Chlorovirus and myovirus diversity in permafrost thaw ponds

Alice V. Lévesque[1,2,3], Warwick F. Vincent[2,4], Jérôme Comte[2,5], Connie Lovejoy[3,4,6], Alexander I. Culley[1,2,3]

1 Département de biochimie, microbiologie et bio-informatique, Université Laval, Québec, QC G1V 0A6, Canada
2 Centre d'études nordiques (CEN), Université Laval, Québec, QC G1V 0A6, Canada
3 Institut de Biologie Intégrative et des Systèmes, Université Laval, Québec, QC G1V 0A6, Canada
4 Département de biologie, Université Laval, Québec, QC G1V 0A6, Canada
5 Centre − Eau Terre Environnement, Institut national de la recherche scientifique, Québec, QC G1K 9A9, Canada 6Québec-Océan, Université Laval, Québec, QC G1V 0A6, Canada

Accepted for publication in *Aquat Microb Ecol* 82: 209–224, October 20, 2018

Abstract

Permafrost thaw ponds occur in high abundance across the northern landscape of Canada and are sites of intense microbial activity, resulting in carbon dioxide and methane emissions to the atmosphere. In this study, we focused on viruses as largely unstudied agents of top-down control in these high-latitude microbial ecosystems. Specifically, we compared the diversity of myovirus, chlorovirus and host microbial communities in an organic soil palsa valley pond and a mineral soil lithalsa valley pond. These 2 subarctic permafrost landscapes are both common in northern Québec, Canada. Sequence analysis of ribosomal small subunit RNA genes showed that the community structure of bacteria and microbial eukaryotes differed significantly between the 2 ponds, which both differed from microbial communities in a rock-basin lake (whose formation was not related to permafrost thawing and which we used as a reference pond) in the same region. The viral assemblages included 439 OTUs in the uncultured *Myoviridae* category and 41 OTUs in the family *Phycodnaviridae*. Phylogenetic analysis of the latter based on an amino acid sequence alignment revealed a single large clade related to chloroviruses, consistent with the abundant presence of chlorophytes in these waters. As there was for the bacterial and eukaryotic communi-ties, there were also significant differences in the community structure of these viral groups among the 3 ponds. These results suggest that host community composition is influenced by environmental filtering, which in turn contributes to driving viral diversity across landscape types.

Keywords: Arctic, Biodiversity, Microbial ecology, Permafrost, Thaw ponds, Viral diversity

INTRODUCTION

Viruses are an abundant component of aquatic ecosystems (Suttle 2005), where they may regulate host abundance and community structure, catalyze evolution through the mediation of gene exchange, and modify biogeochemical cycles (Wilhelm & Suttle 1999, Brussaard 2004, Rohwer & Thurber 2009, Gao et al. 2016, Joli et al. 2017). At high northern latitudes, viruses have been reported from a diversity of aquatic habitats, including arctic and subarctic lakes (de Carcer et al. 2016, Zeigler Allen et al. 2017), cryoconite holes on glaciers (Anesio et al. 2007), sea-ice brines (Wells & Deming 2006), arctic marine sediments (Colangelo-Lillis et al. 2016) and the Arctic Ocean (Guixa-Boixereu et al. 2002).

In some northern permafrost landscapes, global warming has accelerated the formation of small lakes formed by thermokarst: the thawing and collapse of ice-rich soils (Pienitz et al. 2008, Vincent et al. 2017). These thermokarst waterbodies (thaw ponds) are strong emitters of the greenhouse gases carbon dioxide and methane to the atmosphere (Laurion et al. 2010, Matveev et al. 2016). The net production of these gases is controlled by the activity of microbes (Crevecoeur et al. 2015, 2016), which are strongly shaped by local environmental conditions (e.g. Crevecoeur et al. 2015, Comte et al. 2016b, Przytulska et al. 2016) but also subject to predation by grazers and viruses. Despite the fact that thermokarst ponds are among the most abundant freshwater ecosystems in the circumpolar North (Olefeldt et al. 2016), the presence and diversity of the viral community re mains largely unknown.

In the Hudson Bay region of subarctic Québec, thermokarst ponds are subjected to extremes in light, organic carbon availability, and oxygen. These ponds are typically vertically stratified, with a high-light oxygenated surface layer and a low-light hypoxic or anoxic stratum of bottom water in the summer. During winter, when the ponds are ice-covered, light is limited or absent and the water column becomes completely anoxic (Deshpande et al. 2015). Two types of permafrost landscapes are found in the Québec subarctic: (1) palsa thaw ponds formed in hummocks of peatland, and (2) lithalsa ponds that form in hummocks of mineral-rich soils (Gurney 2001, Calmels et al. 2008). In addition, non-permafrost lakes with rocky basins can be found in adjacent regions at similar latitudes. The type of landscape ultimately affects the quantity, nature and availability of organic material in the pond (Watan-abe et al. 2011, Kiikkila et al. 2014), and appears to be a principal driver of microbial community composition (Comte et al. 2016b).

Viruses are intracellular obligate parasites and therefore their dynamics are inextricably linked to those of their hosts. The diversity of *Bacteria* (Comte et al. 2016a,b), *Archaea* (Crevecoeur et al. 2015, 2016) and protists (Przytulska et al. 2016, 2017, Bégin & Vincent 2017), the presumptive primary hosts of the subarctic thermokarst viruses, has recently been characterized in several Québec subarctic ponds. Crevecoeur et al. (2015) reported that the largest differences in bacterial diversity were between the 2 permafrost landscape types (palsa vs. lithalsa), and Comte et al. (2016b) concluded that the composition of these communities appears to be primarily determined by environmental filtering and dispersal limitation. The species composition of archaeal communities was significantly different between palsa and lithalsa ponds (Crevecoeur et al. 2017), similar to the pattern exhibited by bacterial communities. Photo-synthetic groups in these ponds include *Cyanobacteria* (across all size classes, from picoplankton to colonial taxa), chlorophytes, cryptophytes and dinoflagellates, as well as photosynthetic sulphur bacteria (Przytulska et al. 2016). There was no clear difference in algal communities between palsa and lithalsa waters, based on microscopy and morphotypes (Przy-tulska et al. 2017).

The aim in the present study was to characterize 2 very different classes of viruses, chloroviruses and myoviruses, which are globally distributed and infect potential eukaryotic and prokaryotic host types that have been reported in thaw lakes and ponds (Przytulska et al. 2016, Crevecoeur et al. 2017). With one exception (Brussaard 2004), chlorophytes (primary endosymbiotic-origin oxygenic phototrophs) are primarily in fected by viruses from a genus of large double-stranded DNA viruses (*Chlorovirus*) within the family *Phycodnaviridae* (Short 2012). The availability of specific primers (Short et al. 2011) targeting a fragment of the polymerase gene *polB* from this genus has resulted in a large comparative database of *polB*-containing viruses in freshwater, making *polB* a logical choice for this first study of the potential occurrences of chloroviruses in the subarctic. The approach described by Short et al. (2011) yields amplification products of both known and divergent virus phylotypes. Bacteriophages are abundant, ubiquitous and highly diverse in aquatic environments (Breit bart et al. 2007, Matteson et al. 2013). Myoviruses (a type of bacteriophage) are widely reported from freshwater environments, since primers targeting a conserved structural gene (Matteson et al. 2011) are available and have been widely used. For example, myo viruses are present in extreme, temperate and tropical environments (Wilhelm & Matteson 2008). The lack of universal viral primers means that the best available primers for widely distributed groups will provide a window into differences between lakes. We therefore tested these known primers for chloroviruses and myoviruses to address whether: (1) thaw pond viruses are divergent from known types, (2) there were distinct viral signatures in ponds from different landscape types (palsa vs. lithalsa vs. rock basin), and (3) whether the viral communities be tween sur-face and bottom layers of the ponds differ.

Our approach in this study was to analyze samples of virus communities taken from the surface and deeper waters of palsa, lithalsa and shallow rock-basin ponds in subarctic Québec during the summer ice-free period of maximum biological activity. Viral communities were identified using high-throughput sequencing (Illumina MiSeq) of the fragments of the *gp23* structural gene of myoviruses and *polB* polymerase genes of chloroviruses. We also amplified the V6−V8 region of the bacterial and archaeal 16S rRNA gene and the V4 region of the eukaryotic 18S rRNA gene to identify informative patterns between viruses and their putative hosts.

MATERIALS AND METHODS

Study site and sampling

Samples were collected from August 27 to 30, 2015, from the 2 types of permafrost thaw ponds in northern Québec (Nunavik, Canada) (Fig. 1). The SAS2A site is a thaw pond in a palsa valley adjacent to the Sasapimakwananisikw River (55° 13’ N, 77° 42’ W), located in the sporadic permafrost zone, where the extent of permafrost ranges from 10 to 50%. The second site, BGR1, is a lithalsa pond (Bhiry et al. 2011) located in the Bundesanstalt für Geo wissen schaften und Rohstoffe valley (56° 37’ N, 76° 13’ W) near the Sheldrake River in a region of discontinuous permafrost where the extent of permafrost ranges from 50 to 90%. As a point of reference, we sampled Lake Olsha (55° 16’ N, 77° 44’ W), a well-mixed shallow rock-basin lake near Whapmagoostui-Kuujjuarapik, whose formation was not related to permafrost thawing.

Triplicate samples were collected from oxygenated depths (10 cm below the surface) referred to as surface samples, and hypoxic/anoxic subsurface depths, collectively referred to as ‘bottom samples’ at 0.5 m (SAS2A), 3.5 m (BGR1) and 1.5 m (Olsha) (Table S1 in the Supplement at www.int-res.com/ articles/ suppl/a082p209_supp.pdf). Temperature, conductivity, dissolved oxygen (DO) and pH were measured with a Hydrolab™ DS5X profiler. Samples were collected from an inflatable boat from the deepest region of each pond. Hydrolab data were not collected from the BGR1 site, but seasonal depth data for this site are available in Deshpande et al. (2015). Surface water was sampled with acid-washed opaque bottles (Nalgene™), while bottom samples were collected with a horizontally mounted Van Dorn bottle (Wilco). Water samples were transferred into acid-washed Cubitainers™ that were stored in opaque bags and transported back to the laboratory via helicopter for processing. Dissolved organic carbon (DOC), total suspended solids (TSS), total nitrogen (TN) and total phosphorus (TP) were analyzed as in Laurion et al. (2010).

Sample processing and molecular analysis

For each replicate, total viruses (i.e. samples that were not pre-filtered and that likely included extracellular viruses, intracellular viruses and viruses adsorbed to particles), bacteria and eukaryotes were collected by direct filtration of the water samples through 0.02 µm pore size Anodisc aluminum oxide (AAO) filters (Anotop 25, Whatman). The volumes of sample filtered through the AAO filters ranged from 5 to 65 ml, 105 to 245 ml, and 53 to 115 ml for the sites SAS2A, BGR1, and Olsha, respectively. These filters were then stored at −80°C until extraction of the total nucleic acids as described in Mueller et al. (2014). After extraction, samples were treated with the Power Clean Pro DNA Clean up kit (MoBio) which removes potential PCR inhibitors.

We initially conducted PCR on a subset of samples with a suite of primer sets (AVS, CHLV, mcp, T7mcp, T4super and CPS) targeting a variety of molecular marker genes from a diversity of groups of viruses (see Adriaenssens & Cowan 2014 for a review). Based on the results of these assays, we performed PCR on all samples with the myovirus primers T4 superF1 and T4superR1 (Chow & Fuhrman 2012) and the chlorovirus primers

CHLVdF and CHLVdR (Short et al. 2011) listed in our Table 1. Each reaction mixture (final volume: 25 µl) contained 200 µM of each dNTP (Bio Basics), 0.4 µM of each primer, 1× Expand buffer containing $MgCl_2$, 2 U of Expand High-fidelity enzyme blend (Roche) and 3 µl of the purified DNA. The following thermocycler conditions were used: 94°C for 120 s, followed by 40 cycles of denaturation at 94°C for 30 s (CHLVd) or 15 s (T4super), annealing at 52°C (CHLVd) or 54°C (T4super) for 30 s and extension at 72°C for 60 s and then a final extension step at 72°C for 7 min. PCR products were separated on a 1% agarose gel, and DNA of the target size was purified using Axygen magnetic beads (Corning Life Sciences). In preparation for sequencing, Illumina TruSeq adaptors and unique barcodes were added with further rounds of PCR for each reaction. Sample barcoded amplicons were subsequently pooled in equimolar concentrations for sequencing on an Illumina MiSeq at the Plateforme d'Analyses Génomiques (IBIS, Université Laval, Québec, Canada). Sub-regions of the bacterial 16S rRNA gene (V6−V8 region, primers B969F and BA1406R) and eukaryotic 18S rRNA gene (V4 region, primers E572F and E1009R) were amplified using the primers described in Comeau et al. (2011, 2016) with template from extracted AAO filters. The raw Illumina sequences have been deposited in the NCBI Sequence Read Archive (SRA) database with the following identifiers: SRP115531 (bacteria), SRP115532 (viruses), and SRP115533 (eukaryotes).

Bioinformatics and statistical analyses

Sequence reads were first analyzed using the UPARSE quality-filtering pipeline to exclude chimeras and singletons (Edgar 2013) as described in Comte et al. (2016a). Briefly, reads with low Q scores, short reads (<300 nucleotides) and singletons, i.e. reads occurring just once in the whole dataset, were discarded. Chimeras were detected using UCHIME (Edgar et al. 2011) as implemented in USEARCH with SILVA 16S and 18S sequence databases (no database available for viruses). The remaining reads were clustered into operational taxonomic units (OTUs) with >97% identity (16S rRNA), >98% identity (myovirus and 18S rRNA) and >99% identity (chlorovirus). The virus % identity cut-offs were determined from a phylogenetic analysis of classified myo- and phycodnaviruses based on the approach described by Culley & Steward (2007) for the classification of environmental picorna-like virus sequences. Bacterial and eukaryotic sequences were classified based on an alignment of OTUs with the SILVA reference database (Pruesse et al. 2007) and a curated reference database of arctic protists (Lovejoy et al. 2016), respectively. Sequences classified as fungi, Streptophyta, metazoan, chloroplasts and *Archaea* (248 OTUs in total) were removed from the dataset. Viral OTUs were classified with DIAMOND (Buchfink et al. 2015) to identify the nearest match in the indexed nrBLAST database. Results were visualized in MEGAN6 (Huson et al. 2016).

Statistical analyses were executed using QIIME (Caporaso et al. 2010) and RStudio v3.3.0 (Oksanen et al. 2016) on a subset of the data (27 396 bacterial, 27 221 eukaryotic, and 28 511 myovirus sequences), with the exception of chloroviruses, where all 39 sequences were used. Alpha-diversity metrics were calculated using the QIIME command alpha_diversity.py (Chao1, Shannon index). The assumptions of normality and homoscedasticity for an ANOVA were confirmed using a Shapiro-Wilk test. After, a 2-way ANOVA and an a *posteriori* Tukey HSD test were used to assess the significance of pond type or depth on alpha-diversity. Venn diagrams were produced using the VennDiagram package in R (Chen & Boutros 2011) to visualize the percentages of unique and shared OTUs among ponds. A UniFrac distance matrix was generated (beta_diversity.py) after subsampling the dataset and was used to build a jackknife phylogenetic tree. Heatmaps representing the relative abundance of the most abundant OTUs (defined as >1% of reads in the dataset) in each sample were produced to identify specific distribution patterns. Principal coordinate analysis (PCoA) scatter plots based on unweighted UniFrac distances were generated to evaluate the environmental variables potentially influencing diversity patterns among the pond microbial communities.

To gain further insight into the taxonomy of the thaw pond chlorovirus phylotypes, a maximumlikelihood (ML) reference tree (RAxML v8.2.9; Stamatakis 2014) was produced based on an alignment (MUSCLE) (Edgar 2004) of representative amino acid viral sequences from the family *Phycodnaviridae*. Alignments were then transformed into distances with RAxML v8.2.9 (Stamatakis 2014). The 39 chlorovirus sequences were subsequently translated and added to the RAxML phylogeny with the evolutionary placement algorithm (EPA; Berger et al. 2011).

RESULTS

Limnological conditions

The data from the palsa thaw pond (Site SAS2A) were consistent with previous studies from the region and indicated that the thaw ponds are generally highly stratified with warmer, oxygenated water at the surface and cooler, anoxic, and more ion-rich water at the bottom (see Table S1). Conversely, the shallow rock-basin lake (Site Olsha) water column was well-mixed, oxygenated from surface to bottom, and had relatively low conductivity (Deshpande et al. 2015). Site SAS2A also had relatively lower pH values and higher DOC, TP and TN concentrations, while Site BGR1 had higher concentrations of TSS and bottom-water chl *a* (Deshpande et al. 2015, Przytulska et al. 2016).

Cellular communities

Proteobacteria followed by *Actinobacteria*, *Bacteroidetes* and *Verrucomicrobia* were the taxonomic groups with the highest proportion of OTUs in all 3 ponds (Fig. 2A). Within the phylum *Proteobacteria*, the *Betaproteobacteria* comprised between approximately 60 and 90% of OTUs. Members of the genus *Polynucleobacter* (*Betaproteobacteria*) had the highest relative abundance at sites SAS2A and Olsha, while *Actinobacteria* were proportionally most abundant in BGR1. Cyanobacteria were detected at all 3 sites but had greater relative abundance in Olsha (10%) than in the thaw ponds (<2%).

At all 3 sampling sites, the major groups of eukaryotes with the highest relative abundances were Ciliophora, Chlorophyta, Chrysophyceae and Cryptomonadales (Fig. 2B). Chlorophytes tended to have higher relative abundances in bottom waters, while the surface waters had high relative abundances of dinoflagellates, cryptomonads and ciliates in SAS2A, BGR1 and Olsha, respectively. In general, autotrophic protists represented a high percentage of the total community, ranging from 42% in Olsha to 79% in SAS2A.

The PCoA based on unweighted UniFrac distances (Fig. 3) showed distinct clustering of bacterial and eukaryotic communities by pond type, but no evident clustering by depth. SAS2A bacterial communities were distinct from BGR1 and Olsha along the *x*-axis (44% of the variance explained, Fig. 3A). Although BGR1 and Olsha communities overlapped along the *x*-axis, these 2 communities were separate along the *y*-axis (28% of the variance explained, Fig. 3A). Olsha eukaryote communities were distinct from the 2 thaw ponds both along the *x*-axis (36% of the variance explained) and *y*-axis (32% of the variance explained, Fig. 3B).

UniFrac UPGMA trees based on the top 1% (Fig. 4) showed well-supported patterns in OTU distribution by pond type, but not by depth, with the exception of BGR1 bacterial and eukaryotic communities (Fig. 4A,B). The dominant bacterial phylotype was a member of the *Puniceicoccaceae* and the dominant eukaryotic phylotype was a member of the Chlorophyceae.

Bacterial diversity patterns were generally consistent with those of the PCoA (Fig. 5). The Shannon indices from all 3 pond types were significantly different (ANOVA, $p < 0.05$), while surface and bottom indices were not. Mean values and ranges for the 2 indices, surface and bottom, for the 3 sites are listed in Table S2 in the Supplement.

Eukaryotic community Shannon indices were significantly different among all 3 sites (ANOVA, $p < 0.05$). Surface and bottom indices from BGR1 were significantly different; however, the indices from Olsha and SAS2A were not. Eukaryotic richness was highest in the bottom waters of Lake Olsha (462 ± 52 mean ± SD) and lowest in the bottom waters of BGR1 (260 ± 31). Mean values and ranges for the 2 indices are listed in Table S2.

Venn diagrams (Fig. 6) showed that 22.2% of bacterial OTUs and 12.9% of eukaryotic OTUs were shared among the 3 sites. The highest percentage of unique OTUs for bacteria (32.1%) and eukaryotes (25.6%) was in site SAS2A and the highest percentage of shared OTUs for bacteria (40.8%) and for eukaryotes (23.1%) was between sites Olsha and SAS2A. A Venn diagram of chlorophytes (Fig. 7) revealed that 21.4% of OTUs were shared among the 3 sites and that the highest percentage of OTUs were shared between sites BGR1 and Olsha (16.8%).

Viral communities

Classification based on DIAMOND revealed that the majority of OTUs matched with sequences in the uncultured *Myoviridae* category (439 out of 872 total OTUs), while fewer OTUs matched with the myoviruses *Enterobacteria* phage T4 and *Caulobacter* phage Cr30 (Fig. 8A). A single phylotype was identified as a cyanophage (*Synechococcus* phage SCRM01), while a large number (221) of the phylotypes were binned in the general category of 'Viruses'.

The family *Phycodnaviridae* was matched by 41 OTUs (Fig. 8B), with 1 assigned to the genus *Chlorovirus*. To further resolve the taxonomy of the chlorovirus phylotypes, we produced an ML phylogeny analysis based on an amino acid sequence alignment that resulted in a single large clade with a bootstrap value of 100 that contained all of the environmental OTUs from the present study, sequences amplified with the same primer set from Lake Ontario (Short et al. 2011), and from the chlorovirus isolate *Acanthocystis turfacea Chlorella virus 1* (Fig. 9). This clade in turn formed a well-supported group (bootstrap value of 100) with the chloroviruses *Paramecium bursaria Chlorella virus 1* and *Paramecium bursaria Chlorella virus NY2A*. Established genera of phycodnaviruses with multiple representatives were resolved into monophyletic clusters with bootstrap values of 100.

The PCoA based on unweighted UniFrac distances showed distinct clustering of myovirus and chlorovirus communities by pond (with 1 exception) but not by depth (Fig. 3C,D). Myovirus communities at all 3 sites were separate along the *x*-axis (42% of the variance explained, Fig. 3C). BGR1 and Olsha myovirus communities overlapped along the *y*-axis (38% of the variance explained), while SAS2A was distinct. The chlorovirus thaw pond communities se parated along the *x*-axis (57% of the variance explained) but overlapped along the *y*-axis (19% of the variance explained), while the Olsha community showed no clear pattern (Fig. 3D).

The same trends exhibited by the PCoA were evident in un weighted UniFrac UPGMA trees where myoviruses formed well-supported clades (bootstrap values = 100) by site (Fig. 4C) but did not cluster by depth. The results were less clear for the chloroviruses (Fig. 4D) and reflected the pattern of the PCoA. While SAS2A and BGR1 chlorovirus communities were distinct, Olsha and one SAS2A surface sample appeared to form a group, albeit poorly supported.

A heatmap (Fig. 4C,D) showed regional patterns in myovirus and chlorovirus OTU distribution, and pinpointed particular phylotypes in individual samples. For example, one replicate of SAS2A from the bottom consisted mainly of the myovirus OTU 126 that was practically absent in the other SAS2A samples. Similarly, the chlorovirus OTU 239 was dominant in one of the Olsha bottom replicates but was not detected elsewhere.

The ANOVA showed significant differences between the myovirus Shannon indices among all pond types ($p < 0.05$) (Fig. 5), while surface and bottom myovirus indices were not significantly different (see Table S2). While the SAS2A vs. BGR1 (thaw ponds) and Olsha vs. BGR1 chlorovirus indices were significantly different, we did not detect significant differences be tween SAS2A and Olsha. Although surface and bottom indices of Olsha chlorovirus were significantly different, we found no difference in diversity in the thaw pond surface and bottom samples. Myovirus richness ranged from 307 ± 86 phylotypes in the bottom of the SAS2A pond to 193 ± 17 phylotypes in the bottom of the BGR1 pond, while chlorovirus richness was highest in the BGR1 bottom sample (26 ± 1) and lowest in the Olsha bottom sample (7 ± 6).

Venn diagrams (Fig. 6) indicated that chloroviruses shared the highest percentage of OTUs (34.1%) among the 3 sites, while myovirus shared the least (5.7%). The highest percentage of unique OTUs for chloroviruses (19.5%) and myoviruses (37.6%) was in the SAS2A pond, while the highest percentage of chlorovirus (48.7%) and myovirus (14.2%) OTUs was shared between the 2 thermokarst ponds (BGR1 and SAS2A).

DISCUSSION

Consistent with observations from many studies elsewhere that viruses are an integral part of aquatic microbial ecosystems, we found yoviruses and chloroviruses in the water column of all 3 sites in this study. Viruses have been detected in a wide variety of cold-water environments, including polar lakes and seas (Lopez-Bueno et al. 2009, Chenard et al. 2015), and their abundance in thaw ponds was therefore to be expected. The presence of viral nucleic acids indicates that viral replication is taking place, which in turn suggests that viruses are active and likely influence the microbial ecology of northern thaw ponds.

We found that the 2 thermokarst ponds from different landscapes harboured distinct myovirus and chlorovirus communities and that landscape type is likely among the factors that drive these differences. Viral production and viral decay determine viral community composition and these processes are affected by biotic factors such as host metabolism and community composition, direct predation, and extracellular enzymatic activity (Chow & Suttle 2015), as well as by abiotic factors such as exposure to UV-B radiation and the concentration of charged particles (Suttle & Feng 1992, Suttle & Chan 1994). Our analyses were based on replicate samples within each waterbody but were limited to a single representative of each waterbody type. In terms of limnological and microbial characteristics, earlier work reported much higher variability between the rock-basin, palsa and lithalsa pond types compared to within them (Crevecoeur et al. 2015, Comte et al. 2016b). Further study is re quired from more ponds within each valley to resolve the role of landscape type in the differences in viral diversity observed in the present study.

DOC concentration may be the primary driver of viral host community structure and thereby viral diversity. Palsa ponds, such as our site SAS2A, are rich in biologically available DOC (Laurion & Mladenov 2013), which has substantial effects on the structure and function of aquatic microbial communities (Judd et al. 2006, Kritzberg et al. 2006). In concordance with our findings, studies of bacterial biogeography that included the same sites sampled in this study (Comte et al. 2016b) concluded that DOC was among the principal determinants of community composition, and that bacterial diversity in palsa and lithalsa ponds differed significantly. Crevecoeur et al. (2016) also found a significant difference between SAS2A and BGR1 archaeal community diversity, although only communities from bottom waters were targeted.

The diversity patterns of chlorovirus and putative hosts, which included all eukaryotes, were less synchronized than myoviruses and bacteria. This result is consistent with our use of the universal 18S primers capturing a large range of eukaryotes (including photosynthetic and non-photosynthetic organisms), compared to the chlorovirus primers that target a much more limited set of viruses, specifically phycodnaviruses that infect chlorophytes. This was also evident in Venn diagram patterns, with greater diversity of the eukaryotic community compared to chloroviruses. By limiting our study to chlorovirus primers, the overlap among the 3 pond communities could be an artifact of the small fraction of chloroviruses within the total virus community that infects eukaryotes. The Venn diagram based on only chlorophyte OTUs (Fig. 7) showed approximately twice the number of shared sequences among all pond types (21%) compared to the proportion of total eukaryotes (12%). It therefore seems likely that changes in chlorovirus diversity follow changes in chlorophyte diversity among the ponds. It should also be noted that the primer-based approach used to characterize microbial diversity in this study is subject to bias. For example, Needham et al. (2017) found that the T4super primer set tended to preferentially amplify smaller *g23* products and D'Amore et al. (2016) demonstrated that the library preparation method and the choice of sequencing platforms and primers influence the observed bacterial community composition based on 16S universal primers.

Another factor to consider is that the criteria used to define viral OTUs will affect the assessment of diversity. In our study, we chose a cut-off of 99% based on an alignment of the regions of the DNA polymerase amplified by the chlorovirus primers based on species in the family *Phycodnaviridae*. However, as was evident in the chlorovirus phylogeny (Fig. 9), all the thaw pond virotypes were closely related (although considered different OTUs according to our criteria). It remains unknown as to whether these OTUs were genetic variants of the same viral strain, or if they represent different strains of viruses with different host ranges, which would imply that the differences in community diversity were linked with function.

We expected to find different viral communities in thaw pond surface and bottom waters, but our analyses indicated that this was generally not the case. In contrast to Olsha, a rock-basin lake with a wellmixed water column, the thaw ponds were characterized by a stable, vertical physico-chemical gradient (Laurion et al. 2010) in the summer. Previous work has demonstrated that warm, oxygenated surface waters overlying cold, hypoxic/anoxic, bottom waters favour different microbial populations in the different depths (Rossi et al. 2013). Contrary to this expectation, there was substantial overlap in communities between surface and bottom layers (Fig. 5). One possible explanation for this result is that there was recent mixing of surface and bottom waters in the ponds, a hypothesis supported by the fact that aerobic bacterial taxa (e.g. *Comamonadaceae*, *Polynucleobacter*) were detected in bottom samples, suggesting transfer by mixing or sedimentation.

The classification of viruses based on environmental sequences is challenging because of the paucity of relevant genomes in the reference database and the mutability and mosaicism of viral genomes. The classification of thaw pond OTUs from BLAST-based analyses was, with a few exceptions, limited to the taxonomic level of family. For example, 97% of the myovirus OTUs could only be classified in the group 'uncultured *Myoviridae*', providing no real insight into the potential hosts of these viruses, albeit strongly suggesting that thaw pond myoviruses were divergent from other myoviruses. Of the 3% of OTUs that were classified to the species level, most of them were grouped with myoviruses that infect humans or marine organisms that are unlikely to be present in a freshwater thaw pond. Similar to the myoviruses, most thaw pond chlorovirus OTUs (93%) could not be classified past the family level and were categorized as unknown members of the family *Phycodnaviridae*. In an effort to gain a better taxonomic resolution, we constructed phylogenetic trees with thaw pond chlorovirus OTUs and representative genomes from the genera of the family *Phycodnaviridae*. All of the OTUs from this study clustered most closely with the chlorovirus species *Acanthocystis turfacea Chlorella virus 1*, a

virus that infects a chlorophyte symbiont of a heliozoan. However, it is likely that the hosts of these viruses are the abundant free-living chlorophytes that comprised a large percentage of the total eukaryotes.

In conclusion, this first viral study of thaw ponds, a major ecosystem type across the northern landscape of Canada, showed that myo- and chlorovirioplankton were present in both palsa and lithalsa waters, as well as in a shallow rock-basin reference lake in subarctic Québec. Although the thaw pond viral OTUs were re lated to viruses in the families *Myoviridae* and *Phycodnaviridae*, they were divergent from known viruses within these groups. There was a pronounced difference in viral community composition between different thaw ponds, raising the possibility that permafrost landscape type selects the bacterial and eukaryotic host communities, which in turn drives viral structure in these microbial ecosystems. Additional ponds from each landscape type will be needed to further evaluate this conjecture.

Acknowledgements

We acknowledge the Natural Sciences and Engineering Research Council of Canada (NSERC), the Northern Scientific Training Program (NSTP), the Fond de Recherche du Québec − Nature et Technologies (FRQNT) and the Network of Centres of Excellence ArcticNet for their financial support. We are also grateful to Claude Tremblay, manager of the Whapmagoostui-Kuujjuarapik CEN station, Vani Mohit, Bethany Deshpande and Alex Matveev for their help with fieldwork, and Amanda Toperoff for editing the figures.

References

Adriaenssens EM, Cowan DA (2014) Using signature genes as tools to assess environmental viral ecology and diver- sity. Appl Environ Microbiol 80:4470–4480

Anesio AM, Mindl B, Laybourn-Parry J, Hodson AJ, Sattler B (2007) Viral dynamics in cryoconite holes on a high Arctic glacier (Svalbard). J Geophys Res Biogeosci 112: G04S31

Bégin PN, Vincent WF (2017) Permafrost thaw lakes and ponds as habitats for abundant rotifer populations. Arctic Sci 3:354−377

Berger SA, Krompass D, Stamatakis A (2011) Performance, accuracy, and web server for evolutionary placement of short sequence reads under maximum likelihood. Syst Biol 60:291−302

Bhiry N, Delwaide A, Allard M, Begin Y and others (2011) Environmental change in the Great Whale River region, Hudson Bay: five decades of multidisciplinary research by Centre d'études nordiques (CEN). Ecoscience 18: 182−203

Breitbart M, Thompson LR, Suttle CA, Sullivan MB (2007) Exploring the vast diversity of marine viruses. Oceanography (Wash DC) 20:135−139

Brussaard CPD (2004) Viral control of phytoplankton populations — a review. J Eukaryot Microbiol 51:125−138

Buchfink B, Xie C, Huson DH (2015) Fast and sensitive pro- tein alignment using DIAMOND. Nat Methods 12:59−60 Calmels F, Allard M, Delisle G (2008) Development and decay of a lithalsa in Northern Quebec: a geomorphological history. Geomorphology 97:287−299

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K and others (2010) QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7:335−336

Chen H, Boutros PC (2011) VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinformatics 12:35

Chenard C, Chan AM, Vincent WF, Suttle CA (2015) Polar freshwater cyanophage S-EIV1 represents a new wide-spread evolutionary lineage of phages. ISME J 9: 2046−2058

Chow CET, Fuhrman JA (2012) Seasonality and monthly dynamics of marine myovirus communities. Environ Microbiol 14:2171−2183

Chow CET, Suttle CA (2015) Biogeography of viruses in the sea. Annu Rev Virol 2:41−66

Colangelo-Lillis J, Wing BA, Whyte LG (2016) Low viral predation pressure in cold hypersaline Arctic sediments and limits on lytic replication. Environ Microbiol Rep 8: 250−260

Comeau AM, Li WKW, Tremblay JE, Carmack EC, Lovejoy C (2011) Arctic Ocean microbial community structure before and after the 2007 record sea ice minimum. PLOS ONE 6:e27492

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

Comte J, Lovejoy C, Crevecoeur S, Vincent WF (2016a) Co-occurrence patterns in aquatic bacterial communities across changing permafrost landscapes. Biogeosciences 13:175−190

Comte J, Monier A, Crevecoeur S, Lovejoy C, Vincent WF (2016b) Microbial biogeography of permafrost thaw ponds across the changing northern landscape. Eco- graphy 39:609−618

Crevecoeur S, Vincent WF, Comte J, Lovejoy C (2015) Bac- terial community structure across environmental gradients in permafrost thaw ponds: methanotroph-rich eco- systems. Front Microbiol 6:192

Crevecoeur S, Vincent WF, Lovejoy C (2016) Environmental selection of planktonic methanogens in permafrost thaw ponds. Sci Rep 6:31312

Crevecoeur S, Vincent WF, Comte J, Matveev A, Lovejoy C (2017) Diversity and potential activity of methanotrophs in high methane-emitting permafrost thaw ponds. PLOS ONE 12:e0188223

Culley AI, Steward GF (2007) New genera of RNA viruses in subtropical seawater, inferred from polymerase gene sequences. Appl Environ Microbiol 73:5937−5944

D'Amore R, Ijaz UZ, Schirmer M, Kenny JG and others (2016) A comprehensive benchmarking study of protocols and sequencing platforms for 16S rRNA community profiling. BMC Genomics 17:55

de Carcer DA, Pedros-Alio C, Pearce DA, Alcami A (2016) Composition and interactions among bacterial, micro-eukaryotic, and T4-like viral assemblages in lakes from both polar zones. Front Microbiol 7:337

Deshpande BN, MacIntyre S, Matveev A, Vincent WF (2015) Oxygen dynamics in permafrost thaw lakes: anaerobic bioreactors in the Canadian subarctic. Limnol Oceanogr 60:1656−1670

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792−1797

Edgar RC (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat Methods 10:996−998

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27:2194−2200

Gao EB, Huang YH, Ning DG (2016) Metabolic genes within cyanophage genomes: implications for diversity and evolution. Genes (Basel) 7:80

Guixa-Boixereu N, Vaque D, Gasol JM, Sanchez-Camara J, Pedros-Alio C (2002) Viral distribution and activity in Antarctic waters. Deep Sea Res II 49:827−845

Gurney SD (2001) Aspects of the genesis, geomorphology and terminology of palsas: perennial cryogenic mounds. Prog Phys Geogr 25:249−260

Huson DH, Beier S, Flade I, Gorska A and others (2016) MEGAN Community Edition — interactive exploration and analysis of large-scale microbiome sequencing data. PLOS Comput Biol 12:e1004957

Joli N, Monier A, Logares R, Lovejoy C (2017) Seasonal pat- terns in Arctic prasinophytes and inferred ecology of Bathycoccus unveiled in an Arctic winter metagenome. ISME J 11:1372−1385

Judd KE, Crump BC, Kling GW (2006) Variation in dissolved organic matter controls bacterial production and community composition. Ecology 87:2068−2079

Kiikkila O, Smolander A, Ukonmaanaho L (2014) Properties of dissolved organic matter in peatland: implications for water quality after harvest. Vadose Zone J 13:vzj2013. 8.0155

Kritzberg ES, Langenheder S, Lindstrom ES (2006) Influence of dissolved organic matter source on lake bacterio-plankton structure and function — implications for seasonal dynamics of community composition. FEMS Micro- biol Ecol 56:406−417

Laurion I, Mladenov N (2013) Dissolved organic matter photolysis in Canadian arctic thaw ponds. Environ Res Lett 8:035026

Laurion I, Vincent W, 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

Lopez-Bueno A, Tamames J, Velazquez D, Moya A, Que- sada A, Alcami A (2009) High diversity of the viral com- munity from an Antarctic lake. Science 326:858−861

Lovejoy C, Comeau A, Thaler M (2016) Curated reference database of SSU rRNA for northern marine and freshwater communities of Archaea, Bacteria and microbial eukaryotes, v. 1.1 (2002-2008). Nordicana D 23

Matteson AR, Loar SN, Bourbonniere RA, Wilhelm SW (2011) Molecular nnumeration of an ecologically impor- tant cyanophage in a Laurentian Great Lake. Appl Environ Microbiol 77:6772−6779

Matteson AR, Rowe JM, Ponsero AJ, Pimentel TM, Boyd PW, Wilhelm SW (2013) High abundances of cyano-myoviruses in marine ecosystems demonstrate ecological relevance. FEMS Microbiol Ecol 84:223−234

Matveev A, Laurion I, Deshpande BN, Bhiry N, Vincent WF (2016) High methane emissions from thermokarst lakes in subarctic peatlands. Limnol Oceanogr 61: S150−S164

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

Needham DM, Sachdeva R, Fuhrman JA (2017) Ecological dynamics and co-occurrence among marine phytoplankton, bacteria and myoviruses shows microdiversity mat- ters. ISME J 11:1614−1629

Oksanen J, Blanchet FG, Kindt R, Legendre P and others (2016) Vegan: community ecology package. R package version 2.3-5

Olefeldt D, Goswami S, Grosse G, Hayes D and others (2016) Circumpolar distribution and carbon storage of thermo- karst landscapes. Nat Commun 7:13043

Pienitz R, Doran PT, Lamoureux SF (2008) Origin and geomorphology of lakes in the polar regions. In: Vincent WF, Laybourn-Parry J (eds) Polar lakes and rivers: limnology of Arctic and Antarctic aquatic ecosystems. Oxford University press, Oxford, p 25–41

Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig WG, Peplies J, Glockner FO (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res 35:7188−7196

Przytulska A, Comte J, Crevecoeur S, Lovejoy C, Laurion I, Vincent WF (2016) Phototrophic pigment diversity and picophytoplankton in permafrost thaw lakes. Biogeo-sciences 13:13−26

Przytulska A, Bartosiewicz M, Vincent WF (2017) Increased risk of harmful cyanobacterial blooms in northern high-latitude lakes through climate warming and enrichment. Freshw Biol 62:1986−1996

R Core Team (2016) R: a language and environment for sta- tistical computing. R Foundation for Statistical Computing, Vienna. www.r-project.org

Rohwer F, Thurber RV (2009) Viruses manipulate the marine environment. Nature 459:207−212

Rossi PG, Laurion I, Lovejoy C (2013) Distribution and identity of Bacteria in subarctic permafrost thaw ponds. Aquat Microb Ecol 69:231−245

Short SM (2012) The ecology of viruses that infect eukaryotic algae. Environ Microbiol 14:2253−2271

Short SM, Rusanova O, Staniewski MA (2011) Novel phycodnavirus genes amplified from Canadian freshwater environments. Aquat Microb Ecol 63:61−67

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312−1313

Suttle CA (2005) Viruses in the sea. Nature 43:356–361

Suttle CA, Chan AM (1994) Dynamics and distribution of cyanophages and their effect on marine Synechococcus spp. Appl Environ Microbiol 60:3167−3174

Suttle CA, Feng C (1992) Mechanisms and rates of decay of marine viruses in seawater. Appl Environ Microbiol 58: 3721−3729

Vincent WF, Lemay M, Allard M (2017) Arctic permafrost landscapes in transition: towards an integrated Earth system approach. Arctic Sci 3:39−64

Watanabe S, Laurion I, Chokmani K, Pienitz R, Vincent WF (2011) Optical diversity of thaw ponds in discontinuous permafrost: a model system for water color analysis. J Geophys Res Biogeosci 116:G02003

Wells LE, Deming JW (2006) Modelled and measured dynamics of viruses in Arctic winter sea-ice brines. Environ Microbiol 8:1115−1121

Wilhelm SW, Matteson AR (2008) Freshwater and marine virioplankton: a brief overview of commonalities and differences. Freshw Biol 53:1076−1089

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

Zeigler Allen L, McCrow JP, Ininbergs K, Dupont CL and others (2017) The Baltic Sea virome: diversity and tran- scriptional activity of DNA and RNA viruses. mSystems 2:e00125-16

Table 1. Sets of primers used for viral amplification. Degenerate nucleotides code: R = A, G; Y = C, T; M = A, C; K = G, T; S = G, T; W = A, T; H = A, C, T; B = C, G, T; V = A, C, G

Virus	Gene	Primer	Sequence (5′→3′)	Amplicon length (bp)	Reference
Myovirus	*gp23*	T4superF1	GAY HTI KSI GGI GTI CAR CCI ATG	400–500	Chow & Fuhrman (2012)
		T4superR1	GCI YKI ARR TCY TGI GCI ARY TC		
Chlorovirus	*polB*	CHLVdF	CCW ATC GCA GCW CTM GAT TTT G	560–575	Short et al. (2011)
		CHLVdR	ATC TCV CCB GCV ARC CAC TT		

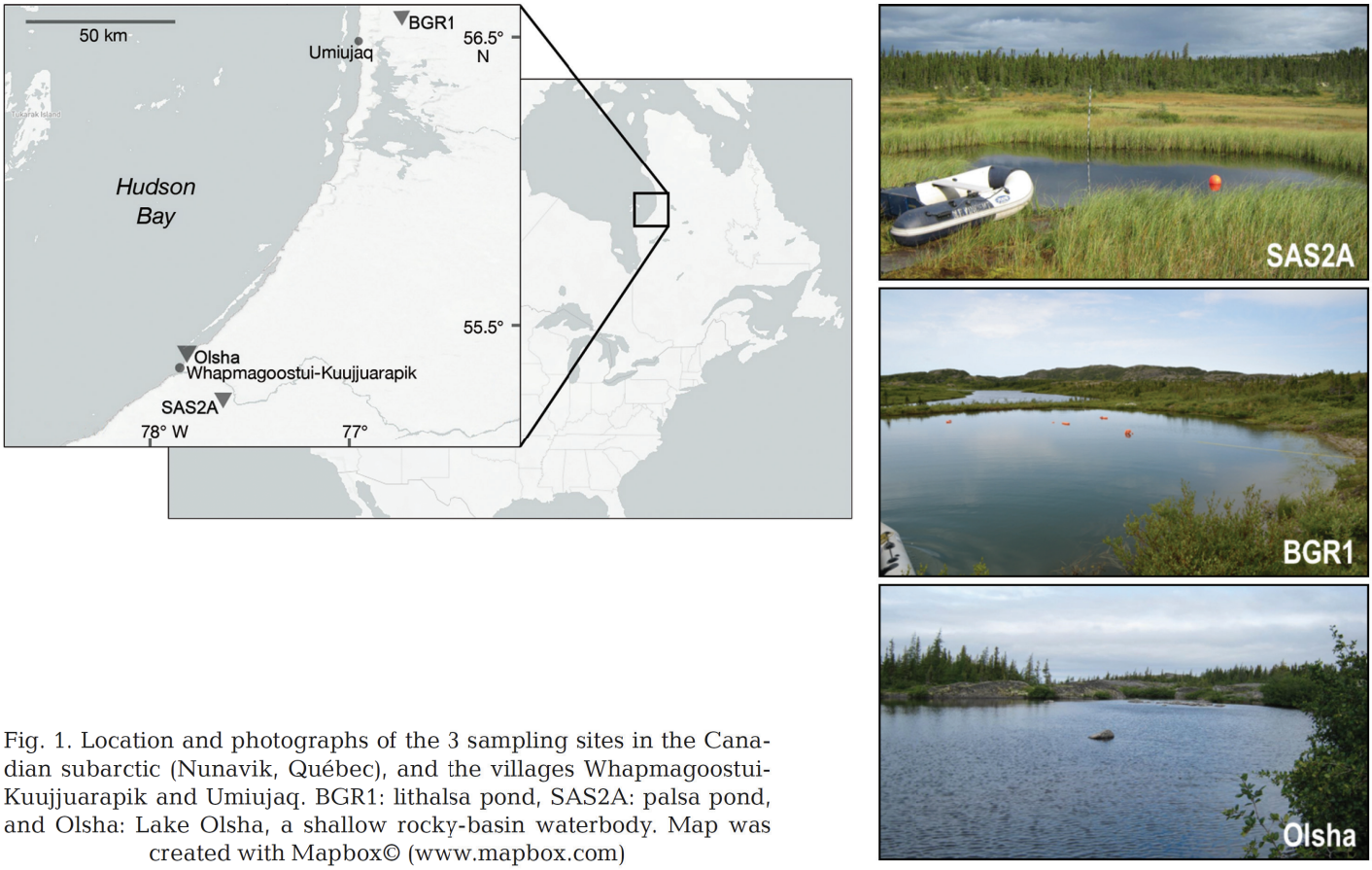

Fig. 1. Location and photographs of the 3 sampling sites in the Canadian subarctic (Nunavik, Québec), and the villages Whapmagoostui-Kuujjuarapik and Umiujaq. BGR1: lithalsa pond, SAS2A: palsa pond, and Olsha: Lake Olsha, a shallow rocky-basin waterbody. Map was created with Mapbox© (www.mapbox.com)

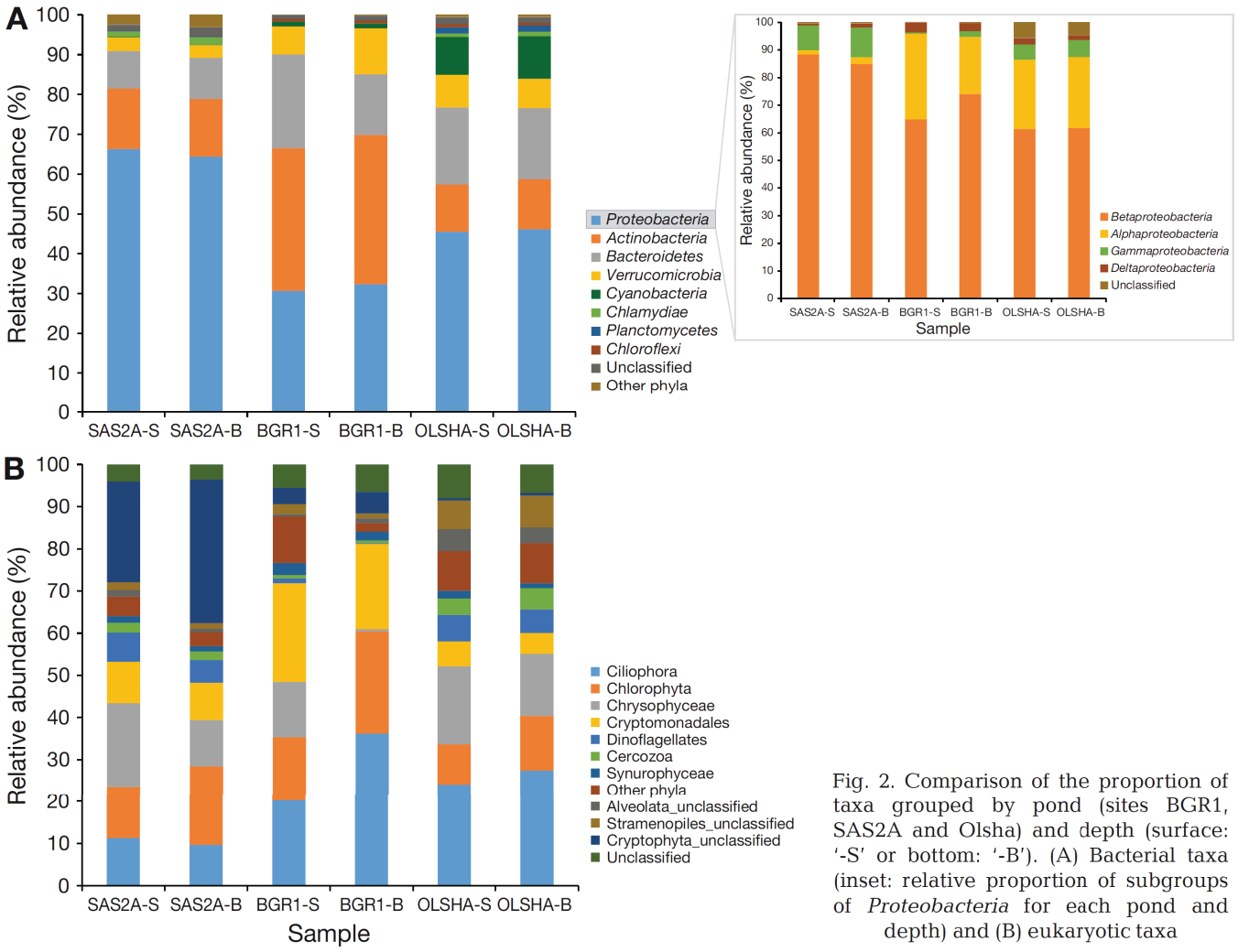

Fig. 2. Comparison of the proportion of taxa grouped by pond (sites BGR1, SAS2A and Olsha) and depth (surface: '-S' or bottom: '-B'). (A) Bacterial taxa (inset: relative proportion of subgroups of *Proteobacteria* for each pond and depth) and (B) eukaryotic taxa

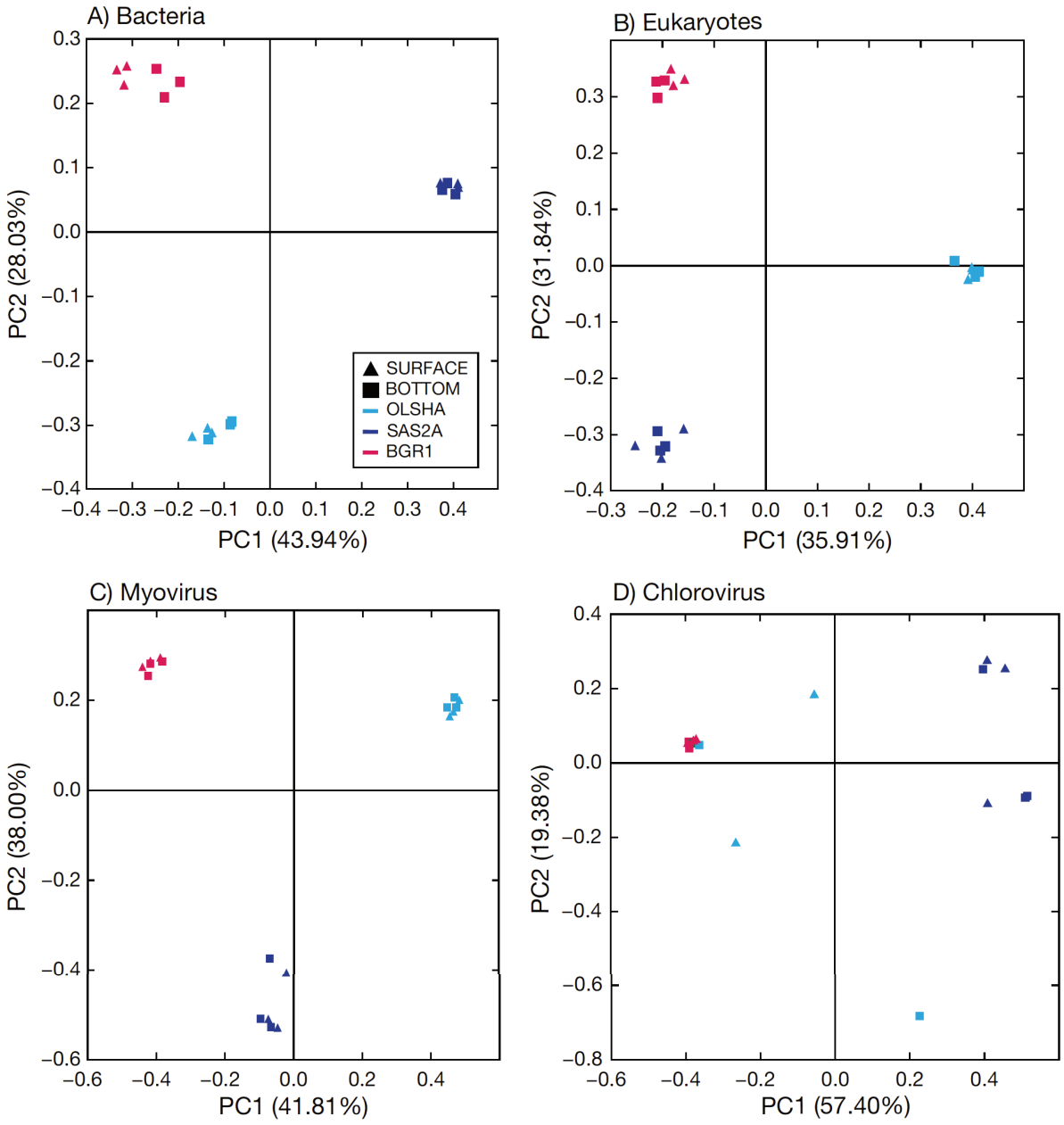

Fig. 3. Principal coordinate analysis generated with unweighted UniFrac distances, for (A) bacterial, (B) eukaryotic, (C) myovirus and (D) chlorovirus communities, by pond type (sites BGR1, SAS2A and Olsha) and depth. Points that are closer on the ordination have communities that are more similar

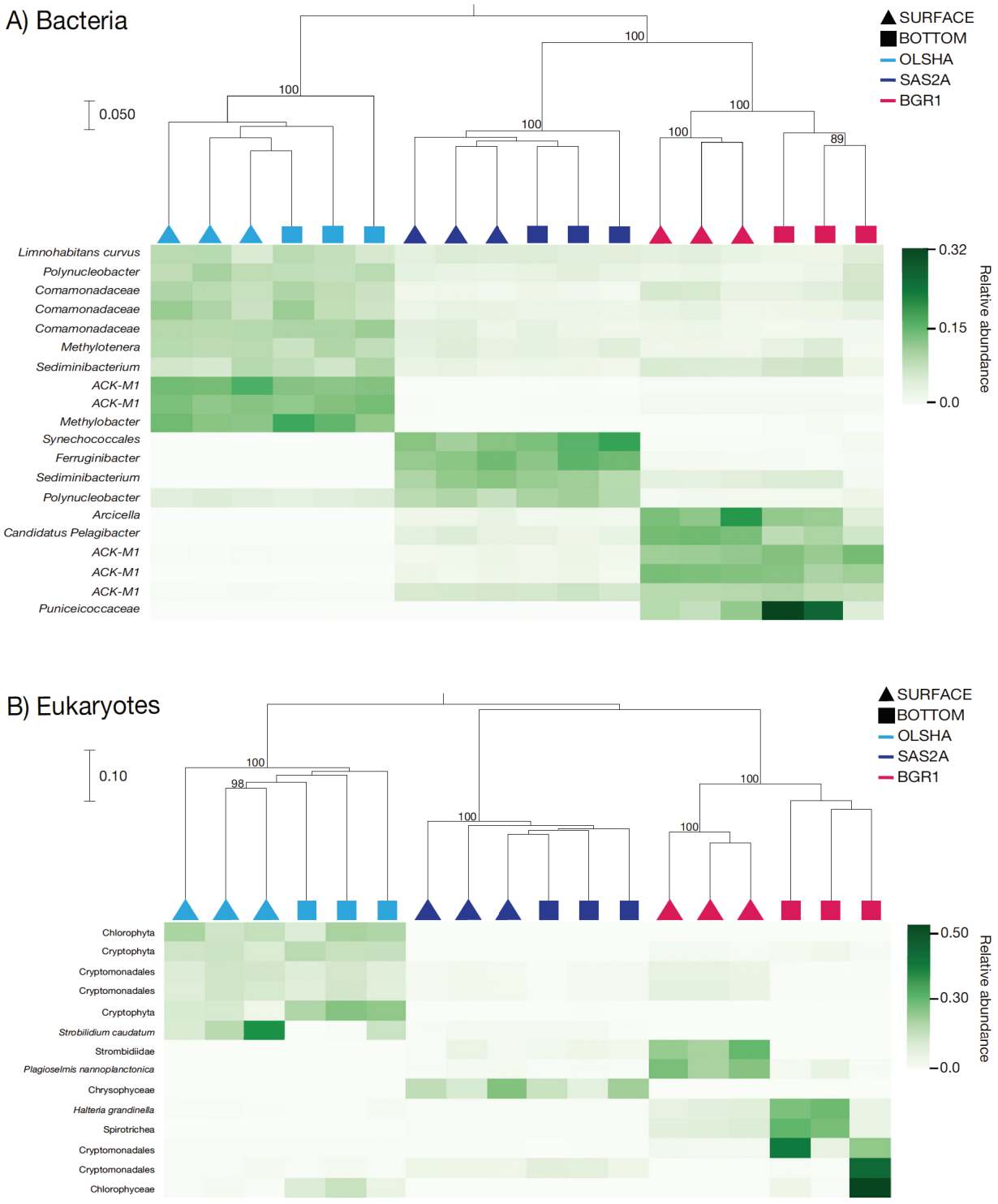

Fig. 4. (Above and following page.) Unweighted UniFrac UPGMA clustering of the samples combined with a heatmap (below the tree) representing the proportion of sequences of the most abundant total OTUs (>1% of the dataset), for (A) bacterial, (B) eukaryotic, (C) myovirus and (D) chlorovirus communities, by pond type (sites BGR1, SAS2A and Olsha) and depth. The shade of green in the heatmap corresponds to the relative abundance of the taxa (bacteria and eukaryotes) or OTU (viruses) shown on the left side of the tree. Bootstrap values > 50 are shown

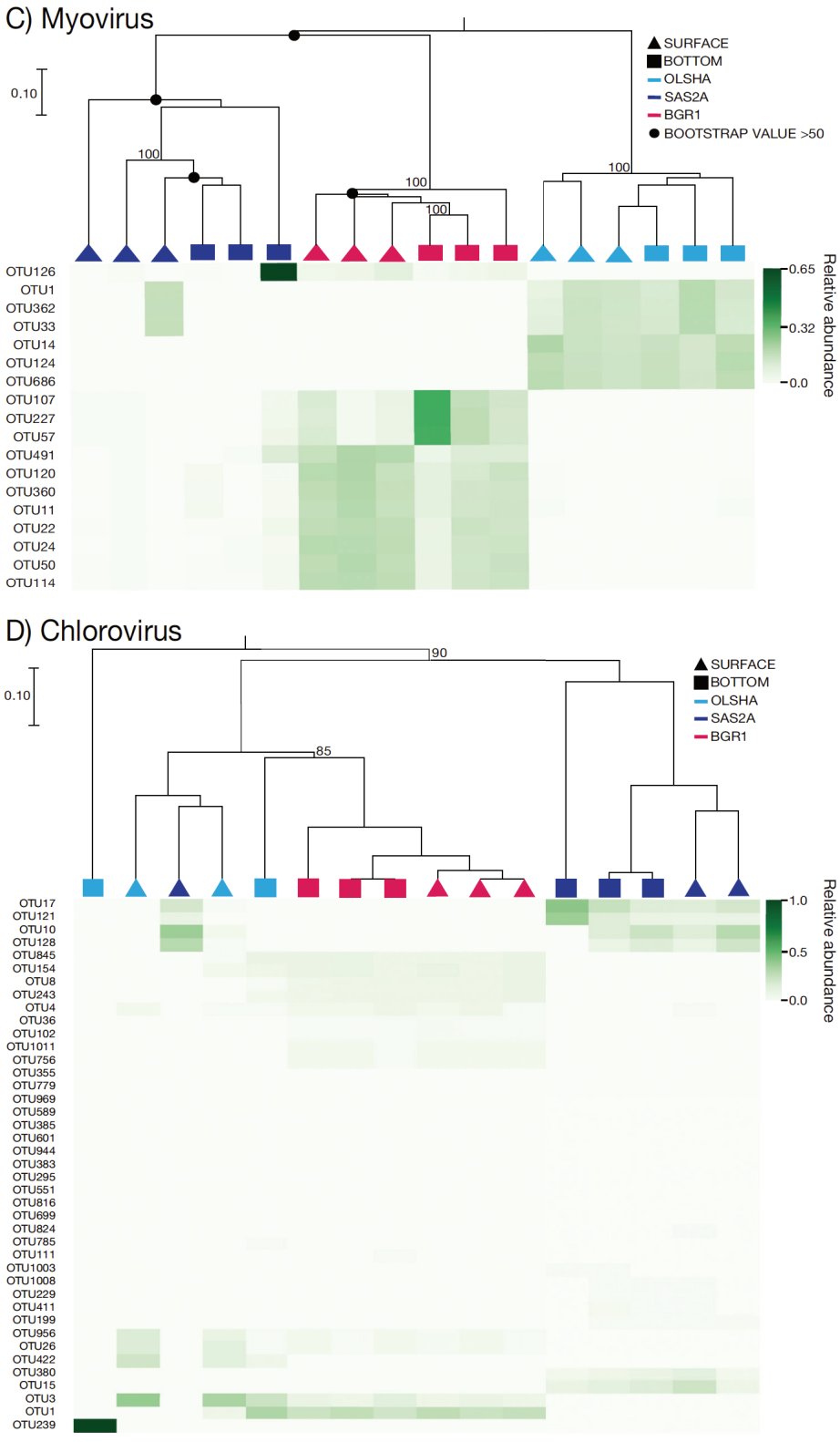

Fig. 4. (continued)

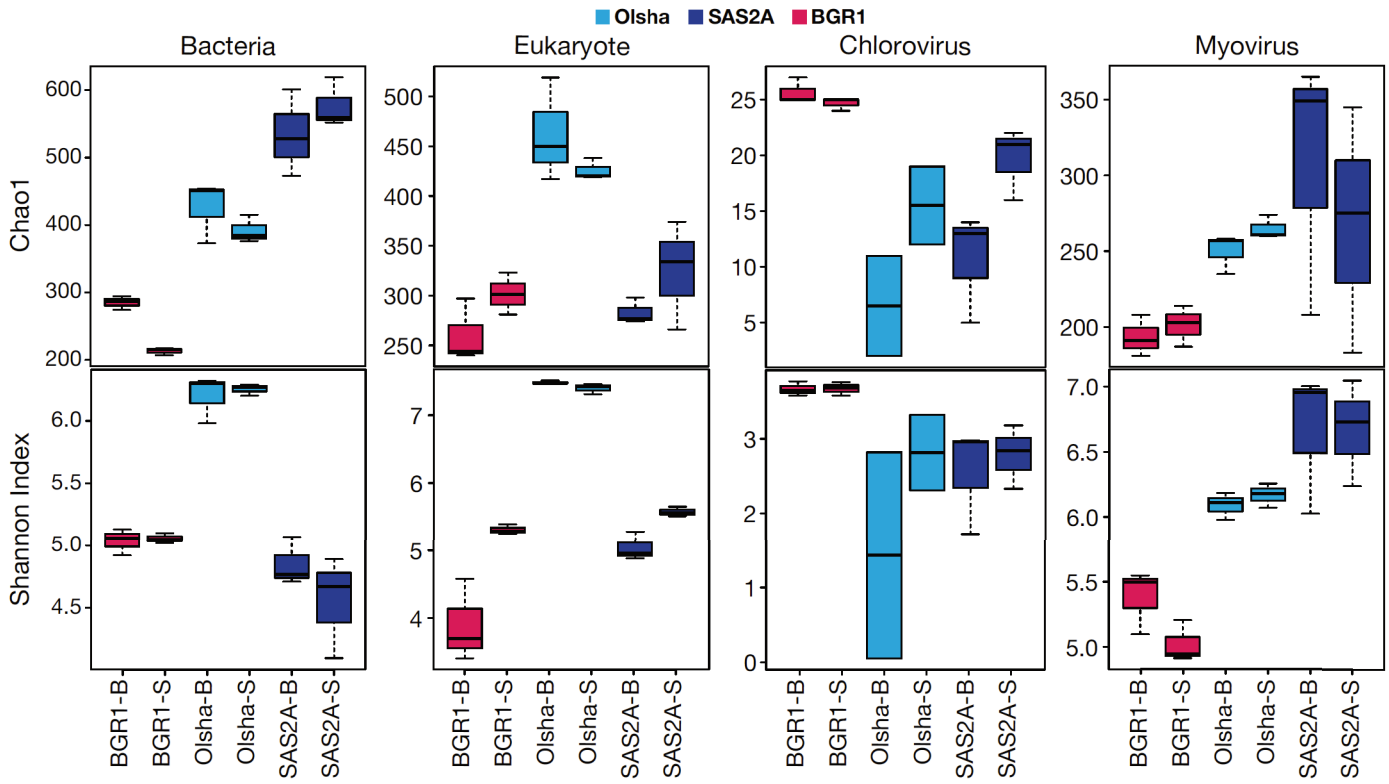

Fig. 5. Alpha-diversity indices for the bacterial, eukaryotic, myovirus and chlorovirus communities, by pond type (sites SAS2A, BGR1 and Olsha) and depth (surface: '-S' or bottom: '-B'). Richness and Shannon indices shown on the left for Chao1 and Shannon plots, respectively (note the different *y*-axis scales). Ends of the box: upper and lower quartiles; line within the box: median; and extent of upper and lower whiskers: the largest and smallest values

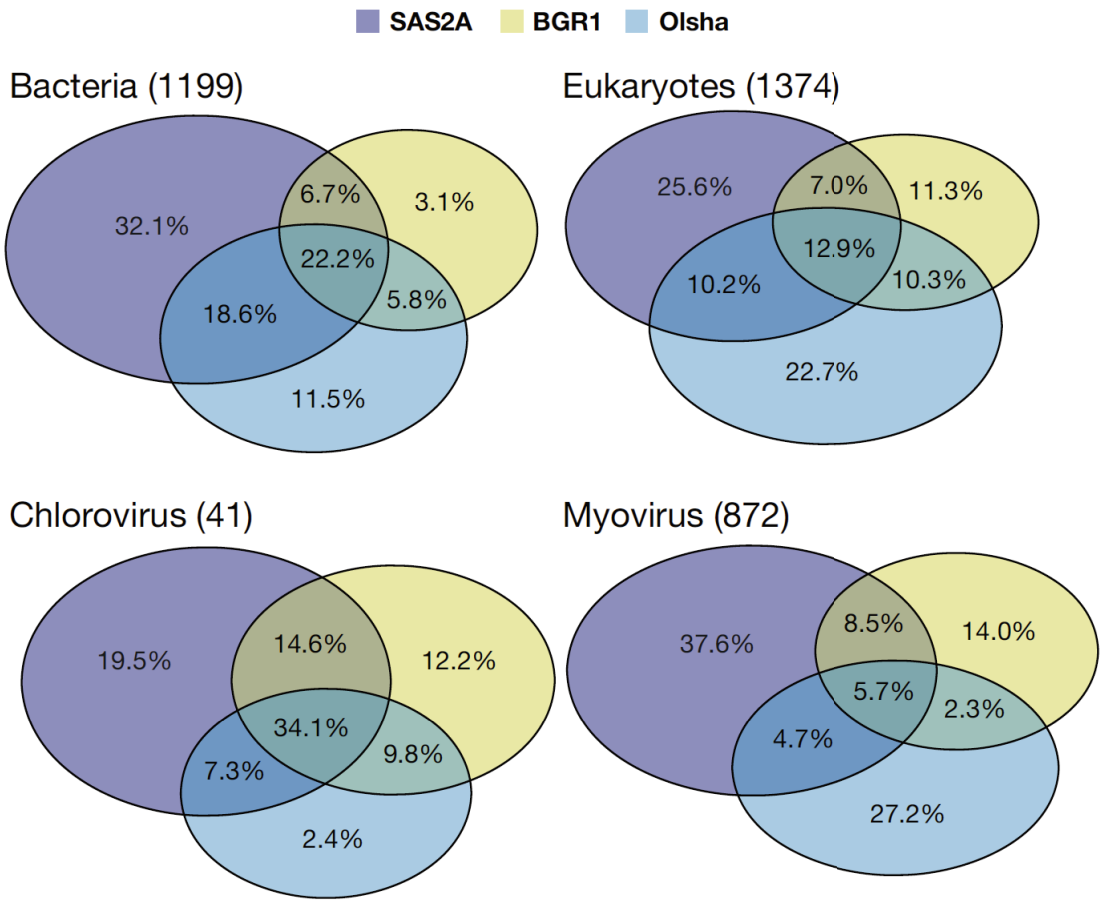

Fig. 6. Proportion of shared and unique OTUs among ponds (sites SAS2A, BGR1 and Olsha) for the bacterial, eukaryotic, myovirus and chlorovirus communities. Total number of OTUs for each group shown in parentheses

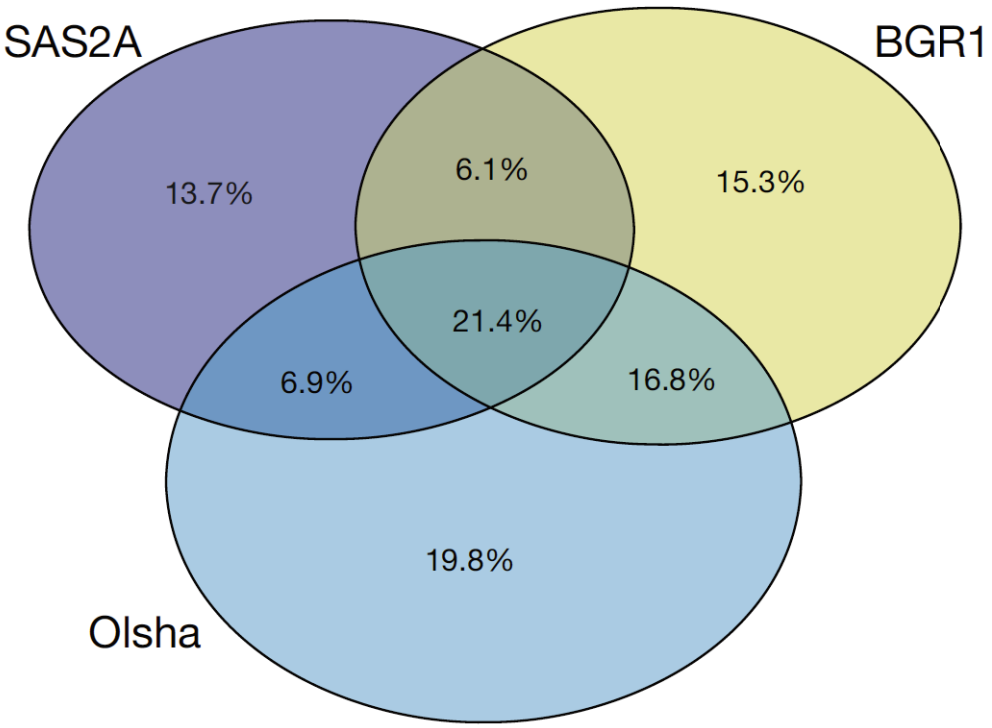

Fig. 7. Proportion of shared and unique chlorophyte OTUs among the 3 sites (SAS2A, BGR1 and Olsha)

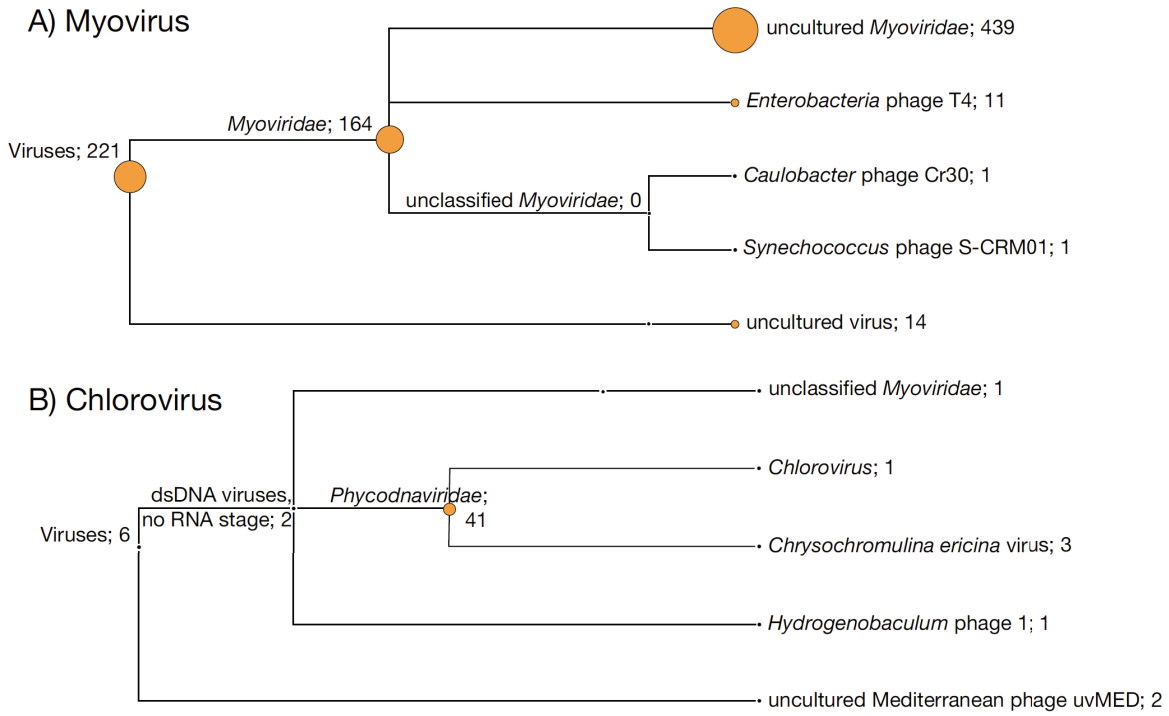

Fig. 8. Classification of the combined OTUs for (A) myovirus and (B) chlorovirus from all 3 ponds (sites SAS2A, BGR1 and Olsha) generated with MEGAN. Number of OTUs in each category is listed after the branch or node name and is reflected in the area of the circle at the corresponding node

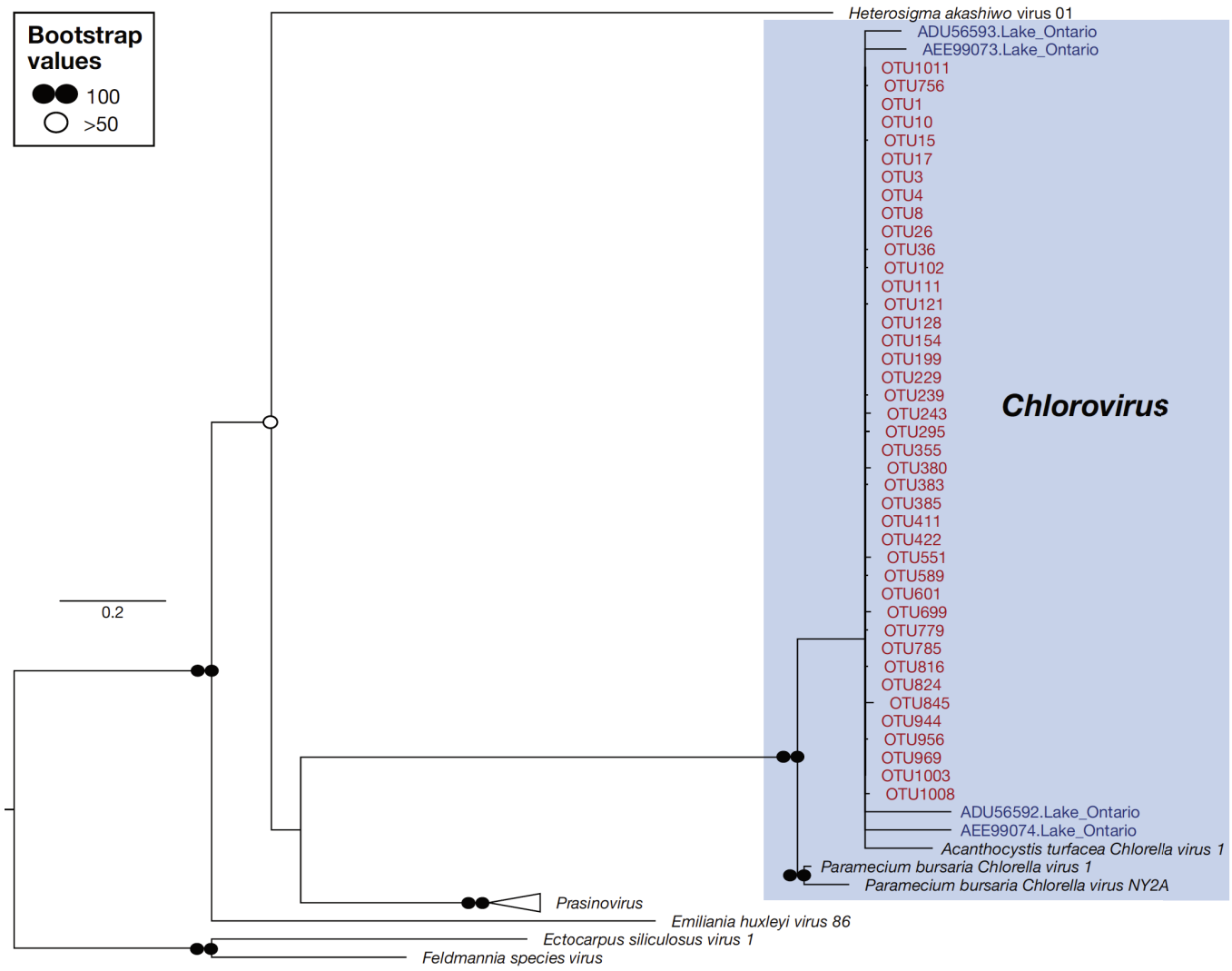

Fig. 9. Chlorovirus phylogenetic tree using the evolutionary placement algorithm in RAxML that includes all the OTUs from this study. Reference tree constructed based on an alignment of complete DNA polymerase sequences from representative viruses from the established taxa in the family *Phycodnaviridae*, as well as sequences amplified from Lake Ontario (Short et al. 2011). Blue shaded box: clade corresponding to the genus *Chlorovirus*. Red font: OTUs from the present study. Blue font: environment viral OTUs from Lake Ontario (Short et al. 2011) Scale bar: 0.2 substitutions site^{-1}