Dynamics in coastal RNA viruses and bacteriophages are driven by shifts in the community phylogenetic structure

Julia A. Gustavsen[1†] and Curtis A. Suttle[1,2,3,4*]

[1] Department of Earth, Ocean and Atmospheric Sciences, University of British Columbia, Vancouver, B.C., Canada

[2] Department of Botany, University of British Columbia, Vancouver, B.C., Canada

[3] Department of Microbiology & Immunology, University of British Columbia, Vancouver, B.C., Canada

[4] Institute for the Oceans and Fisheries, University of British Columbia, Vancouver, B.C., Canada

[†] Present address: Sophia Genetics, Rue du Centre 172, 1025 Saint-Sulpice, V.D., Switzerland

*Correspondence: Curtis A. Suttle, Department of Earth, Ocean & Atmospheric Sciences, The University of British Columbia, 2020 - 2207 Main Mall, Vancouver, B.C., V6T 1Z4, Canada suttle@science.ubc.ca

keywords: virus / phylogeny / coastal / time series / bacteria / phytoplankton / killing the winner / seed bank / 18S / 16S / dynamics / quasispecies / *Picornavirales* / *Myoviridae*

Section: microbial ecology

Running title: Phylogenetic shifts in dominant viral communities

Abstract

Marine microbes including viruses are an essential part of the marine ecosystem that forms the base of the foodweb, and drives biogeochemical cycles. Marine viral communities display repeatable changes in abundance and community composition throughout time; however, whether these changes reflect shifts in dominance within evolutionarily related groups of viruses and their hosts is unexplored. To examine these dynamics, changes in the composition and phylogenetic makeup of two ecologically important groups of viruses, and their potential hosts, were sampled every two weeks for 13 months at a coastal site in British Columbia, Canada. Changes in the taxonomic composition within DNA bacteriophages related to T4-like viruses and marnavirus-like RNA viruses infecting eukaryotic phytoplankton, as well as bacteria and eukaryotes, were examined using amplicon sequencing of gene fragments encoding the major capsid protein (*gp23*), the RNA-dependent RNA polymerase (*RdRp*) and the 16S and 18S ribosomes, respectively. The results showed that for both viral marker genes, the dominant groups of phylogenetically-related viruses shifted over time and contained many transient taxa and few persistent taxa; yet, different community structures were observed in these different viral communities. Additionally, with strong lagged correlations between viral richness and community similarity of putative hosts, the results imply that viruses influence the composition of the host communities.

Importance

Using high-throughput sequencing of coastal seawater collected every two weeks for one year, the dynamics of two groups of ecologically important groups of viruses were described in the context of their putative hosts and the environment. There was a large diversity of viruses and putative hosts in this study, and groups of phylogenetically-related viruses showed temporal dynamics in dominance. Examining the richness of viruses by phylogenetic groups showed different dynamics of either boom-bust or continued persistence. At the OTU-level, some members of these related groups persisted throughout time, while others were more ephemeral. These findings were put in context of potential quasispecies behaviour, and the dynamics of putative hosts. These results showed that temporal dynamics of viral communities have a phylogenetic signal which is important for understanding the ecology of these viruses since it elucidated one of the drivers of the community structure.

Introduction

Understanding diversity, its maintenance, and its drivers is a continued theme in ecology. In microbial systems there has been extensive exploration and discussion about the mechanisms responsible for the observed high diversity (1). Many studies on microbial diversity and dynamics come from the marine milieu, where it has been argued that community composition is driven by environmental factors (2–5). Against this backdrop are viruses, which are obligate parasites that are the most abundant biological entities in the world's oceans, and account for much of its diversity (6).

This high viral diversity arises since viruses have many different lifestyles (7), morphologies (8), and infection strategies. Some viruses infect specific strains or species of hosts, whereas others have broad host-ranges (9). As well, some groups of viruses show particularly high genetic diversity because of their low fidelity of replication (10), while others have high rates of horizontal gene transfer (11). The role of viruses as obligate pathogens, often with high host specificity, implies that they are important drivers of host composition and diversity (12); yet, our understanding of their roles as drivers of marine microbial diversity remains relatively unexplored.

Marine viruses have repeatable seasonal dynamics as revealed by measures of abundance, infectious units, and taxonomic composition. Some seasonal studies in coastal waters report that viral abundances are higher in summer than in winter (13, 14), while other multi-year time series data show that viral production and viral abundances are highest in early spring and summer (15, 16). Moreover, viral dynamics can be associated with putative hosts (16) and specific subsets of the overall coastal viral communities can show seasonal community composition dynamics (17, 18). As well, viruses infecting cyanobacteria show temporal dynamics (19, 20), with communities from the same season resembling each other more than communities sampled in the same year (21, 22), and communities being more stable in winter than in the summer and spring (23).

Viruses affect community composition in laboratory studies by reducing the abundance of the dominant host, thereby allowing others to grow up (24–26); thus, viruses promote diversity among hosts (even at the strain level) and can be responsible for large shifts in the dominant species in bacterial populations (12, 27) . These dynamics have been termed

"Killing the Winner" (KtW), a model in which the most actively growing hosts are killed by viruses and replaced by other strains or species (28, 29) and that co-evolution allows these dynamics to continue over time (30). There is evidence of KtW dynamics in field studies, as illustrated by a study on a solar saltern, in which coarsely-defined bacterial and viral taxa (akin to genus level grouping) were relatively constant over time, but showed KtW dynamics at a finer taxonomic scale (31). Common members of the myoviral community showed greater microdiversity over time and correlations to hosts were stronger when microdiversity within OTUs was examined in a time series in coastal California to revealing potential strain-specific effects of viruses on hosts (32). Using complete viral metagenomes recovered from a freshwater lake, Arkhipova *et al* (33) examined the dynamics of viruses over a year and found peaks in viral relative abundance before, during and after peaks in host abundance. Thus there can be interactions that do not follow the Killing the Winner model and more studies illuminating these interactions are needed as, so far, few environmental studies have examined potential Killing the Winner dynamics since few have compared hosts and viruses (34).

Examining the temporal dynamics of marine viruses and their hosts has yielded insights about the ecology of these viruses, yet little attention has been paid to the phylogenetic relationships within these communities and how they are shaped. An exception is a study by Goldsmith *et al* (35), near Bermuda, where the phylogenetic makeup of related groups of viruses over time and depth was found to be highly uneven and variable. There were differences between fall and winter attributable to stratification, with much of the variability due to one phylogenetic group of cyanophages (35). In coastal California, groups of cyanophages belonging to different phylogenetic clades shifted in their relative

dominance over time (23). Knowing more about the phylogenetic diversity of the viral communities will allow us to better interpret these temporal dynamics.

Phylogenetic relatedness can be correlated to ecological relatedness in plants and animals (36, 37) and microbes have shown phylogenetic patterning in distribution and abundance (38, 39), yet little is known about these patterns in viral communities. To examine these phylogenetic patterns over time, the following hypotheses were tested: First, it was hypothesized that phylogenetic patterns in abundance would be detected in the viral communities, as has been found in putative host communities, since viruses can be driving Killing the Winner dynamics (28) or be responding to their hosts (33, 40). Second, the structure of viral communities has been purported to follow a "seed bank" distribution, where there are many more rare viral operational taxonomic units (OTUs) than abundant ones (35, 41); therefore, the ranking within phylogenetically-related viral OTUs could also follow this pattern over time. Examining the temporal relationship between phylogeny and relative abundance will reveal if genetic relatedness influences dominance in viral and, potentially, in host communities. Uncovering the alpha diversity over time in communities and how it relates to other communities and to the environment will illuminate the drivers of community structure and diversity.

To test these hypotheses, the temporal dynamics of the phylogenetic make-up of two ecologically important groups of marine viruses and their potential hosts were followed in samples taken every two weeks over thirteen months, using amplified marker genes and high-throughput sequencing. The community similarities, richness, phylogenetic diversity, and the relative abundance of phylogenetically-related groups of OTUs were examined over

time in viral and putative host communities. Two groups of viruses were selected and examined because they have shown evidence of continued production in coastal waters and high diversity of viral genotypes in this environment which would be helpful in the phylogenetic patterns related to community dynamics and their diversity over time. The first group was T4-like viruses, DNA viruses that infect bacteria, including cyanobacteria. The amplification target used was *gp23*, the gene encoding the capsid (42). The second group of viruses were marnavirus-like viruses (MLVs), marine RNA viruses in the order *Picornavirales* that infect eukaryotic phytoplankton and possibly heterotrophic protists; they were targeted by amplifying part of the RNA dependent RNA polymerase (*RdRp*) gene (43). These viruses infect ecologically important phytoplankton, such as diatoms belonging to the genera *Rhizoselenia*, *Chaetoceros*, and the toxic bloom-forming raphidophyte *Heterosigma akashiwo* (44), and have been shown continued production and high diversity in oceanic communities on the west coast of North America (43, 45–47), in the North Pacific (48), in Antarctic waters (49) and in the Baltic Sea (50), and in freshwater lakes (51). Furthermore, when examined along a salinity gradient there was little overlap between the communities recovered and man unique viral OTUs recovered per site from six different sites (52). The choice of these two groups also has the advantage of encompassing groups from both DNA and RNA viruses. The dynamics of putative hosts were examined by sequencing amplified marker genes for eukaryotes (18S rRNA gene) and bacteria (16S rRNA gene). This contribution demonstrates that temporal changes in the phylogenetic make-up of viruses infecting bacteria and eukaryotic algae are related to environmental changes and to seasonal fluctuations in the communities of potential hosts.

Materials and methods

Sample collection

Seawater samples were collected from Jericho Pier (49° 16'36.73N, 123° 12'05.41W) in British Columbia, Canada. Jericho Pier (JP) is adjacent to the shoreline, in a well-mixed location with mixed semi-diurnal tides, and significant freshwater influence from the Fraser River. To get a representative sample of water and enough material for viral extraction, 60L of water was pumped from the 1-m depth every two weeks at the daytime high tide between June 2010 and July 2011 (33 samples). Salinity and temperature were measured using a YSI probe (Yellow Springs, Ohio, USA). For all samples, the water was pre-filtered through a 65-μm Nitex mesh and filtered sequentially through 142-mm diameter, 1.2-μm nominal pore-size glass-fibre (GC50 Advantec MFS, Dublin, CA., USA) and 0.22-μm pore-size polyvinyldine (Millipore, Bedford, MA, USA) filters. Viral size particles in the filtrate were concentrated to ~500 mL (viral concentrate) using tangential flow ultrafiltration with a 30kDa MW prep-scale Spiral Wound TFF-6 cartridge (Millipore) (53).

Nutrients

Phosphate, silicate and nitrate+nitrite concentrations were determined in duplicate 15-mL seawater samples filtered through 0.45 μm pore-size HA filters (Millipore) and stored at -20°C until air-segmented continuous-flow analysis on a AutoAnalyzer 3 (Bran+Luebbe, Norderstedt, Germany). Chlorophyll *a* (Chl *a*) was determined in triplicate by filtering 100 mL of seawater onto 0.45 μm pore-size HA filters (Millipore), and storing the filters in the dark at -20°C until acetone extraction and then analysed fluorometrically (54).

Enumeration of bacteria and viruses

Samples for viral and bacterial abundances were taken at each sampling point by fixing duplicate cryovials containing 980μL of sample with final concentration of 0.5% glutaraldehyde (EM-grade), freezing in liquid nitrogen and storing at -80°C until processing. Flow cytometry samples were processed as in Brussaard *et al* (55). Briefly, viral samples were diluted 1:10 to 1:10 000 in sterile 0.1 μm filtered 1X TE, stained with SYBR Green I (Invitrogen, Waltham, MA, USA) at a final concentration of 0.5 x 10^{-4} of commercial stock, heated for 10 minutes at 80° C and then cooled in the dark for 5 minutes before processing. Bacterial samples were diluted up to 1:1000 in sterile 0.1 μm filtered 1xTE, stained with SYBR Green I (Invitrogen) at a final concentration of 0.5 x 10^{-4} of commercial stock, and incubated in the dark for 15 minutes before processing. All samples were processed on a FACScalibur (Becton-Dickinson, Franklin Lakes, New Jersey, USA) with viral and bacterial samples run for 1 min at a medium or high flow rate, respectively. Event rates were kept between 100 to 1000 events per second and green fluorescence and side scatter detectors were used. Data were processed and gated using Cell-Quest software (Becton-Dickinson).

Extraction of viral nucleic acids

The viral concentrate was filtered twice through 0.22-μm pore-size Durapore PVDF filters (Millipore) in a sterile Sterivex filter unit (Millipore). Viral sized particles in the filtrate were pelleted by ultracentrifugation (Beckman-Coulter, Brea, California, USA) in a SW40 rotor at 108 000 *g* for 5 h at 12°C. The pellet was resuspended overnight in 100 μL of

supernatant at 4°C. To digest free DNA, the pellets were incubated with 1U μL^{-1} DNAse with a final concentration 5 mM $MgCl_2$ for 3 h at room temperature. Nucleic acids were extracted using a Qiamp Viral Minelute spin kit (Qiagen, Hilden, Germany) according to the manufacturer's directions.

PCR amplification of T4-like virus marker gene

To target the marine T4-like virus capsid protein gene (*gp23*), PCRs were set up as in Filee *et al* (42). Briefly, each reaction mixture (final volume, 50 μL) consisted of 2 μL template DNA, 1x (final concentration) PCR buffer (Invitrogen, Carlsbad, California, USA), 1.5 mM $MgCl_2$, 0.2 mM of each deoxynucleoside triphosphate (Bioline, London, UK), 40 pmol of MZIA1bis and 40pmol of MZIA6, and 1 U Platinum Taq DNA polymerase (Invitrogen) and was amplified using the program conditions as in Table 1.

PCR amplification of marnavirus-like marker gene

Half of each viral extract was used to synthesize complementary DNA (cDNA). To remove DNA, the extracted viral pellets digested with amplification-grade DNase 1 (Invitrogen). The reaction was terminated by adding 2.5 mM EDTA (final concentration) and incubating for 10 min at 65°C. Complementary DNA (cDNA) was generated using Superscript III reverse transcriptase (Invitrogen) with random hexamers (50 ng μL^{-1}) as per the manufacturer.

PCR was performed using primer set MPL-2 to target the *RdRp* of marnavirus-like viruses (43). Each reaction mixture (final volume, 50 μL) consisted of 50 ng of cDNA, 1x (final concentration) PCR buffer, 2 mM $MgCl_2$, 0.2 mM of each deoxynucleoside triphosphate, 1

212 μM of each primer, and 1 U Platinum Taq DNA polymerase. The reaction was run in a PCR
213 Express thermocycler (Hybaid, Ashford, UK) with program conditions as in Table 1.
214 Products were run on a 0.5X TBE 1% low melt gel, excised and extracted using Zymoclean
215 Gel DNA Recovery Kit (Zymo, Irvine, California, USA) as per the manufacturer and a final
216 elution step of 2x10 μL EB buffer (Qiagen).

217 Filtration and extraction of marine bacteria and eukaryotes

218 One liter from each 60-L seawater sample was filtered through a 0.22-μm pore-size
219 Durapore PVDF 47-mm filter (Millipore) in a sterile Sterivex filter unit (Millipore). The
220 filter was either stored at -20°C until extraction or immediately extracted as follows (56):
221 Briefly, filters were aseptically cut and incubated with lysozyme (Sigma-Aldrich, St. Louis,
222 MO, USA) at a final concentration of 1mg mL^{-1} for 2 h at 37°C. Sodium dodecyl sulfate was
223 added at a final concentration of 0.1% (w/v) and each filter was put through three freeze-
224 thaw cycles. Proteinase K (Qiagen) was then added to a final concentration of 100 μg mL^{-1}
225 and incubated for 1 h at 55°C. DNA was sequentially extracted using equal volumes of
226 phenol:chloroform:isoamyl alcohol (25:24:1), and chloroform:isoamyl alcohol (24:1). DNA
227 was precipitated by adding NaCl to a final concentration of 0.3M and by adding 2X the
228 extract volume of ethanol. Samples were incubated at -20°C for at least 1 h and then
229 centrifuged for 1 h at 20 000 *g* at 4°C. Extracts were washed with 70% ethanol and
230 resuspended in 50 μL EB buffer.

PCR amplification of bacterial and eukaryotic ribosomal sequences

PCR targeting eukaryotes used primers Euk1209f and Uni1392r (57). These primers target positions 1423 to 1641 and include the variable region V8. Each reaction mixture (final volume, 50 μL) consisted of 2 μL template, 1x (final concentration) PCR buffer, 1.5 mM $MgCl_2$, 0.2 mM of each deoxynucleoside triphosphate, 0.3 μM of each primer, and 2.5 U Platinum Taq DNA polymerase. The reaction was run in a PCR Express thermocycler with program conditions as in Table 1.

PCR targeting bacteria used primers 341F (58) and 907R (59). These primers target the v3 to v5 regions. PCRs were run with the following conditions: each reaction mixture (final volume, 50 μL) consisted of 2 μL template, 1x (final concentration) PCR buffer, 1.5 mM $MgCl_2$, 0.2 mM of each deoxynucleoside triphosphate, 0.4 µM of each primer, and 1 U Platinum Taq DNA polymerase. The reaction was run in a PCR Express thermocycler with program conditions as in Table 1.

Sequencing library preparation

Construction

PCR products not requiring gel excision were purified after PCR using AMPure XP beads (Beckman Coulter) at a ratio of 1.2:1 beads:product. Cleaned products were resuspended in 30 μL EB buffer (Qiagen). All products were quantified using the Picogreen dsDNA (Invitrogen) assay using Lambda DNA (Invitrogen) as a standard. Sample concentrations were read using iQ5 (Bio-Rad, Hercules, CA, USA) and CFX96 Touch systems (Bio-Rad).

Pooled libraries were constructed using one of each of the amplicons at a concentration so that their molarity would be similar and the total product of the pool to be ~700-900 ng. Pooled amplicons were concentrated using AMPure XP beads (Beckman Coulter) at a ratio of 1.2:1 beads:product. NxSeq DNA sample prep kit 2 (Lucigen, Middleton, WI, USA) was used as per manufacturer's directions with either NEXTFlex 48 barcodes (BioO, Austin, USA), NEXTflex 96 HT barcodes (BioO), or TruSeq adapters (IDT, Coralville, Iowa). Libraries were cleaned up using AMPure XP beads (Beckman Coulter) at a ratio of 0.9:1 beads:library.

Quantification and quality control of libraries

Libraries were checked for small fragments (primer dimers and/or adapter dimers) using a 2100 Bionanalyzer (Agilent, Santa Clara, CA, USA) with the High Sensitivity DNA kit (Agilent). The concentration of libraries was quantified using Picogreen dsDNA assay as above. The libraries were quantified and checked for amplifiable adapters using the Library Quantification DNA standards 1-6 (Kappa Biosystems, Wilmington, USA) with the SsoFast EvaGreen qPCR supermix (Bio-Rad) using 10 μL EvaGreen master mix, 3 μL of 0.5 μM F primer, 3 μL of 0.5 μm R primer and 4 μL of 1:1000, 1:5000 and 1:10000 dilutions of the libraries in triplicate on iQ5 (Bio-Rad) and CFX96 Touch qPCR machines. Cycling parameters were as follows: 95°C for 30s, 35 cycles of 95°C for 5s, 60°C for 30s, and the melt curve generation from 65°C to 95°C in 0.5°C steps (10s/step). Quantification from both Picogreen and qPCR assays were used to determine final pooling of all libraries before sequencing. Libraries were sequenced using 2x250bp PE Miseq (Illumina, San Diego, USA) sequencing at Génome Québec Innovation Centre at the McGill University (Montreal, QC,

Canada), and 2x300bp PE Miseq (Illumina) sequencing at UBC Pharmaceutical Sciences Sequencing Centre (Vancouver, BC, Canada) and at UCLA's Genoseq (Los Angeles, CA, USA).

Sequence processing

Libraries were either split by the sequencing center using CASAVA (Illumina) or with Miseq Reporter software (Illumina). Sequence quality was initially examined using FastQC (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/). Contaminating sequencing adapters were removed using Trimmomatic version 0.32 (60) and the quality of the sequencing library further examined using fastx_quality (61). Libraries were further split into individual amplicons (i.e. 18S, 16S, *gp23* and *RdRp*) and then, if the expected overlap of the paired-end reads was 40bp or more, the paired reads were merged using PEAR (62). For 16S the expected overlap was right around the cut-off, therefore instead the sequences were aligned to the SILVA database and then assigned to either the forward or reverse primer based on their position. Then the forward and reverse reads were concatenated. However, the reverse primer sequenced less readily than the forward and thus much information was lost if only sequences with both forward and reverse were kept, therefore only the forward reads were used for further analysis. All sequences were then quality trimmed using Trimmomatic with the default quality settings. Sequences were aligned to known sequences (Silva 119 database (63) for 16S and 18S rRNA genes) using align.seqs in mothur 1.33.3 (64) and those not aligned were removed. Viral sequences were queried using BLAST against databases containing the gene markers of interest from viral isolates and from other environmental surveys and sequences with hits with an e-value below 10^{-3} were kept.

The 16S and 18S rRNA gene sequences were checked for chimeras using USEARCH version 8.0.1517 (65) with the Gold reference database. Unique, non-chimeric sequences were clustered at 97% similarity. Taxonomy for the 16S and 18S rRNA gene sequences was assigned using mothur (Wang-type algorithm) using the Silva 119 taxonomy (63). For the viral targets sequences were chimera-checked using USEARCH denovo and reference (65). Viral sequences were then translated using FragGeneScan 1.20 (66). Viral reads were clustered using USEARCH (65) at 95% similarity for MVL, and 95% similarity for T4-like viruses. Since OTU clustering sequence similarity cut-offs are not well-defined for viruses, percent similarities (at the protein level) were examined at 50% to 100% for both MVL and T4-like viruses to examine an appropriate level these values chosen for similarity. For gp23, genera can have 35% protein similarity (67). When the number of OTUs compared to the percent similarity at 95% the rise in the number of OTUs was still linear and above 95% the slope more quickly (indicating many more OTUs picked with each per percent similarity increase) thus this level it was still sensitive without splitting potential groups and similar levels and approach as in other studies (68, 69). For MLV it was as previously described (70) and chosen at 95% sequence similarity at the protein level. This approach is similar to another OTU-based study targetting ssRNA viruses in coral photosymbionts (71). Operational taxonomic unit (OTU) tables for all targets were constructed using USEARCH (65). Rarefaction curves were generated using vegan (72). Sequences were normalized for by date and by target using vegan's rarefy (72). All of the initial sequence files have been deposited in NCBI’s BioProject database under ID PRJNA406940.

Data analysis, multivariate statistics and phylogenetic analysis

Environmental parameters with missing data, because of instrument malfunction or unavailability, were mean imputed to fill in missing values. Day length data were retrieved using R package geosphere (73). Adonis was used as implemented in vegan (72) to test whether community matrices showed seasonal differences. Bray-Curtis distance matrices were constructed from the normalized OTU abundance tables. Mantel tests were performed by comparing the community distance matrices to each other and to distance matrices of environmental parameters as implemented in vegan (72).

NCBI CDD domain alignments for *RdRp* and *gp23* were retrieved and used as hidden Markov models via HMMER (74) to align translated OTUs with Clustal Omega (75). Environmental sequences for both *gp23* (17, 42, 76–81) and *RdRp* (43, 45) were retrieved from Genbank to give context to the OTUs.

Alignments were viewed and manually curated with aliview (82). Automated trimming of the alignments was done using Trimal (83). Initial phylogenetic trees were built with Fast Tree (84). Final maximum likelihood trees were generated using RAxML (85) with 1000 bootstraps, with VT with the PROTGAMMA model for the RdRp gene tree and JTT with the PROTGAMMA model for the *gp23* gene tree chosen using Prottest (86). Faith's phylogenetic diversity (87) was calculated as implemented in picante (88). The package ggtree (89) was used for visualizing and annotating trees and all other plots were created using R (90). All scripts used for processing the data are available at: https://github.com/jooolia/phylo_temporal_jericho (doi: 10.5281/zenodo.2554772).

Results

Variability of environmental characteristics

Chlorophyll *a* (chl *a*) concentrations varied over time with a maximum observed concentration during a eukaryotic phytoplankton bloom (46.5 μg L^{-1} in June 2011)(Fig. S1). The second highest chl *a* occurred during the annual spring bloom in late April 2011 (5.88 μg L^{-1}) which is mainly composed of diatoms belonging to *Thallassiosira* sp. (91, 92). The minimum chl *a* value of 0.05 μg L^{-1} occurred in May and the chlorophyll levels remained below 1 μg L^{-1} from September to March.

Nutrient concentrations were also highly dynamic ranging between 6.1 μM to 67.3 μM for silicate, from below 0.1 μM to 2.3 μM for phosphate and from below 0.1 μM to 27.7 μM for nitrate+nitrite (Fig. S1). Overall, nutrient concentrations were high and stable over winter, dipped in late April and then were followed by a large increase in silicate commencing in May.

Variability of viral and bacterial abundance

During the time series, the viral abundance ranged from 5.41×10^6 to 4.70×10^7 particles mL^{-1} while the bacterial abundance was one order of magnitude lower ranging from 6.59×10^5 to 4.43×10^6 cells mL^{-1} (Fig. S1).

Shared viral and microbial OTUs over time

Amplicon target reads representing marnavirus-like viruses (*RdRp*) and T4-like viruses (*gp23*) were translated into amino acids and the reads were normalized to the library with fewest reads. The 16S and 18S reads were normalized the same way using the nucleotide sequences. This resulted in 566 OTUs at 95% sequence similarity for the marnavirus-like viruses, with between 59 and 142 OTUs per timepoint with an average of 6% of these OTUs shared over time (Fig. S2). For the T4-like viruses there were a total of 1737 OTUs at 95% sequence similarity, with a minimum of 149 and a maximum of 484 OTUs per timepoint (Fig. S2). On average, 6% of the T4-like viruses OTUs were shared among all timepoints. There were 813 bacterial OTUs (97% sequence similarity) with an average of 10% shared over time. The lowest number of OTUs seen per timepoint was 84 and the highest number was 269.The phylum-level classifications showed a dominance of Proteobacteria and Bacteriodetes and Cyanobacteria showing large fluctuations (Fig. S3). In the eukaryotic community a total of 1115 OTUs (97% sequence similarity) were found with 6% shared on average, a minimum of 55 and a maximum of 298 (Fig. S2).The phylum-level classifications showed a dominance of the SAR group with Ophisthokonta, Haptophyta and Cryptophycaea making up the majority of the sequences recovered (Fig. S4). Rarefaction curves for individual samples did not flatten ("saturate") indicating that not all possible OTUs were sequenced in these samples, but when considered together the curves saturated indicating that even if all the diversity was not captured in one sample, the overall community of OTUs was captured (Fig. S5).

Dynamics of phylogenetically-related viral OTUs

Viral OTUs were placed into a phylogenetic context and their dynamics examined over time (Fig. 1,Fig. S6 and Fig. S7). Well-defined and well-supported phylogenetic groups (A-H) of marnavirus-like viruses (Fig. 1A and B) showed strong temporal dynamics and differed by season (Adonis R= 0.503, *P*-value= 0.001). Group H, which included isolated viruses infecting the diatoms *Chaetoceros* sp., and *Rhizoselenia* sp., were constant members of the communities although their relative abundance was highest in late November. Group A was always present and included many environmental sequences, as well as sequences from a virus that infects the raphidophyte, *Heterosigma akashiwo*; the group was most abundant between August and September. The winter months from October to February were dominated by OTUs in Group E, with a smaller contribution by Group H. The structure of the MVL OTU phylogenetic groups closely mirrored the structure of the top 20 OTUs found over time (Fig. S8), demonstrating that this community contained few dominant OTUs.

The T4-like virus OTUs were also placed in a phylogenetic context and categorized into groups of related OTUs (Fig. 1 C). In the fall, Group I dominated the community, followed by Group G; both groups included viral isolates infecting cyanobacteria. In January 35.7% of the relative abundance of the T4-like viruses was represented by Group B which contains no known isolates. Unlike the MVL community, the T4-like virus community had very different patterns among the top 20 OTUs and the phylogenetic groups over time (Fig. S9). When there was a large increase in nutrients in late September (Fig. S1), the dominant groups in the community shifted. The community returned to its previous state by the next

396 sampling time. The T4-like virus communities showed small differences by season (Adonis
397 R = 0.23, *P*-value = 0.001)

398 Richness of related viruses over time

399 Within phylogenetic groupings the richness of OTUs varied over time in the viral
400 communities (Fig. 2, Fig. S10 and Fig. 3, Fig. S11). Often, within the groups, the richness and
401 relative abundance increased at the same timepoint indicating that at periods when these
402 groups of OTUs appeared to dominate the overall communities, there was an increase in
403 alpha diversity within the group.

404 Examining the richness over time for MVL viruses, for OTUs in group A there was a peak in
405 relative abundance and richness in the Fall and later the richness dramatically decreased
406 but this group showed continued presence over time. Similar patterns were observed for
407 groups E and G where there was a peak in their abundance and richness at one point in the
408 time series and otherwise a low, but detectable level of viruses. Conversely, for group H the
409 pattern is different whereby the richness and relative abundance fluctuate but the peaks
410 and troughs are less steep than in the other groups indicating a larger number of persistent
411 OTUs. Group H comprised a large proportion of the overall OTUs (47% percent overall)
412 which affects the relative dynamics observed in the MVL community.

413 For the T4-like viruses, Groups A, C and D had similar dynamics during the time series
414 where they showed high richness for their group in June-August, then lower richness from
415 August until February, and then in February there was a small peak in richness. These
416 groups did not show large increases or decreases in the relative abundance but made up a

small proportion overall of the relative abundance throughout the time series. Groups B, G, H and I had dynamics whereby they showed a continued presence throughout the time series with fluctuations between timepoints and at times a high relative abundance attributed to these groups. Groups E and F showed higher richness in the Fall than the rest of the year and a small peak in February.

There were many persistent and ephemeral OTUs over time in the viral communities (Fig. 2 and Fig. 3). Furthermore, many of the ephemeral viral OTUs were also related to the persistent viral OTUs. In the T4-like viruses there were many more OTUs than for the marnavirus-like viruses, however, both communities fluctuated over time (0.9% of the OTUs were found at 90% of the timepoints for MVL, vs. 0.8% of the OTUs in the T4-like viruses). In the T4-like viruses Groups A, B, G and F contained both persistent and ephemeral OTUs. Conversely group C had few OTUs that persisted. In the MVL viruses there were fewer OTUs overall and also fewer OTUs that persisted over time. Group A had one OTU that persisted over time and other OTUs that were only found at 4-6 sampling points, which encompassed 8 to 12 weeks at the sampling site.

Lagged correlations with hosts over time

Raphidophytes and marnavirus-like viral Group A

Group A (Fig. 1A), which includes HaRNAV, a virus which infects the raphidophyte *Heterosigma akashiwo*, increased in relative abundance after the relative abundance of eukaryotic sequences classified as raphidophytes increased (Fig. 4, Spearman rank correlation over entire time series: 0.43 and *P*-value 0.07). There were further peaks in the

relative abundance of raphidophyte sequences, but they did not coincide with increases in the relative abundance of marnavirus-like viral Group A. Of all the OTUs in Group A, OTU 1 was relatively most abundant, and when aligned with other closely related sequences (Fig. S12) these sequences showed changes in the amino acid D to E in the palm region of the *RdRp* (93).

Cyanobacteria and T4-like virus Group I

Comparing the T4-like virus Group I, which contains cyanophage isolates, to cyanobacterial OTUs (Fig. 5), showed that the relative abundance of viruses increased in the fall after peaks in the relative abundance of cyanobacterial OTUs (correlation over entire time series: 0.56, *P*-value: 0.03). The lags in relative abundances of putative cyanophages relative to cyanobacteria continued, and after the spring diatom bloom (April 8th) there was a lag before the relative abundance of a different putative cyanophage in Group I increased, showing succession in the viral community.

The community similarity of the T4-like viruses had a strong lagged negative correlation to the richness of the bacterial community (Fig. S13 and Fig. 6). The correlations to the MVL community were much stronger to other communities when lagged than when directly compared. Viral abundance was negatively correlated to bacterial community similarity.

Mantel tests among community similarity

Mantel tests were used to examine concurrent community and environmental changes over time (Fig. 6). The overall bacterial and eukaryotic community compositions fluctuated strongly together. The marnavirus-like community changes were predicted most strongly

by changes in the eukaryotic community with a lag of two weeks and less strongly by the bacterial and myoviral communities. The T4-like viruses showed changes most strongly with the bacterial communities with a two-week lag, however also the eukaryotic and marnavirus-like communities with lags were also strong predictors of the myoviral communities.

Discussion

Temporal shifts in community dominance by groups of related viruses

A major theme in microbial ecology is understanding the causes of temporal shifts in community composition. Previously, a study at the Bermuda Atlantic Time series Study (BATS) used deep sequencing of amplicons of the viral gene marker *pho h*, to resolve phylogenetically distinct groups of viral populations that differed between fall and winter; these differences were attributed to a phylogenetic group containing cyanobacteria-infecting viruses and to water stratification (35). In our study, in coastal British Columbia, the temporal dynamics were driven by shifts in groups of phylogenetically-related OTUs (Fig. 1) and there was a consistently high diversity of OTUs in the viral communities. The dynamics of MVL viruses were similar when examined as groups of related viruses (Fig. 1B) and as the top 20 MVL OTUs over time (Fig. S8) indicating an uneven community as has been previously seen (46, 47, 70). Conversely, in the T4-like viruses the patterns among phylogenetically-related groups did not resemble the patterns of the top 20 OTUs, indicating that the community was more diverse and even. Even with these differences, the phylogenetic dynamics fluctuated in both the T4-like and MVL viruses even though viral

richness stayed relatively constant over time (Fig. 2, Fig. 3,Fig. S2). Dynamics within the phylogenetic groups showed several different patterns that could be classified as having one large peak per year, several peaks or a constant presence in the community. These different dynamics could represent differences in lifestyles (r- or K-selected) (94), hosts, or the influence of environmental parameters. This would help explain how constant co-evolution or co-presence of viruses that is likely a factor in the Killing the Winner model (30).

The dynamics of the phylogenetic groups reflect that viral and microbial communities can be stable at the genus level and above, but are dynamic at the strain or species level (31). Furthermore, taxonomically related species can have similar niches and ecology (36, 37). For example, although there are exceptions, viruses infecting marine cyanobacteria and eukaryotic phytoplankton fall into different phylogenetic clusters depending on the host that they infect (95, 96), consistent with the hosts representing specific niches. Alternatively, given that the genetic similarity of marker gene sequences predicts gene content in at least some groups of double-stranded DNA viruses (96), and that viral gene content or GC ratios may result in light-dependent effects on growth or nutrient effects (97, 98), it is likely that environmental factors may have different influences on different phylogenetic groups of viruses. For example, Moniruzzaman *et al* (99) saw increased activity of ssRNA viruses with the termination of a phytoplankton bloom. Consequently, the phylogenetic signal is a very important part of the community structure.

Previously it has been found that the largest OTUs were persistent throughout time and it was suggested that transient viruses were dispersed from different areas (31). This seems

unlikely in our study since it was frequent that the ephemeral viruses were closely related to the persistent viruses. Furthermore, both persistent and ephemeral viruses of phytoplankton were found in a freshwater lakes (100, 101), and at a coastal site (56), where some ephemeral viruses were correlated with shifts in environmental parameters. In another freshwater study it appeared that most viruses were transient and the system saw a very quick turnover (33). Additionally, in marine T4-like-virus communities, some OTUs persisted but many more were ephemeral in 3-year (17) and 2-year (18) time series. Furthermore, in a daily study of T4-like viral communities, few large-scale increase or decrease events were seen and overall the changes in both the bacterial and T4-like-viral communities were slow (102). Additionally, the data show that some OTUs are persistent and thus continually successful, while others are ephemeral (Fig. S10 and Fig. S11) with some of the viral groups having no persistent OTUs (T4-like C and D and MVL B, C, D, F and G). Therefore, both eukaryotic and bacterial marine viruses show the pattern of ephemeral and persistent viruses and adding the phylogenetic relatedness gives a deeper understanding of these dynamics.

Potential quasispecies behaviour

The population structure of RNA viruses is proposed to be a mixture of genotypes called quasispecies that are produced through polymerase errors; they encompass the community of genotypes theoretically produced from one infection (103, 104). It is proposed to be a result of higher mutation rates and higher burst sizes found in RNA viruses compared to double-stranded DNA viruses (105). However, it is theoretically possible for quasispecies to exist in DNA bacteriophage populations (106). In high

524 confidence viral DNA metagenomes there is heterogeneity in assembled reads beyond
525 expected sequencing errors (107), and there is site-specific variation in DNA viral genome
526 populations studied in humans (108). Thus a sequenced viral genome could be considered
527 to be an "average" of individual genotypes. In viral RNA metagenomic sequences observed
528 in an Antarctic lake, the ecological setting likely influences the presence of quasispecies
529 (109) since there are more single-nucleotide variations (SNVs) in the lake-water
530 metagenomes than in the microbial mats that they studied. This is explained either by
531 higher turnover in the lake water, and thus more ecological niches/diversity in these
532 samples, or as the result of convergence of water from more locations (109). When the
533 OTUs in the marnavirus-like Group A (group which includes viral isolate *HaRNAV*) were
534 visualized as an alignment (Fig. S12), a small number of the differences between the
535 ephemeral and persistent OTUs were randomly spread across the gene fragment; however,
536 most of the differences were in the "palm" section of the catalytic site C (93) where at a
537 specific position the most persistent viral sequence had D and the ephemeral sequences
538 had E. These ephemeral sequences point to a population that is marginally successful while
539 the most abundant OTU (OTU_2, retaining the D amino acid) remains persistent. This
540 suggests that the high diversity of related-viruses means a successful large-scale infection
541 event as seen in clinical studies (104) and this suggests that this phenomenon could be
542 prevalent in marine settings (and should be incorporated into current ecological theories
543 (e.g. KtW)).

Implications for theories related to community structure and dynamics

In the Killing the Winner model, viruses infect the most active organism (28). One interpretation is that hosts compete for limiting resources which determine the community composition at a particular site, and viruses determine the specific host species and the abundance of these hosts (110). With the lagged dynamics observed in the raphidophytes and viruses related to raphidophyte-infecting viruses, the "winner" has been killed (Fig. 4). Subsequently, there is a small increase in the relative abundance of raphidophytes but no associated increase in RdRp viral Group A. There are several explanations for the increase in raphidophytes without an observed increase in viral Group A. Possible explanations are that the surviving raphidophytes were resistant to specific viruses after infection (111), in this case members from viral Group A, or that the number of susceptible hosts was too low to allow for a detectable increase in viruses. Alternatively, a different subset of viruses, such as the DNA virus HAV (112), or protist grazing (113) could be keeping populations low, and thus preventing the replication of the Group A viruses. Similar patterns were detected in T4-like myoviral Group I (including isolates from Prochlorococcus phage *PSSM2* and Synechococcus phage *SSM2*) and cyanobacteria (Fig. 5). These types of patterns have been observed in other studies where there were many viral OTUs detected with strong time-lagged correlations to bacterial OTUs (40). Also, in a mesocosm experiment with *Emiliania huxleyi* there was a peak in host abundance and then four days later a peak in *Emiliania huxleyi* Virus (EhV) abundance (114). Rapid shifts in fine-scale viral dynamics and stability at coarse scale viral dynamics suggest that the Killing the Winner theory is

566 operating at the strain level (31, 32, 115) preserving bacterial strain level diversity (12,
567 29).

568 In the marine environment, viruses, at the coarse scale of family or genus, could show
569 Killing the Winner dynamics with finer scale dynamics at the strain level. This bank or seed
570 bank model (41) explains how high local viral diversity (shuffling of viruses) can be
571 consistent with low overall global diversity (represented by the most abundant viruses) by
572 a constant local production of viruses has been supported by many studies (17, 116–118).
573 Previously it has been found that the viral community was mostly dominated (>50%) by a
574 few successful OTUs and the rest of the OTUs were rare and contained in the "bank" (35).
575 We hypothesized that the communities would have a "seed bank" where there is shuffling
576 in rank of phylogenetically-related viruses along a rank-abundance curve, however, when
577 examined with the phylogenetic signal it is not a shuffling of rank of these related viruses
578 as there are few viral OTUs that dominate within each phylogenetic group that dominate
579 over time and the other OTUs are ephemeral. Thus the seed bank glosses over important
580 information revealed by the phylogenetic examination of this communities. This seed bank
581 idea could be effective at different time scales, but at the scale of one year with samples
582 every two weeks there was no evidence of shuffling of rank of viruses within these viral
583 groups. Our study has deepened this understanding by showing that the relatedness of the
584 viruses is crucial to understanding their dynamics and reveals that the seed bank does not
585 appear to be operating as described. Our data reveal that the viruses within phylogenetic
586 groups can be ephemeral and related to persistent viruses and that groups of related
587 viruses can become abundant through ecological processes (e.g. through habitat filtering
588 where closely related species can persist in a particular environment (119)).

Caveats

Although the challenges with viral gene markers (43, 70) and PCR in general (120) have been discussed, it remains an excellent approach for examining population structures. Although it could be argued that rare and putative quasispecies OTUs could result from PCR errors (121) or sampling anomalies (122), this is unlikely given that these OTUs were seen multiple times in different samples, implying that they are not spurious. To increase confidence in the results, some libraries with fewer reads were excluded so that more sequence could be used overall when normalizing samples (121) even though this has the trade-off of decreasing the number of samples. The sequences were checked for chimeras since chimeras can form as a result of high cycle number (123). Hence, although the read abundance of OTUs is semi-quantitative, it is a good approach for comparing richness and diversity among samples (but not for absolute counts of genes) (121). Thus, we are confident that our results conservatively reflect the changes that occurred in these microbial communities.

Conclusions

Phylogenetically-related viruses showed temporal patterns of dominance within the viral communities over time. As well, viral communities showed evidence of time-lagged dynamics related to the potential host communities, where, for example, marnavirus-like viral (MVL) group A (containing isolate *HaRNAV*) was correlated in a lagged fashion to the raphidophytes, and myoviral virus group I (containing isolates *PSSM2* and *SSM2*) was correlated with cyanobacteria. MVL-like and myoviral viral communities differed in fine-

scale community structure illustrated by differences in the proportion of OTUs that persisted over time, the evenness, and the diversity of the communities.

The MVL, which contained isolates with high burst sizes, exhibited potential quasispecies-type behaviour whereby OTUs at one timepoint were composed of many different closely related viruses and the differences between the ephemeral and persistent OTUs differed mostly in a protein residue at a catalytic site suggesting that this diversity could originate within one infection.

Previously, the link between dominance, persistence and phylogeny of virus-host communities has largely been overlooked. Other studies have found that many viral OTUs are ephemeral and that few are persistent; whereas, this study demonstrated that most of the ephemeral viral OTUs were closely related to a persistent viral OTU, and that over time the community was dominated by different phylogenetic viral groups composed of related ephemeral and persistent OTUs.

Hence, the observed changes were not the result of the presence of a seed bank that led to an overall shuffling in ranks of OTUs in the community, but rather the changes were related to fluctuations in the dominance of different phylogenetic groups of viruses over time. These dynamics thus add an important insight about the structure of viral communities and are thus crucial for understanding these viral communities.

Acknowledgements

Suttle lab and Beaty Biodiversity members assisted greatly through discussions, manuscript improvements, assisting in field work and facilitating laboratory and analytical

work. In particular the contributions of A.M. Chan, C. Chénard, C-E Chow, J.L. Clasen, C. Deeg, J.F. Finke, T.J. Heger, X.A. Tian, M. Vlok, and X. Zhong are gratefully acknowledged. C. Payne conducted the nutrient analyses and R. Pawlowicz provided the YSI probe. Thanks also to A.M. Comeau and A.I. Culley for providing reference alignments for *gp23* and *RdRp*, respectively. The research was made possible through funding from NSERC through PGS-M and PGS-D awards to JAG, and Discovery Research and Ship-time Grants to CAS, fellowships to JAG from UBC and the Beaty Biodiversity Center through a BRITE award from an NSERC CREATE grant. The Canada Foundation for Innovation and the British Columbia Knowledge Development Fund funded equipment purchases that allowed the work to be done.

Conflict of Interest

The authors declare no conflict of interest.

References

1. Lynch MDJ, Neufeld JD. 2015. Ecology and exploration of the rare biosphere. Nature Reviews Microbiology 13:217–229.

2. DuRand MD, Olson RJ, Chisholm SW. 2001. Phytoplankton population dynamics at the Bermuda Atlantic Time-series station in the Sargasso Sea. Deep Sea Research Part II: Topical Studies in Oceanography 48:1983–2003.

3. Morris RM, Vergin KL, Cho J-C, Rappé MS, Carlson CA, Giovannoni SJ. 2005. Temporal and spatial response of bacterioplankton lineages to annual convective overturn at the Bermuda Atlantic Time-series Study site. Limnology and Oceanography 50:1687–1696.

4. Fuhrman JA, Hewson I, Schwalbach M, Steele J, Brown M, Naeem S. 2006. Annually reoccurring bacterial communities are predictable from ocean conditions. Proceedings of the National Academy of Sciences 103:13104–13109.

5. Gilbert JA, Field D, Swift P, Newbold L, Oliver A, Smyth T, Somerfield PJ, Huse S, Joint I. 2009. The seasonal structure of microbial communities in the Western English Channel. Environmental Microbiology 11:3132–3139.

6. Suttle CA. 2005. Viruses in the sea. Nature 437:356–361.

7. Paul JH. 2008. Prophages in marine bacteria: Dangerous molecular time bombs or the key to survival in the seas? The ISME Journal 2:579–589.

8. Brum JR, Schenck RO, Sullivan MB. 2013. Global morphological analysis of marine viruses shows minimal regional variation and dominance of non-tailed viruses. The ISME Journal 7:1738–1751.

9. Breitbart M. 2012. Marine viruses: Truth or dare. Annual Review of Marine Science 4:425–448.

10. Lang AS, Rise ML, Culley AI, Steward GF. 2009. RNA viruses in the sea. FEMS Microbiology Reviews 33:295–323.

667 11. Moreau H, Piganeau G, Desdevises Y, Cooke R, Derelle E, Grimsley N. 2010. Marine
668 Prasinovirus genomes show low evolutionary divergence and acquisition of protein
669 metabolism genes by horizontal gene transfer. Journal of Virology 84:12555–12563.

670 12. Rodriguez-Valera F, Martin-Cuadrado A-B, Rodriguez-Brito B, Pašić L, Thingstad TF,
671 Rohwer F, Mira A. 2009. Explaining microbial population genomics through phage
672 predation. Nature Reviews Microbiology 7:828–836.

673 13. Bergh O, Borsheim KY, Bratbak G, Heldal M. 1989. High abundance of viruses found in
674 aquatic environments. Nature 340:467–468.

675 14. Jiang SC, Paul JH. 1994. Seasonal and diel abundance of viruses and occurrence of
676 lysogeny/bacteriocinogeny in the marine environment. Marine Ecology-Progress Series
677 104:163–172.

678 15. Winget DM, Helton RR, Williamson KE, Bench SR, Williamson SJ, Wommack KE. 2011.
679 Repeating patterns of virioplankton production within an estuarine ecosystem.
680 Proceedings of the National Academy of Sciences 108:11506–11511.

681 16. Parsons RJ, Breitbart M, Lomas MW, Carlson CA. 2012. Ocean time-series reveals
682 recurring seasonal patterns of virioplankton dynamics in the northwestern Sargasso Sea.
683 The ISME Journal 6:273–284.

684 17. Chow C-ET, Fuhrman JA. 2012. Seasonality and monthly dynamics of marine myovirus
685 communities: Marine myovirus community dynamics at SPOT. Environmental Microbiology
686 14:2171–2183.

687 18. Pagarete A, Chow C-ET, Johannessen T, Fuhrman JA, Thingstad TF, Sandaa RA. 2013.
688 Strong seasonality and interannual recurrence in marine myovirus communities. Applied
689 and Environmental Microbiology 79:6253–6259.

690 19. Waterbury JB, Valois FW. 1993. Resistance to co-occurring phages enables marine
691 *Synechococcus* communities to coexist with cyanophages abundant in seawater. Applied
692 and Environmental Microbiology 59:3393–3399.

693 20. Suttle CA, Chan AM. 1994. Dynamics and distribution of cyanophages and their effect on
694 marine *Synechococcus* spp. Applied and Environmental Microbiology 60:3167–3174.

695 21. Marston MF, Taylor S, Sme N, Parsons RJ, Noyes TJE, Martiny JBH. 2013. Marine
696 cyanophages exhibit local and regional biogeography. Environmental Microbiology
697 15:1452–1463.

698 22. Sandaa R-A, E. Storesund J, Olesin E, Lund Paulsen M, Larsen A, Bratbak G, Ray J. 2018.
699 Seasonality drives microbial community structure, shaping both eukaryotic and
700 prokaryotic host–viral relationships in an Arctic marine ecosystem. Viruses 10:715.

701 23. Clasen J, Hanson C, Ibrahim Y, Weihe C, Marston M, Martiny J. 2013. Diversity and
702 temporal dynamics of Southern California coastal marine cyanophage isolates. Aquatic
703 Microbial Ecology 69:17–31.

704 24. Middelboe M, Hagström A, Blackburn N, Sinn B, Fischer U, Borch NH, Pinhassi J, Simu K,
705 Lorenz MG. 2001. Effects of bacteriophages on the population dynamics of four strains of
706 pelagic marine bacteria. Microbial Ecology 42:395–406.

707 25. Bouvier T, Giorgio PA del. 2007. Key role of selective viral-induced mortality in
708 determining marine bacterial community composition. Environmental Microbiology
709 9:287–297.

710 26. Marston MF, Pierciey FJ, Shepard A, Gearin G, Qi J, Yandava C, Schuster SC, Henn MR,
711 Martiny JBH. 2012. Rapid diversification of coevolving marine *Synechococcus* and a virus.
712 Proceedings of the National Academy of Sciences 109:4544–4549.

713 27. Hewson I, Vargo GA, Fuhrman JA. 2003. Bacterial diversity in shallow oligotrophic
714 marine benthos and overlying waters: Effects of virus infection, containment, and nutrient
715 enrichment. Microbial Ecology 46:322–336.

716 28. Thingstad TF. 2000. Elements of a theory for the mechanisms controlling abundance,
717 diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. Limnology
718 and Oceanography 45:1320–1328.

719 29. Thingstad TF, Pree B, Giske J, Våge S. 2015. What difference does it make if viruses are
720 strain-, rather than species-specific? Frontiers in Microbiology 6:320.

721 30. Xue C, Goldenfeld N. 2017. Coevolution maintains diversity in the stochastic "Kill the
722 Winner" model. Physical Review Letters 119:268101.

723 31. Rodriguez-Brito B, Li L, Wegley L, Furlan M, Angly F, Breitbart M, Buchanan J, Desnues
724 C, Dinsdale E, Edwards R, Felts B, Haynes M, Liu H, Lipson D, Mahaffy J, Martin-Cuadrado
725 AB, Mira A, Nulton J, Pašić L, Rayhawk S, Rodriguez-Mueller J, Rodriguez-Valera F, Salamon
726 P, Srinagesh S, Thingstad TF, Tran T, Thurber RV, Willner D, Youle M, Rohwer F. 2010. Viral

727 and microbial community dynamics in four aquatic environments. The ISME Journal 4:739–
728 751.

729 32. Needham DM, Sachdeva R, Fuhrman JA. 2017. Ecological dynamics and co-occurrence
730 among marine phytoplankton, bacteria and myoviruses shows microdiversity matters. The
731 ISME Journal 11:1614–1629.

732 33. Arkhipova K, Skvortsov T, Quinn JP, McGrath JW, Allen CC, Dutilh BE, McElarney Y,
733 Kulakov LA. 2018. Temporal dynamics of uncultured viruses: A new dimension in viral
734 diversity. The ISME Journal 12:199–211.

735 34. Winter C, Bouvier T, Weinbauer MG, Thingstad TF. 2010. Trade-offs between
736 competition and defense specialists among unicellular planktonic organisms: The "Killing
737 the Winner" hypothesis revisited. Microbiology and Molecular Biology Reviews 74:42–57.

738 35. Goldsmith DB, Parsons RJ, Beyene D, Salamon P, Breitbart M. 2015. Deep sequencing of
739 the viral *phoH* gene reveals temporal variation, depth-specific composition, and persistent
740 dominance of the same viral *phoH* genes in the Sargasso Sea. PeerJ 3:e997.

741 36. Harvey P, Purvis A. 1991. Comparative methods for explaining adaptations. Nature
742 351:619–624.

743 37. Srivastava DS, Cadotte MW, MacDonald AAM, Marushia RG, Mirotchnick N. 2012.
744 Phylogenetic diversity and the functioning of ecosystems. Ecology Letters 15:637–648.

745 38. Horner-Devine MC, Bohannan BJ. 2006. Phylogenetic clustering and overdispersion in
746 bacterial communities. Ecology 87:S100–S108.

747 39. Lennon JT, Aanderud ZT, Lehmkuhl BK, Schoolmaster Jr DR. 2012. Mapping the niche
748 space of soil microorganisms using taxonomy and traits. Ecology 93:1867–1879.

749 40. Chow C-ET, Kim DY, Sachdeva R, Caron DA, Fuhrman JA. 2014. Top-down controls on
750 bacterial community structure: Microbial network analysis of bacteria, T4-like viruses and
751 protists. The ISME Journal 8:816–829.

752 41. Breitbart M, Rohwer F. 2005. Here a virus, there a virus, everywhere the same virus?
753 Trends in Microbiology 13:278–284.

754 42. Filée J, Tétart F, Suttle CA, Krisch HM. 2005. Marine T4-type bacteriophages, a
755 ubiquitous component of the dark matter of the biosphere. Proceedings of the National
756 Academy of Sciences 102:12471–12476.

757 43. Culley AI, Steward GF. 2007. New genera of RNA viruses in subtropical seawater,
758 inferred from polymerase gene sequences. Applied and Environmental Microbiology
759 73:5937–5944.

760 44. Tomaru Y. 2015. Marine diatom viruses and their hosts: Resistance mechanisms and
761 population dynamics. Perspectives in Phycology 2:69–81.

762 45. Culley AI, Lang A, Suttle CA. 2003. High diversity of unknown picorna-like viruses in the
763 sea. Nature 424:1054–1057.

764 46. Culley AI, Lang A, Suttle CA. 2006. Metagenomic analysis of coastal RNA virus
765 communities. Science 312:1795–1798.

766 47. Culley AI, Mueller JA, Belcaid M, Wood-Charlson EM, Poisson G, Steward GF. 2014. The
767 characterization of RNA viruses in tropical seawater using targeted PCR and metagenomics.
768 mBio 5:e01210–14.

769 48. Urayama S-i, Takaki Y, Nishi S, Yoshida-Takashima Y, Deguchi S, Takai K, Nunoura T.
770 2018. Unveiling the RNA virosphere associated with marine microorganisms. Molecular
771 Ecology Resources 18:1444–1455.

772 49. Miranda JA, Culley AI, Schvarcz CR, Steward GF. 2016. RNA viruses as major
773 contributors to Antarctic virioplankton: RNA viruses in the Antarctic. Environmental
774 Microbiology 18:3714–3727.

775 50. Allen LZ, McCrow JP, Ininbergs K, Dupont CL, Badger JH, Hoffman JM, Ekman M, Allen
776 AE, Bergman B, Venter JC. 2017. The Baltic Sea Virome: Diversity and Transcriptional
777 Activity of DNA and RNA Viruses. MSystems 2:e00125–16.

778 51. Hewson I, Bistolas KS, Button JB, Jackson EW. 2018. Occurrence and seasonal dynamics
779 of RNA viral genotypes in three contrasting temperate lakes. PloS ONE 13:e0194419.

780 52. Labbé M, Raymond F, Lévesque A, Thaler M, Mohit V, Audet M, Corbeil J, Culley A. 2018.
781 Communities of phytoplankton viruses across the transition zone of the St. Lawrence
782 estuary. Viruses 10:672.

783 53. Suttle CA, Chan AM, Cottrell MT. 1991. Use of ultrafiltration to isolate viruses from
784 seawater which are pathogens of marine phytoplankton. Applied and Environmental
785 Microbiology 57:721–726.

786 54. Parsons T, Maita Y, Lalli C. 1984. A manual of chemical and biological methods for
787 seawater analysis. Pergamon Press, New York.

788 55. Brussaard CPD. 2004. Optimization of procedures for counting viruses by flow
789 cytometry. Applied and Environmental Microbiology 70:1506–1513.

790 56. Short SM, Suttle CA. 2003. Temporal dynamics of natural communities of marine algal
791 viruses and eukaryotes. Aquatic Microbial Ecology 32:107–119.

792 57. Diez B, Pedros-Alio C, Marsh TL, Massana R. 2001. Application of denaturing gradient
793 gel electrophoresis (DGGE) to study the diversity of marine picoeukaryotic assemblages
794 and comparison of DGGE with other molecular techniques. Applied and Environmental
795 Microbiology 67:2942–2951.

796 58. Baker G, Smith J, Cowan D. 2003. Review and re-analysis of domain-specific 16S
797 primers. Journal of Microbiological Methods 55:541–555.

798 59. Muyzer G, Teske A, Wirsen CO, Jannasch HW. 1995. Phylogenetic relationships of
799 *Thiomicrospira* species and their identification in deep-sea hydrothermal vent samples by
800 denaturing gradient gel electrophoresis of 16S rDNA fragments. Archives of Microbiology
801 164:165–172.

802 60. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic a flexible trimmer for Illumina
803 sequence data. Bioinformatics 30:2114–2120.

804 61. Gordon A. 2014. Fastx_toolkit.

805 62. Zhang J, Kobert K, Flouri T, Stamatakis A. 2014. PEAR: A fast and accurate Illumina
806 Paired-End reAd mergeR. Bioinformatics 30:614–620.

807 63. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glockner FO. 2013.
808 The SILVA ribosomal RNA gene database project: Improved data processing and web-based
809 tools. Nucleic Acids Research 41:D590–D596.

810 64. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA,
811 Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF.
812 2009. Introducing mothur: Open-source, platform-independent, community-supported
813 software for describing and comparing microbial communities. Applied and Environmental
814 Microbiology 75:7537–7541.

815 65. Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST.
816 Bioinformatics 26:2460–2461.

817 66. Rho M, Tang H, Ye Y. 2010. FragGeneScan predicting genes in short and error-prone
818 reads. Nucleic Acids Research 38:e191–e191.

819 67. Lavigne R, Darius P, Summer EJ, Seto D, Mahadevan P, Nilsson AS, Ackermann HW,
820 Kropinski AM. 2009. Classification of Myoviridae bacteriophages using protein sequence
821 similarity. BMC Microbiology 9:224.

822 68. Hou W, Wang S, Briggs BR, Li G, Xie W, Dong H. 2018. High diversity of myocyanophage
823 in various aquatic environments revealed by high-throughput sequencing of major capsid
824 protein gene with a new set of primers. Frontiers in Microbiology 9:887.

825 69. Buerger P, Weynberg K, Wood-Charlson E, Sato Y, Willis B, Oppen M van. 2019. Novel
826 T4 bacteriophages associated with black band disease in corals: Bacteriophages in black
827 band disease. Environmental Microbiology doi:10.1111/1462-2920.14432.

828 70. Gustavsen JA, Winget DM, Tian X, Suttle CA. 2014. High temporal and spatial diversity in
829 marine RNA viruses implies that they have an important role in mortality and structuring
830 plankton communities. Frontiers in Microbiology 5:703.

831 71. Montalvo-Proaño J, Buerger P, Weynberg KD, Oppen MJH van. 2017. A PCR-based assay
832 targeting the major capsid protein gene of a dinorna-like ssRNA virus that infects coral
833 photosymbionts. Frontiers in Microbiology 8:1665.

834 72. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL,
835 Solymos P, Stevens MHH, Wagner H. 2015. Vegan: Community Ecology Package.

836 73. Hijmans RJ. 2015. Geosphere: Spherical Trigonometry.

837 74. Johnson LS, Eddy SR, Portugaly E. 2010. Hidden Markov model speed heuristic and
838 iterative HMM search procedure. BMC Bioinformatics 11:431.

839 75. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert
840 M, Soding J, Thompson JD, Higgins DG. 2014. Fast, scalable generation of high-quality
841 protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology
842 7:539–539.

843 76. Jia Z, Ishihara R, Nakajima Y, Asakawa S, Kimura M. 2007. Molecular characterization of
844 T4-type bacteriophages in a rice field. Environmental Microbiology 9:1091–1096.

845 77. López-Bueno A, Tamames J, Velázquez D, Moya A, Quesada A, Alcamí A. 2009. High
846 diversity of the viral community from an Antarctic lake. Science 326:858–861.

847 78. Comeau AM, Arbiol C, Krisch HM. 2010. Gene network visualization and quantitative
848 synteny analysis of more than 300 marine T4-like phage scaffolds from the GOS
849 metagenome. Molecular Biology and Evolution 27:1935–1944.

850 79. Liu J, Wang G, Wang Q, Liu J, Jin J, Liu X. 2012. Phylogenetic diversity and assemblage of
851 major capsid genes (*g23*) of T4-type bacteriophages in paddy field soils during rice
852 growth season in Northeast China. Soil Science and Plant Nutrition 58:435–444.

853 80. Bellas CM, Anesio AM. 2013. High diversity and potential origins of T4-type
854 bacteriophages on the surface of Arctic glaciers. Extremophiles 17:861–870.

855 81. Butina TV, Belykh OI, Potapov SA, Sorokovikova EG. 2013. Diversity of the major capsid
856 genes (g23) of T4-like bacteriophages in the eutrophic Lake Kotokel in East Siberia, Russia.
857 Archives of Microbiology 195:513–520.

858 82. Larsson A. 2014. AliView: A fast and lightweight alignment viewer and editor for large
859 datasets. Bioinformatics 30:3276–3278.

860 83. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. 2009. trimAl: A tool for automated
861 alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972–1973.

862 84. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2–approximately maximum-likelihood
863 trees for large alignments. PloS ONE 5:e9490.

85. Stamatakis A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313.

86. Darriba D, Taboada GL, Doallo R, Bangor D. 2011. ProtTest 3: Fast selection of best-fit models of protein evolution. Bioinformatics 27:1164–1165.

87. Faith DP. 1992. Conservation evaluation and phylogenetic diversity. Biological Conservation 61:1–10.

88. Kembel SW, Cowan PD, Helmus MR, Cornwell WK, Morlon H, Ackerly DD, Blomberg SP, Webb CO. 2010. Picante: R tools for integrating phylogenies and ecology. Bioinformatics 26:1463–1464.

89. Yu G, Smith D, Zhu H, Guan Y, Lam TT-Y. 2017. ggtree: An R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods in Ecology and Evolution 8:28–36.

90. R Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

91. Harrison PJ, Fulton JD, Taylor FJR, Parsons TR. 1983. Review of the biological oceanography of the Strait of Georgia: Pelagic environment. Canadian Journal of Fisheries and Aquatic Sciences 40:1064–1094.

92. Allen S, Wolfe M. 2013. Hindcast of the timing of the spring phytoplankton bloom in the Strait of Georgia, 1968-2010. Progress in Oceanography 115:6–13.

883 93. Velthuis AJW te. 2014. Common and unique features of viral RNA-dependent
884 polymerases. Cellular and Molecular Life Sciences 71:4403–4420.

885 94. Suttle CA. 2007. Marine viruses - major players in the global ecosystem. Nature Reviews
886 Microbiology 5:801–812.

887 95. Chénard C, Suttle CA. 2008. Phylogenetic diversity of sequences of cyanophage
888 photosynthetic gene *psba* in marine and freshwaters. Applied and Environmental
889 Microbiology 74:5317–5324.

890 96. Finke J, Winget D, Chan A, Suttle C. 2017. Variation in the genetic repertoire of viruses
891 infecting *Micromonas pusilla* reflects horizontal gene transfer and links to their
892 environmental distribution. Viruses 9:116.

893 97. Kellogg CA, Paul JH. 2002. Degree of ultraviolet radiation damage and repair capabilities
894 are related to G+ C content in marine vibriophages. Aquatic microbial ecology 27:13–20.

895 98. Bragg JG, Chisholm SW. 2008. Modeling the fitness consequences of a cyanophage-
896 encoded photosynthesis gene. PLoS ONE 3:e3550.

897 99. Moniruzzaman M, Wurch LL, Alexander H, Dyhrman ST, Gobler CJ, Wilhelm SW. 2017.
898 Virus-host relationships of marine single-celled eukaryotes resolved from
899 metatranscriptomics. Nature Communications 8:16054.

900 100. Short SM, Short CM. 2009. Quantitative PCR reveals transient and persistent algal
901 viruses in Lake Ontario, Canada. Environmental Microbiology 11:2639–2648.

902 101. Rozon RM, Short SM. 2013. Complex seasonality observed amongst diverse
903 phytoplankton viruses in the Bay of Quinte, an embayment of Lake Ontario. Freshwater
904 Biology 58:2648–2663.

905 102. Needham DM, Chow C-ET, Cram JA, Sachdeva R, Parada A, Fuhrman JA. 2013. Short-
906 term observations of marine bacterial and viral communities: Patterns, connections and
907 resilience. The ISME Journal 7:1274–1285.

908 103. Holmes EC. 2010. The RNA Virus Quasispecies: Fact or Fiction? Journal of Molecular
909 Biology 400:271–273.

910 104. Domingo E, Sheldon J, Perales C. 2012. Viral quasispecies evolution. Microbiology and
911 Molecular Biology Reviews 76:159–216.

912 105. Milo R, Jorgensen P, Moran U, Weber G, Springer M. 2010. BioNumbers the database of
913 key numbers in molecular and cell biology. Nucleic Acids Research 38:D750–D753.

914 106. Weitz JS, Hartman H, Levin SA. 2005. Coevolutionary arms races between bacteria and
915 bacteriophage. Proceedings of the National Academy of Sciences 102:9535–9540.

916 107. Dutilh BE, Cassman N, McNair K, Sanchez SE, Silva GGZ, Boling L, Barr JJ, Speth DR,
917 Seguritan V, Aziz RK, Felts B, Dinsdale EA, Mokili JL, Edwards RA. 2014. A highly abundant
918 bacteriophage discovered in the unknown sequences of human faecal metagenomes.
919 Nature Communications 5:4498.

920 108. Renzette N, Pokalyuk C, Gibson L, Bhattacharjee B, Schleiss MR, Hamprecht K,
921 Yamamoto AY, Mussi-Pinhata MM, Britt WJ, Jensen JD, Kowalik TF. 2015. Limits and

patterns of cytomegalovirus genomic diversity in humans. Proceedings of the National Academy of Sciences 112:E4120–E4128.

109. López-Bueno A, Rastrojo A, Peiró R, Arenas M, Alcamí A. 2015. Ecological connectivity shapes quasispecies structure of RNA viruses in an Antarctic lake. Molecular Ecology 24:4812–4825.

110. Storesund JE, Erga SR, Ray JL, Thingstad TF, Sandaa R-A. 2015. Top-down and bottom-up control on bacterial diversity in a western Norwegian deep-silled fjord. FEMS Microbiology Ecology 91:fiv076.

111. Tarutani K, Nagasaki K, Yamaguchi M. 2000. Viral impacts on total abundance and clonal composition of the harmful bloom-forming phytoplankton *Heterosigma akashiwo*. Applied and Environmental Microbiology 66:4916–4920.

112. Nagasaki K, Yamaguchi M. 1997. Isolation of a virus infectious to the harmful bloom causing microalga *Heterosigma akashiwo* (Raphidophyceae). Aquatic Microbial Ecology 13:135–140.

113. Harvey EL, Menden-Deuer S. 2012. Predator-induced fleeing behaviors in phytoplankton: A new mechanism for harmful algal bloom formation? PLoS ONE 7:e46438.

114. Schroeder DC, Oke J, Hall M, Malin G, Wilson WH. 2003. Virus succession observed during an *Emiliania huxleyi* bloom. Applied and Environmental Microbiology 69:2484–2490.

941 115. Emerson JB, Thomas BC, Andrade K, Heidelberg KB, Banfield JF. 2013. New
942 approaches indicate constant viral diversity despite shifts in assemblage structure in an
943 Australian hypersaline lake. Applied and Environmental Microbiology 79:6755–6764.

944 116. Short CM, Rusanova O, Short SM. 2010. Quantification of virus genes provides
945 evidence for seed-bank populations of phycodnaviruses in Lake Ontario, Canada. The ISME
946 Journal 5:810–821.

947 117. Zhong X, Jacquet S. 2014. Differing assemblage composition and dynamics in T4-like
948 myophages of two neighbouring sub-alpine lakes. Freshwater Biology 59:1577–1595.

949 118. Brum JR, Ignacio-Espinoza JC, Roux S, Doulcier G, Acinas SG, Alberti A, Chaffron S,
950 Cruaud C, Vargas C de, Gasol JM, Gorsky G, Gregory AC, Guidi L, Hingamp P, Iudicone D, Not
951 F, Ogata H, Pesant S, Poulos BT, Schwenck SM, Speich S, Dimier C, Kandels-Lewis S, Picheral
952 M, Searson S, Bork P, Bowler C, Sunagawa S, Wincker P, Karsenti E, Sullivan MB. 2015.
953 Patterns and ecological drivers of ocean viral communities. Science 348:1261498.

954 119. Koeppel AF, Wu M. 2013. Species matter: The role of competition in the assembly of
955 congeneric bacteria. The ISME Journal 8:531–540.

956 120. Lee CK, Herbold CW, Polson SW, Wommack KE, Williamson SJ, McDonald IR, Cary SC.
957 2012. Groundtruthing next-gen sequencing for microbial ecology–biases and errors in
958 community structure estimates from PCR amplicon pyrosequencing. PLoS ONE 7:e44224.

959 121. Pinto AJ, Raskin L. 2012. PCR biases distort bacterial and archaeal community
960 structure in pyrosequencing datasets. PLoS ONE 7:e43093.

961 122. Shade A, Jones SE, Caporaso JG, Handelsman J, Knight R, Fierer N, Gilbert JA. 2014.
962 Conditionally rare taxa disproportionately contribute to temporal changes in microbial
963 diversity. mBio 5:e01371–14.

964 123. Qiu X, Wu L, Huang H, McDonel PE, Palumbo AV, Tiedje JM, Zhou J. 2001. Evaluation of
965 PCR-Generated chimeras, mutations, and heteroduplexes with 16S rRNA gene-based
966 cloning. Applied and Environmental Microbiology 67:880–887.

967

968 # Tables

969 Table 1: PCR parameters used in this study.

Marker gene	Target	Primer names	PCR initial	PCR denaturation	Annealing temperature	Extension	Cycles	Final extension	Reference
gp23	T4-like virus	MZIA1bis and MZIA6	94°C for 90 s	94°C for 45 s	50°C	72°C for 45 s	35	5 min at 72°C	42
RdRp	Marnavirus -like viruses	MPL-2F and MPL-2R	94°C for 75 s	94°C for 45 s	43°C	72°C for 60 s	40	9 min at 72°C	43
18S rRNA gene	Eukaryotes	Euk1209f and Uni1392r	94°C for 75 s	94°C for 1 min	65°C touchdown for 10 cycles followed by 55°C	72°C for 60 s	10 + 20	9 min at 72°C	57
16S rRNA gene	Bacteria	341F and 907R	94°C for 75 s	94°C for 1 min	64°C, 12cycles followed by 54°C	72°C for 60 s	12 + 25	10 min at 72°C	59 ; 58

970

Figures and Figure Legends

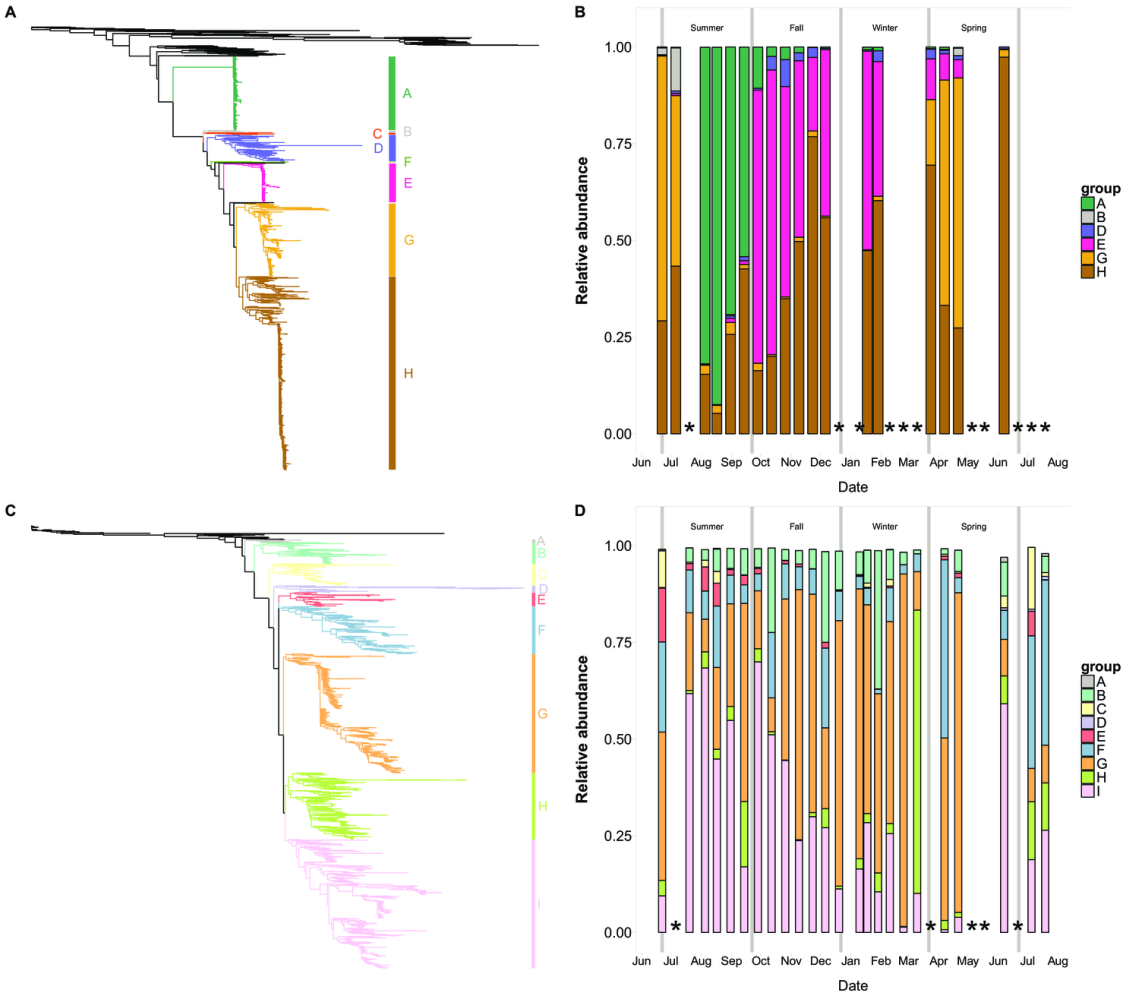


Fig. 1: Maximum likelihood RAxML phylogenetic trees and barplots of closely-related phylogenetic groups of OTUs. A) Tree of marnavirus-like virus *RdRp* sequences including reference sequences and OTUs generated in this study. Outgroup is virus Equine rhinitis B virus (*Picornaviridae*). B) Barplot of the relative abundances of marnavirus-like virus phylogenetic groups over time. C) Tree of T4-like virus major capsid protein sequences including reference sequences and OTUs generated in this study. Outgroup is Enterobacteria phage T4. D) Barplot of the relative abundances of T4-like virus phylogenetic groups from over time. Grey vertical lines indicate seasonal boundaries. X's indicate missing or removed samples. More detailed phylogenetic trees are available in Supplemental figures.

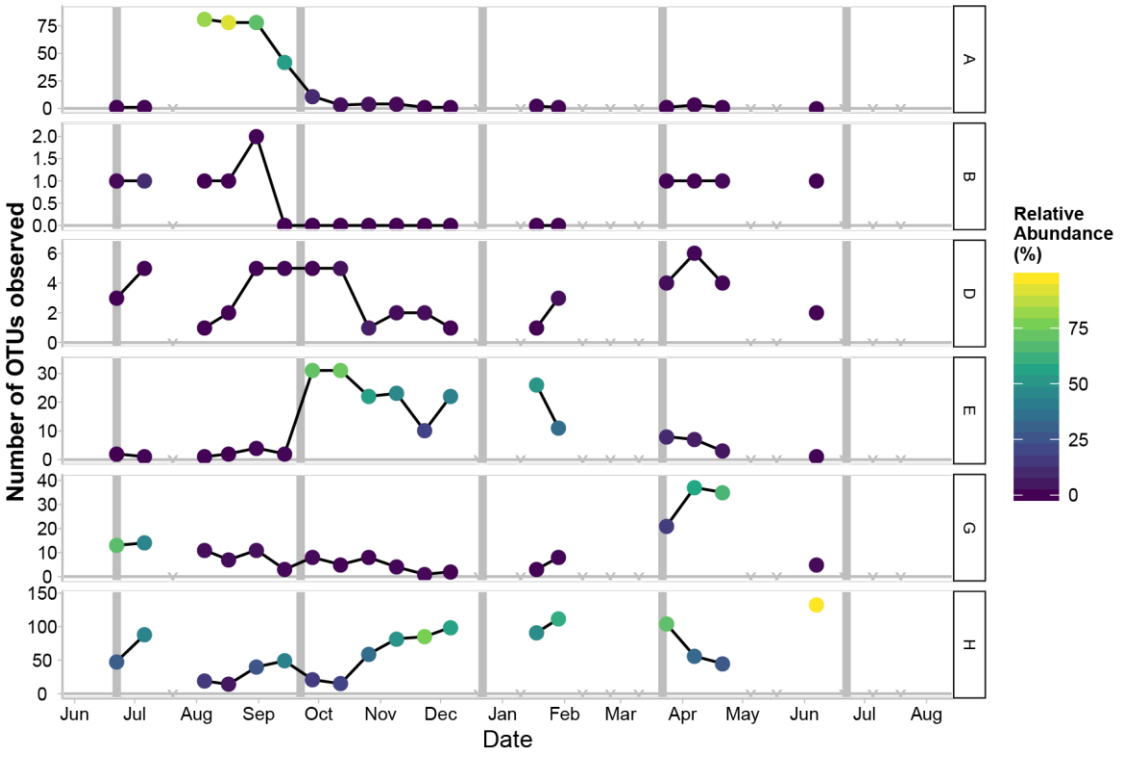
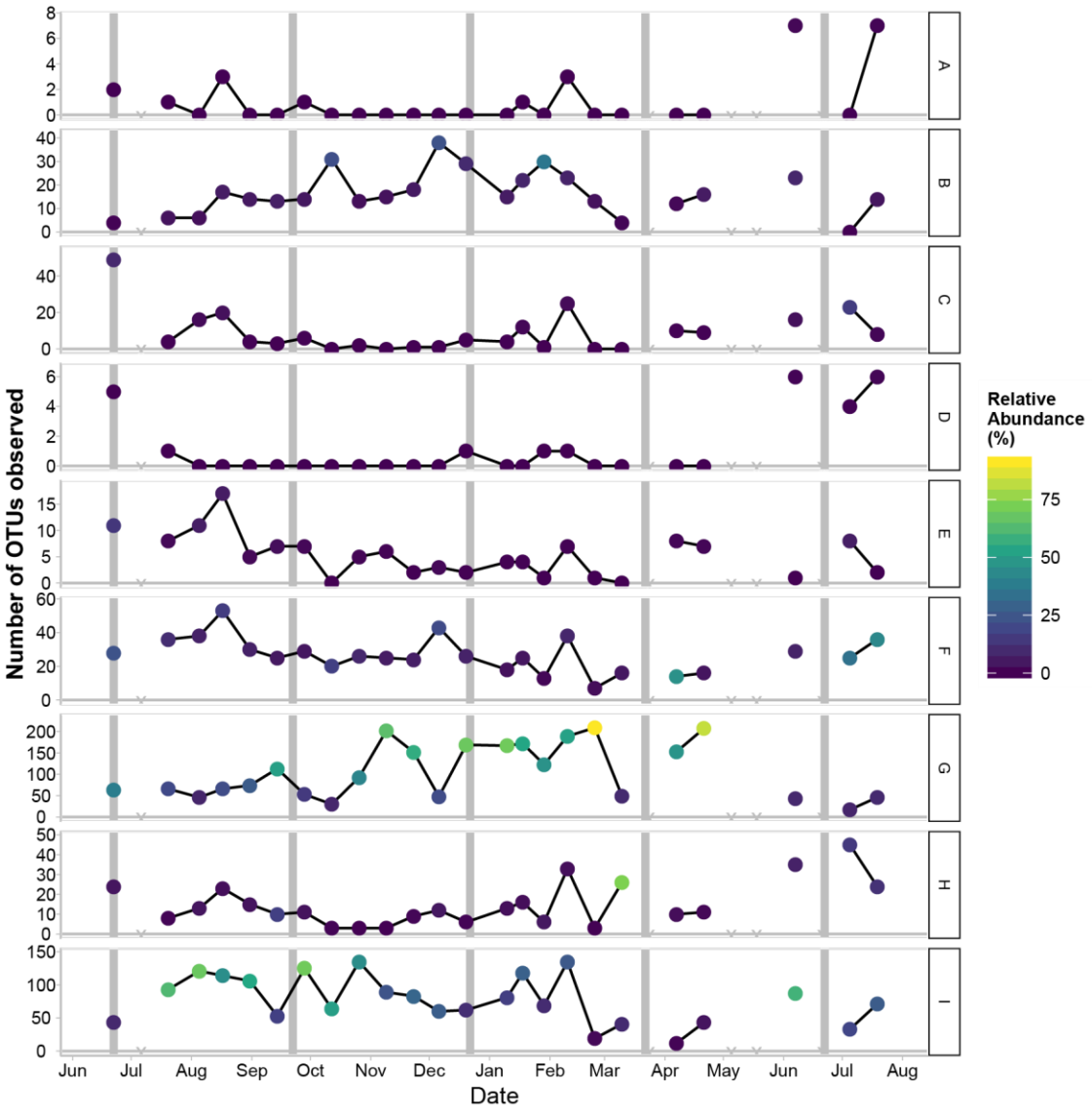


Fig. 2: Richness of observed marnavirus-like viral OTUs (95% amino-acid similarity). X's indicate missing or removed samples.



985

986 *Fig. 3: Richness of observed T4-like virus OTUs (95% amino-acid similarity). X's indicate*
987 *missing or removed samples.*

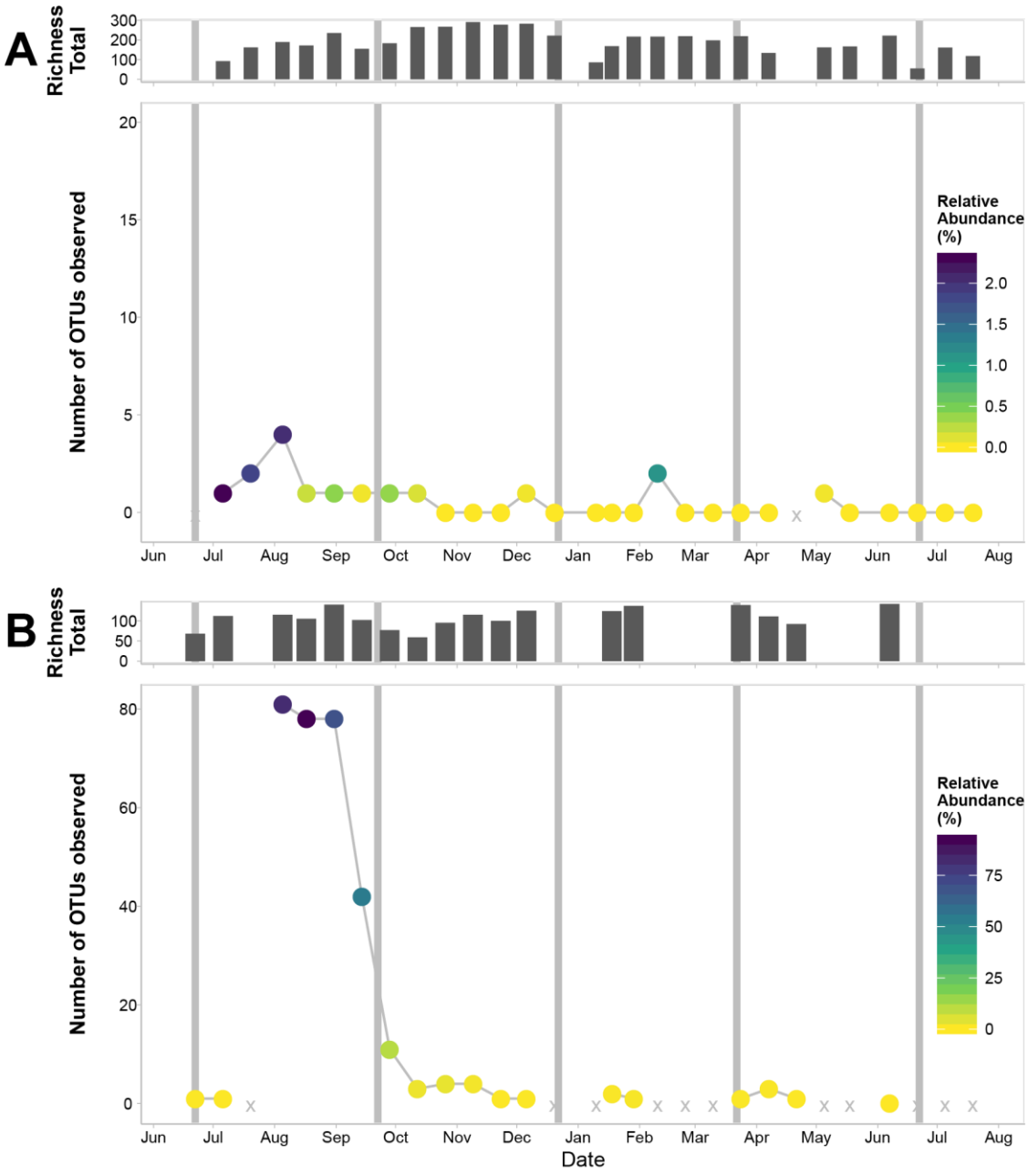


Fig. 4: Marine marnavirus-like viral Group A compared to OTUs classified as raphidophytes over time. A) Relative abundance of OTUs (97%) classified as raphidophytic OTUs over time. Richness of all 18S OTUs at each time point is plotted above the richness of the raphidophytes.. B) Relative abundance of marnavirus-like virus Group A OTUs (95% amino acid) over time. Richness of all marnavirus-like virus OTUs is plotted above the richness of the Group A OTUs. Grey vertical lines indicate boundaries between seasons. X's indicate missing or removed samples.

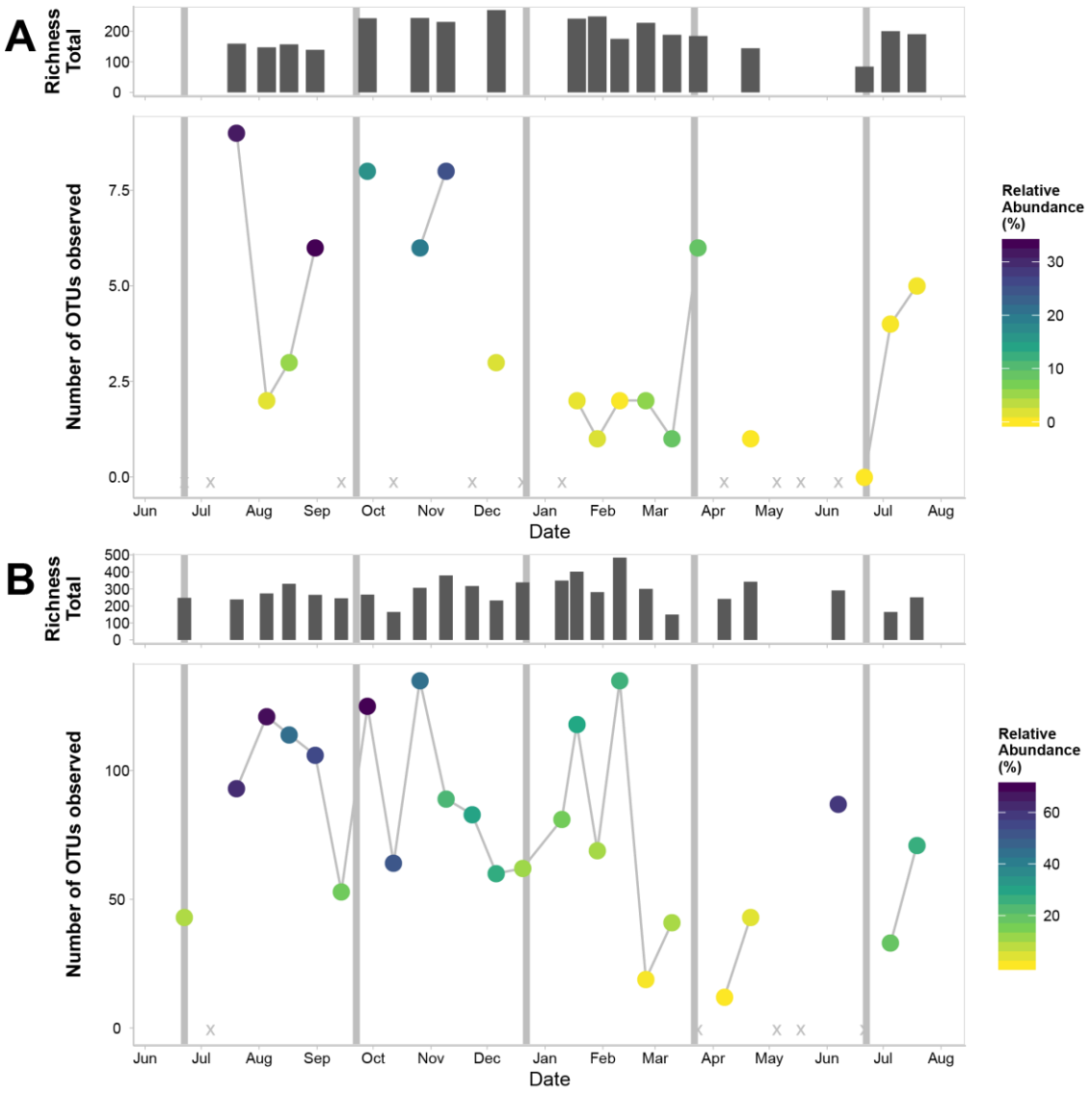
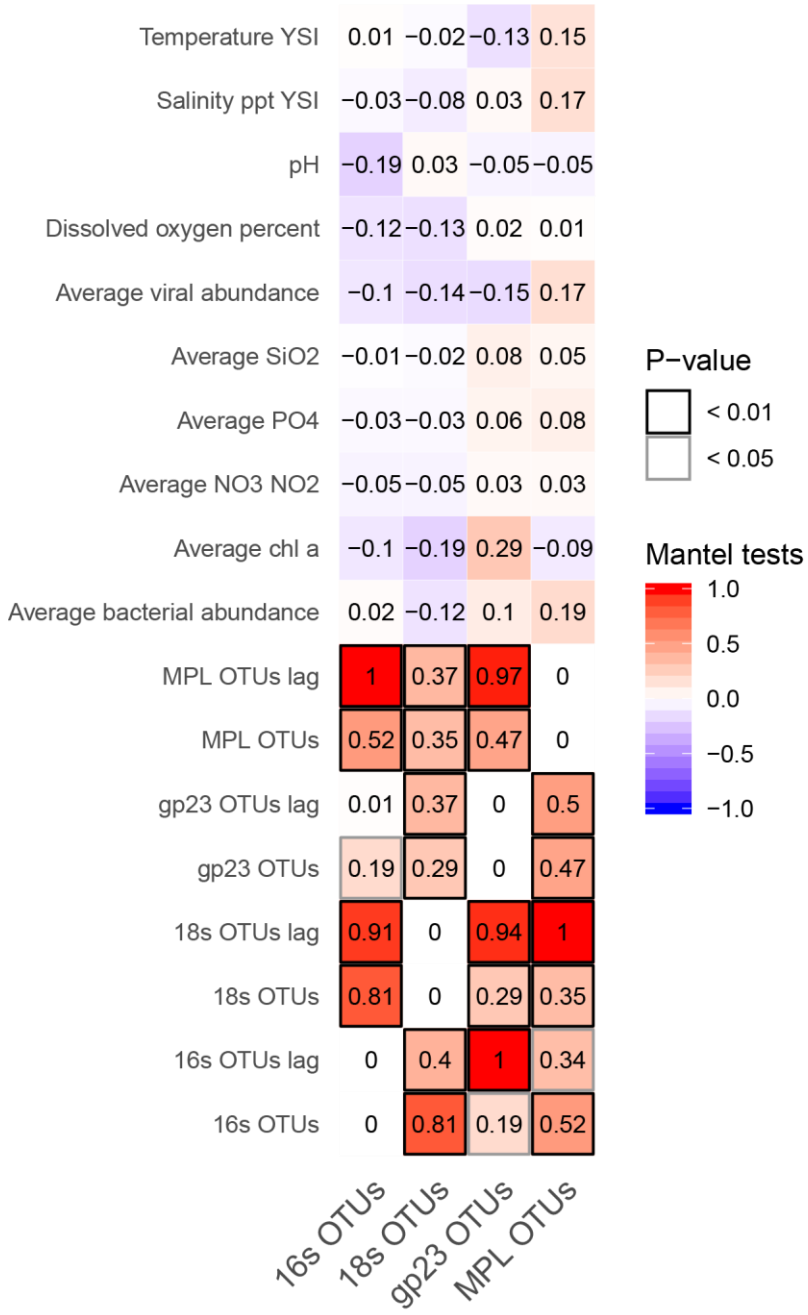


Fig. 5: T4-like virus Group I compared to bacterial OTUs classified as cyanobacteria over time. A) Relative abundance of bacterial OTUs (97%) classified as cyanobacterial OTUs over time. Richness of all 16S OTUs at each time point is plotted above the richness of the cyanobacteria. B) Relative abundance of T4-like virus Group I OTUs (95% amino acid) over time. Richness of all T4-like virus OTUs is plotted above the richness of the Group I OTUs. Grey vertical lines indicate boundaries between seasons. X's indicate missing or removed samples.



1004

1005 *Fig. 6: Mantel tests among community similarity matrices and distance matrices of*
1006 *environmental data.*

1007