent WF, Bernier L, Lovejoy C. 2016. Novel chytrid lineages dominate fungal sequences in diverse marine and freshwater habitats. Sci Rep 6:30120. https://doi.org/10.1038/srep30120.

15. Chénard C, Chan AM, Vincent WF, Suttle CA. 2015. Polar freshwater cyanophage S-EIV1 represents a new widespread evolutionary lineage of phages. ISME J 9:2046–2058. https://doi.org/10.1038/ismej.2015.24.

16. Middelboe M, Brussaard CP. 2017. Marine viruses: key players in marine ecosystems. Viruses 9:E302. https://doi.org/10.3390/v9100302.

17. Suttle CA. 2016. Environmental microbiology: viral diversity on the global stage. Nat Microbiol 1:16205. https://doi.org/10.1038/nmicrobiol.2016.205.

18. Middelboe M, Jacquet S, Weinbauer M. 2008. Viruses in freshwater ecosystems: an introduction to the exploration of viruses in new aquatic habitats. Freshwater Biol 53:1069–1075. https://doi.org/10.1111/j.1365-2427.2008.02014.x.

19. Wommack EK, Colwell RR. 2000. Virioplankton: viruses in aquatic ecosystems. Microbiol Mol Biol Rev 64:69–114. https://doi.org/10.1128/mmbr.64.1.69-114.2000.

20. Rohwer F, Thurber R. 2009. Viruses manipulate the marine environment. Nature 459:207–212. https://doi.org/10.1038/nature08060.

21. Weitz JS, Wilhelm SW. 2012. Ocean viruses and their effects on microbial communities and biogeochemical cycles. F1000 Biol Rep 4:17. https://doi.org/10.3410/B4-17.

22. Yau S, Seth-Pasricha M. 2019. Viruses of polar aquatic environments. Viruses 11:E189. https://doi.org/10.3390/v11020189.

23. Gregory A, Zayed A, Jang HB, Conceicao-Neto N, Temperton B, Bolduc B, Alberti A, Ardyna M, Arkhipova K, Carmichael M, Cruaud C, Dimier C, Dominguez-Huerta G, Ferland J, Kandels-Lewis S, Liu Y, Marec C, Pesant S, Picheral M, Pisarev S, Poulain J, Tremblay J-E, Vik D, Tara Oceans Coordinators, Babin M, Bowler C, Culley AI, Vargas C, Dutilh BE, Iudicone D, Karp-Boss L, Roux S, Sunagawa S, Wincker P, Sullivan MB. 2019. Marine DNA viral macro- and micro-diversity from pole to pole. Cell 177: 1109–1123. https://doi.org/10.1016/j.cell.2019.03.040.

24. Aguirre de Cárcer D, López-Bueno A, Pearce DA, Alcamí A. 2015. Biodiversity and distribution of polar freshwater DNA viruses. Sci Adv 1:e1400127. https://doi.org/10.1126/sciadv.1400127.

25. Kim KH, Bae JW. 2011. Amplification methods bias metagenomic libraries of uncultured single-stranded and double-stranded DNA viruses. Appl Environ Microbiol 77:7663–7668. https://doi.org/10.1128/AEM.00289-11.

26. Allen L, McCrow JP, Ininbergs K, Dupont CL, Badger JH, Hoffman JM, Ekman M, Allen AE, Bergman B, Venter CJ. 2017. The Baltic Sea virome: diversity and transcriptional activity of DNA and RNA viruses. mSystems 2:e00125-16. https://doi.org/10.1128/mSystems.00125-16.

27. López-Bueno A, Tamames J, Velázquez D, Moya A, Quesada A, Alcamí A. 2009. High diversity of the viral community from an Antarctic lake. Science 326:858–861. https://doi.org/10.1126/science.1179287.

28. Madan NJ, Marshall WA, Laybourn-Parry J. 2005. Virus and microbial loop dynamics over an annual cycle in three contrasting Antarctic lakes. Freshwater Biol 50:1291–1300. https://doi.org/10.1111/j.1365-2427.2005.01399.x.

29. Lauro FM, DeMaere MZ, Yau S, Brown MV, Ng C, Wilkins D, Raftery MJ, Gibson JA, Andrews-Pfannkoch C, Lewis M, Hoffman JM, Thomas T, Cavicchioli R. 2011. An integrative study of a meromictic lake ecosystem in Antarctica. ISME J 5:879–895. https://doi.org/10.1038/ismej.2010.185.

30. Yau S, Lauro FM, DeMaere MZ, Brown MV, Thomas T, Raftery MJ, Andrews-Pfannkoch C, Lewis M, Hoffman JM, Gibson JA, Cavicchioli R. 2011. Virophage control of Antarctic algal host-virus dynamics. Proc Natl Acad Sci U S A 108:6163–6168. https://doi.org/10.1073/pnas.1018221108.

31. Nishimura Y, Yoshida T, Kuronishi M, Uehara H, Ogata H, Goto S. 2017.

ViPTree: the viral proteomic tree server. Bioinformatics 33:2379–2380. https://doi.org/10.1093/bioinformatics/btx157.

32. Paez-Espino D, Chen I-M, Palaniappan K, Ratner A, Chu K, Szeto E, Pillay M, Huang J, Markowitz VM, Nielsen T, Huntemann M, Reddy TK, Pavlopoulos GA, Sullivan MB, Campbell BJ, Chen F, McMahon K, Hallam SJ, Denef V, Cavicchioli R, Caffrey SM, Streit WR, Webster J, Handley KM, Salekdeh GH, Tsesmetzis N, Setubal JC, Pope PB, Liu W-T, Rivers AR, Ivanova NN, Kyrpides NC. 2017. IMG/VR: a database of cultured and uncultured DNA viruses and retroviruses. Nucleic Acids Res 45:D457–D465. https://doi.org/10.1093/nar/gkw1030.

33. Tara Oceans Consortium, Coordinators; Tara Oceans Expedition, Participants. 2015. Biodiversity context of all samples from the Tara Oceans Expedition (2009–2013). PANGAEA https://doi.org/10.1594/PANGAEA.853809.

34. Choi I, Ponsero A, Bomhoff M, Youens-Clark K, Hartman J, Hurwitz B. 28 December 2018. Libra: scalable k-mer based tool for massive all-vs-all metagenome comparisons. GigaScience https://doi.org/10.1093/gigascience/giy165.

35. Van Hove P, Swadling KM, Gibson JA, Belzile C, Vincent WF. 2001. Farthest north lake and fjord populations of calanoid copepods Limnocalanus macrurus and Drepanopus bungei in the Canadian High Arctic. Polar Biol 24:303–307. https://doi.org/10.1007/s003000000207.

36. Suttle CA, Chen F. 1992. Mechanisms and rates of decay of marine viruses in seawater. Appl Environ Microbiol 58:3721–3729. https://doi.org/10.1128/AEM.58.11.3721-3729.1992.

37. Chow C-E, Suttle CA. 2015. Biogeography of viruses in the sea. Annu Rev Virol 2:41–66. https://doi.org/10.1146/annurev-virology-031413-085540.

38. Okazaki Y, Nishimura Y, Yoshida T, Ogata H, Nakano S. 2019. Genome-resolved viral and cellular metagenomes revealed potential key virus-host interactions in a deep freshwater lake. Environ Microbiol 21:4740–4754. https://doi.org/10.1111/1462-2920.14816.

39. Wu Y-T, Yang C-Y, Chiang P-W, Tseng C-H, Chiu H-H, Saeed I, Baatar A, Rogozin D, Halgamuge S, Degermendzhi A, Tang S. 2018. Comprehensive insights into composition, metabolic potentials, and interactions among archaeal, bacterial, and viral assemblages in meromictic Lake Shunet in Siberia. Front Microbiol 9:1763–1021. https://doi.org/10.3389/fmicb.2018.01763.

40. Watkins SC, Kuehnle N, Ruggeri CA, Malki K, Bruder K, Elayyan J, Damisch K, Vahora N, O'Malley P, Ruggles-Sage B, Romer Z, Putonti C. 2016. Assessment of a metaviromic dataset generated from nearshore Lake Michigan. Mar Freshw Res 67:1700–1708. https://doi.org/10.1071/MF15172.

41. Fancello L, Trape S, Robert C, Boyer M, Popgeorgiev N, Raoult D, Desnues C. 2013. Viruses in the desert: a metagenomic survey of viral communities in four perennial ponds of the Mauritanian Sahara. ISME J 7:359–369. https://doi.org/10.1038/ismej.2012.101.

42. Arkhipova K, Skvortsov T, Quinn JP, McGrath JW, Allen CC, Dutilh BE, McElarney Y, Kulakov LA. 2018. Temporal dynamics of uncultured viruses: a new dimension in viral diversity. ISME J 12:199–211. https://doi.org/10.1038/ismej.2017.157.

43. Moon K, Kang I, Kim S, Kim S, Cho J. 2017. Genome characteristics and environmental distribution of the first phage that infects the LD28 clade, a freshwater methylotrophic bacterial group. Environ Microbiol 19:4714–4727. https://doi.org/10.1111/1462-2920.13936.

44. Green J, Rahman F, Saxton M, Williamson K. 2015. Metagenomic assessment of viral diversity in Lake Matoaka, a temperate, eutrophic freshwater lake in southeastern Virginia, USA. Aquat Microb Ecol 75:117–128. https://doi.org/10.3354/ame01752.

45. Tseng C-H, Chiang P-W, Shiah F-K, Chen Y-L, Liou J-R, Hsu T-C, Maheswararajah S, Saeed I, Halgamuge S, Tang S-L. 2013. Microbial and viral metagenomes of a subtropical freshwater reservoir subject to climatic disturbances. ISME J 7:2374–2386. https://doi.org/10.1038/ismej.2013.118.

46. Roux S, Enault F, Robin A, Ravet V, Personnic S, Theil S, Colombet J, Sime-Ngando T, Debroas D. 2012. Assessing the diversity and specificity of two freshwater viral communities through metagenomics. PLoS One 7:e33641. https://doi.org/10.1371/journal.pone.0033641.

47. Lang AS, Rise ML, Culley AI, Steward GF. 2009. RNA viruses in the sea. FEMS Microbiol Rev 33:295–323. https://doi.org/10.1111/j.1574-6976.2008.00132.x.

48. Colson P, De Lamballerie X, Yutin N, Asgari S, Bigot Y, Bideshi DK, Cheng X-W, Federici BA, Van Etten JL, Koonin EV, La Scola B, Raoult D. 2013. 'Megavirales', a proposed new order for eukaryotic nucleocytoplasmic large DNA viruses. Arch Virol 158:2517–2521. https://doi.org/10.1007/s00705-013-1768-6.

49. Winter C, Köstner N, Kruspe C, Urban D, Muck S, Reinthaler T, Herndl GJ.

50. 2018. Mixing alters the lytic activity of viruses in the dark ocean. Ecology 99:700–713. https://doi.org/10.1002/ecy.2135.

50. Laybourn-Parry J, Anesio AM, Madan N, Säwström C. 2013. Virus dynamics in a large epishelf lake (Beaver Lake). Freshw Biol 58:1484–1493. https://doi.org/10.1111/fwb.12146.

51. Magagnini M, Corinaldesi C, Monticelli LS, Domenico E, Danovaro R. 2007. Viral abundance and distribution in mesopelagic and bathypelagic waters of the Mediterranean Sea. Deep Sea Res Part 1 Oceanogr Res Pap 54:1209–1220. https://doi.org/10.1016/j.dsr.2007.05.006.

52. Coutinho F, Gregoracci G, Walter J, Thompson C, Thompson FL. 2018. Metagenomics sheds light on the ecology of marine microbes and their viruses. Trends Microbiol 26:955–965. https://doi.org/10.1016/j.tim.2018.05.015.

53. Aguirre de Cárcer D, López-Bueno A, Alonso-Lobo JM, Quesada A, Alcamí A. 2016. Metagenomic analysis of lacustrine viral diversity along a latitudinal transect of the Antarctic Peninsula. FEMS Microbiol Ecol 92:fiw074. https://doi.org/10.1093/femsec/fiw074.

54. Li W, Morgan-Kiss RM. 2019. Influence of environmental drivers and potential interactions on the distribution of microbial communities from three permanently stratified Antarctic lakes. Front Microbiol 10:1067. https://doi.org/10.3389/fmicb.2019.01067.

55. Laurion I, Vincent WF, MacIntyre S, Retamal L, Dupont C, Francus P, Pienitz R. 2010. Variability in greenhouse gas emissions from permafrost thaw ponds. Limnol Oceanogr 55:115–133. https://doi.org/10.4319/lo.2010.55.1.0115.

56. Brussaard CPD, Payet JP, Winter C, Weinbauer MG. 2010. Quantification of aquatic viruses by flow cytometry, p 102–109. In Wilhelm SW, Weinbauer MG [ed], Manual of aquatic viral ecology. ASLO, Waco, TX.

57. Thaler M, Vincent WF, Lionard M, Hamilton AK, Lovejoy C. 2017. Microbial community structure and interannual change in the last epishelf lake ecosystem in the north polar region. Front Mar Sci 3:275. https://doi.org/10.3389/fmars.2016.00275.

58. Mueller JA, Culley AI, Steward GF. 2014. Variables influencing extraction of nucleic acids from microbial plankton (viruses, bacteria, and protists) collected on nanoporous aluminum oxide filters. Appl Environ Microbiol 80:3930–3942. https://doi.org/10.1128/AEM.00245-14.

59. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

60. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

61. Allam A, Kalnis P, Solovyev V. 2015. Karect: accurate correction of substitution, insertion and deletion errors for next-generation sequencing data. Bioinformatics 31:3421–3428. https://doi.org/10.1093/bioinformatics/btv415.

62. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. 2017. metaSPAdes: a new versatile metagenomic assembler. Genome Res 27:824–834. https://doi.org/10.1101/gr.213959.116.

63. Roux S, Enault F, Hurwitz BL, Sullivan MB. 2015. VirSorter: mining viral signal from microbial genomic data. PeerJ 3:e985. https://doi.org/10.7717/peerj.985.

64. Ren J, Ahlgren NA, Lu Y, Fuhrman JA, Sun F. 2017. VirFinder: a novel k-mer based tool for identifying viral sequences from assembled metagenomic data. Microbiome 5:69. https://doi.org/10.1186/s40168-017-0283-5.

65. Bolduc B, Youens-Clark K, Roux S, Hurwitz BL, Sullivan MB. 2017. iVirus: facilitating new insights in viral ecology with software and community data sets imbedded in a cyberinfrastructure. ISME J 11:7–14. https://doi.org/10.1038/ismej.2016.89.

66. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923.

67. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

68. Brum JR, Ignacio-Espinoza JC, Roux S, Doulcier G, Acinas SG, Alberti A, Chaffron S, Cruaud C, de Vargas C, Gasol JM, Gorsky G, Gregory AC, Guidi L, Hingamp P, Iudicone D, Not F, Ogata H, Pesant S, Poulos BT, Schwenck SM, Speich S, Dimier C, Kandels-Lewis S, Picheral M, Searson S, Tara Oceans Coordinators, Bork P, Bowler C, Sunagawa S, Wincker P, Karsenti E, Sullivan MB. 2015. Ocean plankton. Patterns and ecological drivers of ocean viral communities. Science 348:1261498. https://doi.org/10.1126/science.1261498.

69. Oksanen J. 2011. Multivariate analysis of ecological communities in R: vegan tutorial. R package version 1.17–7. https://github.com/vegandevs/vegan.

70. Ogle DH, Wheeler P, Dinno A. 2019. FSA: Fisheries Stock Analysis. R package version 0.8.25. https://github.com/droglenc/FSA.

71. Paradis E, Claude J, Strimmer K. 2004. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics 20:289–290. https://doi.org/10.1093/bioinformatics/btg412.

72. Oliveros JC. 2015. Venny. An interactive tool for comparing lists with Venn's diagrams. https://bioinfogp.cnb.csic.es/tools/venny/index.html.

73. Chen H, Boutros PC. 2011. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinformatics 12:35. https://doi.org/10.1186/1471-2105-12-35.

74. Taiyun W, Simko V. 2017. R package "corrplot": visualization of a Correlation Matrix. R package version 0.84. https://github.com/taiyun/corrplot.

75. Wickham H. 2009. ggplot2: elegant graphics for data analysis. Use R! Springer, New York, NY. https://doi.org/10.1007/978-0-387-98141-3_9.

76. Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9:559. https://doi.org/10.1186/1471-2105-9-559.

77. Guidi L, Chaffron S, Bittner L, Eveillard D, Larhlimi A, Roux S, Darzi Y, Audic S, Berline L, Brum J, Coelho LP, Espinoza JC, Malviya S, Sunagawa S, Dimier C, Kandels-Lewis S, Picheral M, Poulain J, Searson S, Tara Oceans coordinators, Stemmann L, Not F, Hingamp P, Speich S, Follows M, Karp-Boss L, Boss E, Ogata H, Pesant S, Weissenbach J, Wincker P, Acinas SG, Bork P, de Vargas C, Iudicone D, Sullivan MB, Raes J, Karsenti E, Bowler C, Gorsky G. 2016. Plankton networks driving carbon export in the oligotrophic ocean. Nature 532:465–470. https://doi.org/10.1038/nature16942.

78. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

79. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153.

80. Suzek BE, Wang Y, Huang H, McGarvey PB, Wu CH; the UniProt Consortium. 2015. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. Bioinformatics 31:926–932. https://doi.org/10.1093/bioinformatics/btu739.

81. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31:1674–1676. https://doi.org/10.1093/bioinformatics/btv033.

82. Keegan KP, Glass EM, Meyer F. 2016. MG-RAST, a metagenomics service for analysis of microbial community structure and function, p 207–233. In Martin F, Uroz S (ed), Microbial environmental genomics (MEG). Methods in Molecular Biology, vol 1399. Humana Press, New York, NY. https://doi.org/10.1007/978-1-4939-3369-3_13.

83. Northern Ellesmere Island in the Global Environment (NEIGE). 2017. Water column physico-chemical profiles of lakes and fiords along the northern coastline of Ellesmere Island, v. 1.1 (1954–2016). Nordicana D27. (In French and English.) https://doi.org/10.5885/45445CE-7B8194DB81754841.