Role of Phylogenetic Structure in the Dynamics of Coastal Viral Assemblages

Julia A. Gustavsen,^a Curtis A. Suttle^{a,b,c,d}

^aDepartment of Earth, Ocean and Atmospheric Sciences, University of British Columbia, Vancouver, BC, Canada

^bDepartment of Botany, University of British Columbia, Vancouver, BC, Canada

^cDepartment of Microbiology and Immunology, University of British Columbia, Vancouver, BC, Canada

^dInstitute for the Oceans and Fisheries, University of British Columbia, Vancouver, BC, Canada

ABSTRACT Marine microbes, including viruses, are an essential part of the marine ecosystem, forming the base of the food web and driving biogeochemical cycles. Within this system, the composition of viral assemblages changes markedly with time, and some of these changes are repeatable through time; however, the extent to which these dynamics are reflected within versus among evolutionarily related groups of viruses is largely unexplored. To examine these dynamics, changes in the composition of two groups of ecologically important viruses and communities of their potential hosts were sampled every 2 weeks for 13 months at a coastal site in British Columbia, Canada. We sequenced two marker genes for viruses—the gene encoding the major capsid protein of T4-like phages and their relatives (*gp23*) and the RNA-dependent RNA polymerase (*RdRp*) gene of marnavirus-like RNA viruses—as well as marker genes for their bacterial and eukaryotic host communities, the genes encoding 16S rRNA and 18S rRNA. There were strong lagged correlations between viral diversity and community similarity of putative hosts, implying that the viruses influenced the composition of the host communities. The results showed that for both viral assemblages, the dominant clusters of phylogenetically related viruses shifted over time, and this was correlated with environmental changes. Viral clusters contained many ephemeral taxa and few persistent taxa, but within a viral assemblage, the ephemeral and persistent taxa were closely related, implying ecological dynamics within these clusters. Furthermore, these dynamics occurred in both the RNA and DNA viral assemblages surveyed, implying that this structure is common in natural viral assemblages.

IMPORTANCE Viruses are major agents of microbial mortality in marine systems, yet little is known about changes in the composition of viral assemblages in relation to those of the microbial communities that they infect. Here, we sampled coastal seawater every 2 weeks for 1 year and used high-throughput sequencing of marker genes to follow changes in the composition of two groups of ecologically important viruses, as well as the communities of bacteria and protists that serve as their respective hosts. Different subsets of genetically related viruses dominated at different times. These results demonstrate that although the genetic composition of viral assemblages is highly dynamic temporally, for the most part the shuffling of genotypes occurs within a few clusters of phylogenetically related viruses. Thus, it appears that even in temperate coastal waters with large seasonal changes, the highly dynamic shuffling of viral genotypes occurs largely within a few subsets of related individuals.

KEYWORDS virus, phylogeny, coastal, time series, bacteria, phytoplankton, seed bank, 18S rRNA, 16S rRNA, dynamics, *Picornavirales*, *Myoviridae*

Citation Gustavsen JA, Suttle CA. 2021. Role of phylogenetic structure in the dynamics of coastal viral assemblages. *Appl Environ Microbiol* 87:e02704-20. <https://doi.org/10.1128/AEM.02704-20>.

Editor Karyn N. Johnson, University of Queensland

Copyright © 2021 American Society for Microbiology. All Rights Reserved.

Address correspondence to Curtis A. Suttle, suttle@science.ubc.ca.

Received 10 November 2020

Accepted 16 March 2021

Accepted manuscript posted online
19 March 2021

Published 11 May 2021

Understanding diversity, its maintenance, and its drivers is a foundation in ecology. For microbial systems, there has been extensive exploration and discussion about the mechanisms affecting diversity (1). Many temporal studies on microbial diversity come from the marine milieu, where, it has been argued, community composition is driven mainly by environmental factors (2–5). The role of viruses as obligate pathogens, often with high host specificity, implies that they are important drivers of host composition and diversity (6), yet the role of viruses as drivers of marine microbial diversity and the factors affecting viral diversity remain relatively unexplored.

Marine viruses have repeatable seasonal dynamics as revealed by measures of abundance, infectious units, and taxonomic composition. Seasonal studies in coastal waters report that viral abundances are higher in summer than in winter (7, 8). As well, other multiyear time series data show that viral production and viral abundances are highest in early spring and summer (9, 10). Moreover, viral dynamics can be associated with putative hosts (10) and specific phylogenetically related groups of viruses (11–13). As well, viruses infecting cyanobacteria show seasonal dynamics (14, 15): cyanophages from the same season are genetically more similar across years than across seasons within the same year (16, 17) and are more stable in winter than in summer and spring (18). Another time series study from coastal California revealed strong correlations between common myovirus genotypes and bacterial genotypes when microdiversity within operational taxonomic units (OTUs) was examined (19). Furthermore, Arkhipova et al. (20) examined metagenomically assembled viral genomes from a lake over a year and found peaks in relative abundance before, during, and after peaks in host abundance.

Unlike cellular organisms, viruses have no shared universal genes that can be used as proxies for presence and for overall phylogeny, such as the 16S rRNA genes for prokaryotes and the 18S rRNA genes for eukaryotes (21, 22). Nevertheless, several well-conserved marker genes, such as those encoding DNA polymerase (Pol) (23), the major capsid protein (24), and RNA-dependent RNA polymerase (encoded by *RdRp*) (25), have highly conserved and variable regions, allowing the presence, diversity, and phylogeny of specific groups of viruses to be assessed (26–28). Moreover, the conserved genes for the major capsid protein are good surrogates for whole genomes, since their phylogeny is congruent with the phylogeny of the whole viral genomes (24), and the *RdRp* gene enables grouping by phylogeny that is congruent with the current classification of RNA viral families (25, 29).

Marker genes can be used to examine genetic diversity and phylogenetic relationships within groups of related viruses in natural systems. In turn, phylogenetic relatedness can be correlated to ecological relatedness in plants and animals (30, 31), and phylogenetic patterns in distribution and abundance occur in bacterial communities (32, 33), yet few studies have examined phylogenetic patterns in viral and putative host communities. An exception is work in the subtropical Atlantic showing differences in the genetic composition of viruses associated with seasonal changes in the stratification and differences in the composition of a phylogenetic group of cyanophages (34). Similarly, a study in the coastal waters of California showed repeatable shifts in relative dominance among phylogenetically related cyanophages (18). These types of dynamics are captured by the Royal Family framework, in which shifts among dominant phages often occur within a group of closely related phages (35).

In order to gain more insight into the temporal dynamics within and among phylogenetic clusters of viruses, we sampled a coastal environment every 2 weeks for 13 months to follow changes in the relative abundances of viral OTUs within a broad evolutionary group of RNA viruses infecting protists (marnaviruses) and a diverse group of phages infecting bacteria (T4-like phages). These groups of viruses were targeted because of their persistence and ecological importance in coastal waters, where they are pathogens of major groups of prokaryotic and eukaryotic primary producers (25, 36–43), and because of the availability of well-documented PCR primers for marker genes in these viruses (24, 37). We hypothesized that within each group of viruses, close phylogenetic relatives would be abundant at the same time, and that within

those groups of close relatives, a few OTUs would be persistent but most would be ephemeral. Such a finding would be consistent with a “seed bank” model, in which there are many rare viral OTUs that can potentially become abundant, but most remain rare (34, 44). As well, shuffling of dominant OTUs among close relatives would provide evidence for the proposed Royal Family framework (35).

In addition, we followed the OTU dynamics of microbial eukaryotes and prokaryotes by sequencing amplified gene fragments for 18S rRNA and 16S rRNA, respectively, in order to capture changes in the communities harboring host cells. Our data reveal that most viral OTUs were ephemeral but were related to a much smaller number of viruses that were relatively persistent. Nonetheless, a small subset of the ephemeral viruses became abundant in association with environmental changes and fluctuations in the communities of potential hosts. Thus, different phylogenetic clusters had the highest relative abundances at different times, and the within-cluster variation could be associated with putative hosts. These results are consistent with the concept of a seed bank of different virus genotypes, from which individual taxa are favored depending on shifts in the host communities and environmental conditions.

RESULTS

Chlorophyll *a* and nutrient concentrations. Chlorophyll *a* (Chl *a*) concentrations varied over time; the maximum was observed during a eukaryotic phytoplankton bloom (mean $\pm$ standard error [SE], $46.5 \pm 10.17 \mu\text{g liter}^{-1}$ in late June 2011) (see Fig. S1 in the supplemental material) and the second highest during a spring bloom in late April 2011 (mean $\pm$ SE, $5.88 \pm 1.38 \mu\text{g liter}^{-1}$). The minimum Chl *a* value of $0.05 \mu\text{g liter}^{-1}$ occurred in May, and values remained below $1 \mu\text{g liter}^{-1}$ from September to March.

Nutrient concentrations were also highly dynamic, ranging from $6.1 \mu\text{M}$ to $67.3 \mu\text{M}$ for silicate, from $<0.1 \mu\text{M}$ to $2.3 \mu\text{M}$ for phosphate, and from $<0.1 \mu\text{M}$ to $27.7 \mu\text{M}$ for nitrate plus nitrite (Fig. S1). Overall, nutrient concentrations were high and stable over the winter and declined in April, with a large increase in silicate levels commencing in May and remaining through to the end of the study in July.

Viral and bacterial abundances. During the time series, viral abundances ranged from 5.41×10^6 to 4.70×10^7 particles ml^{-1} , while bacterial abundances were generally about an order of magnitude lower, ranging from 6.59×10^5 to 4.43×10^6 cells ml^{-1} (Fig. S1). Abundances were lowest from late fall to early spring.

Few viral and cellular OTUs were shared across all time points. Amplicon reads representing marnavirus-like (MVL) viruses (*RdRp*) and T4-like viruses (*gp23*) were translated into amino acids, and the reads were normalized by rarefaction to the library with the fewest reads. The 16S and 18S rRNA 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 59 to 142 OTUs per time point (Fig. S2) and an average of 6% of OTUs shared among all time points. For the T4-like viruses, there were a total of 1,737 OTUs at 95% sequence similarity, with a minimum of 149 and a maximum of 484 OTUs per time point (Fig. S2). On average, 6% of the T4-like virus OTUs were shared among all time points. There were 813 bacterial OTUs observed (97% sequence similarity), with an average of 10% shared across all time points. The lowest number of OTUs seen per time point was 84, and the highest was 269.

Phylum-level classification of the OTUs showed a dominance of bacteria in the phyla *Proteobacteria*, *Bacteroidetes*, and *Cyanobacteria*, although there were large fluctuations in relative abundances across time points (Fig. S3). In the eukaryotic community, there were a total of 1,115 OTUs (97% sequence similarity), with a minimum of 55 and a maximum of 298 OTUs (Fig. S2), and a mean of 6% of OTUs shared among all time points. Most sequences belonged to four phylum-level classifications, with members of the SAR (Stramenopiles, Alveolates, and Rhizaria) group dominating, followed by Opisthokonta, Haptophyta, and Cryptophyceae (Fig. S4). Rarefaction curves for individual samples did not flatten (“saturate”), indicating that not all OTUs were sequenced in each sample, but when the data were combined, the curves saturated, indicating that overall, the community of OTUs was captured (Fig. S5).

FIG 1 Maximum-likelihood RAXML phylogenetic trees and bar plots of closely-related phylogenetic clusters of OTUs. (A) Tree of marnavirus-like virus *RdRp* gene sequences, including reference sequences and OTUs generated in this study. The outgroup is equine rhinitis B virus (*Picornaviridae*). (B) Bar plot of the relative abundances of marnavirus-like virus phylogenetic clusters over time. (C) Tree of T4-like virus major capsid protein sequences, including reference sequences and OTUs generated in this study. The outgroup is enterobacterial phage T4. (D) Bar plot of the relative abundances of T4-like virus phylogenetic clusters over time. Vertical gray lines indicate seasonal boundaries. Asterisks indicate missing or removed samples. More-detailed phylogenetic trees are available in the supplemental material.

FIG 1 (Continued)

The dominance of major phylogenetic clusters of viruses shifted gradually with time. Viral OTUs were placed in a phylogenetic context and their temporal dynamics examined (Fig. 1; also Fig. S6 and S7). Well-defined and supported phylogenetic clusters (clusters A to H) of marnavirus-like viruses (Fig. 1A and B) showed strong temporal dynamics and differed by season (adonis $R=0.503$; $P=0.001$). Sequences in cluster H, which included viruses infecting the diatoms *Chaetoceros* sp. and *Rhizosolenia* sp., were consistently present, with their highest relative abundance in late November. Sequences in cluster A were also always detected, except at one time point in early June, and they were most abundant in August and September. Cluster A included many sequences from other environmental gene surveys, as well as from a virus that infects the raphidophyte *Heterosigma akashiwo*. October to February was dominated by OTUs in clusters E and H. The structure of the MVL OTU phylogenetic clusters closely mirrored the structure of the top 20 OTUs found over time (Fig. S8), demonstrating that this assemblage comprised a few dominant OTUs.

The T4-like virus OTUs were also placed in a phylogenetic context and categorized into clusters of related OTUs (Fig. 1C). In the early fall, cluster I dominated the assemblage, while cluster G dominated in the late fall and early winter; both clusters included sequences from viral isolates infecting cyanobacteria. In January, 35.7% of the relative abundance of T4-like viruses consisted of cluster B, which contains no known isolates. One time point in March was dominated by cluster H, and the next time point showed a large shift in assemblage composition whereby clusters F and G dominated. Unlike the MVL assemblage, the T4-like virus assemblage had very different patterns for the top 20 OTUs and phylogenetic clusters over time (Fig. S9). The T4-like virus communities showed small differences by season (adonis $R=0.23$; $P=0.001$).

Phylogenetic clusters of viruses were defined by many ephemeral OTUs and a few dominant OTUs. Within phylogenetic clusters, the diversity of OTUs varied over time in the viral assemblages (Fig. 2 and 3; Fig. S10 and S11). For MVL OTUs in cluster A, diversity peaked in August and gradually decreased from fall into winter (Fig. 2; Fig. S10). Clusters D, E, and G showed variable patterns over time, and cluster B showed only one spike in diversity, in September. Conversely, for cluster H, the diversity of OTUs fluctuated, but diversity remained relatively high (above 1.0) in all observations. Cluster H constituted a large proportion of the overall OTUs (averaging 47% of the total), which affected the dynamics of diversity observed within MVLs.

For the T4-like viruses (Fig. 3; Fig. S11), clusters C, E, and H had similar dynamics, with high diversity within each cluster in June to August, lower diversity from August until February, and a small peak in diversity in February. Clusters B, G, and I continued throughout the time series with generally small fluctuations between time points. Clusters A and D displayed higher diversity in the summer than in the rest of the year, and cluster A had a small peak in February.

There were persistent and ephemeral OTUs in MVL and T4-like viruses (Fig. S10 and S11). Furthermore, many ephemeral OTUs were closely related to the persistent OTUs. There were many more OTUs for T4-like viruses than for MVL viruses; however, both fluctuated over time. For example, 0.9% of MVL OTUs and 0.8% of T4-like viral OTUs were found at 90% of the time points. For the T4-like viruses, clusters A, B, G, and F had persistent and ephemeral OTUs, whereas cluster C did not. For MVL viruses, there were fewer OTUs overall and fewer OTUs that persisted over time. MVL cluster A had one persistent OTU and other OTUs that were detected at four to six of the sampling points.

Changes in viral OTUs lagged behind changes in cellular OTUs. The assemblage similarity of the T4-like viruses had a strong lagged negative correlation to the richness of the bacterial community (see Fig. 6; Fig. S13). The MVL assemblage was much more strongly correlated with changes in the composition of the T4-like virus OTUs and the OTUs of their putative hosts when lagged in time than when directly compared. Viral abundance was negatively correlated with bacterial community similarity.

Raphidophytes and marnavirus-like viruses. Cluster A (Fig. 1A), which includes HaRNAV, an RNA virus infecting the raphidophyte *Heterosigma akashiwo*, increased in

FIG 2 Shannon diversity of observed marnavirus-like viral OTUs (95% amino acid similarity). Each row shows data for a phylogenetic cluster as defined in Fig. 1A. X indicates missing or removed samples.

diversity after the diversity of eukaryotic sequences classified as raphidophytes increased (Fig. 4) (Spearman rank correlation over the entire time series, 0.43; $P=0.07$). There were further peaks in the diversity of raphidophyte sequences, but they did not coincide with increases in the diversity of marnavirus-like viral cluster A. Of all the OTUs in cluster A, OTU 1 had the highest relative abundance, and when aligned with other closely related sequences (Fig. S12), these sequences showed a change from the amino acid aspartate to glutamate in the palm region of *RdRp* protein (45).

Cyanobacteria and T4-like virus cluster I. Comparison of the T4-like virus cluster I, containing cyanophage isolates, to cyanobacterial OTUs (Fig. 5) showed that the diversity of viruses increased in the fall after peaks in the relative abundance of cyanobacterial OTUs (correlation over the entire time series, 0.56; $P=0.03$). The cyclical lags in the diversity of putative cyanophages relative to cyanobacteria continued through the time points up to the spring diatom bloom (8 April), when there was a lag before the diversity of a different putative cyanophage in cluster I increased, showing succession in the viral assemblage.

Changes in taxonomic composition among viruses and cells were coupled. Mantel tests were used to examine concurrent changes in the relative abundances of OTUs from viruses and their putative hosts in relation to environmental changes over time (Fig. 6). The overall bacterial and eukaryotic community compositions fluctuated strongly together. The changes in marnavirus-like assemblages were predicted most strongly by changes in the eukaryotic community, with a lag of 2 weeks, and less strongly by the bacterial and T4-like viral assemblages. The T4-like viruses showed

FIG 3 Shannon diversity of observed T4-like virus OTUs (95% amino acid similarity). Each row shows data for a phylogenetic cluster as defined in Fig. 1C. X indicates missing or removed samples.

the strongest associations with bacterial communities, with a 2-week lag; however, the eukaryotic and marnavirus-like communities, with lags, were also strong predictors of the T4-like viral assemblages.

DISCUSSION

Temporal dynamics in the taxonomic composition of two groups of ecologically important viruses, as well as the bacterial and microeukaryotic communities containing their putative hosts, were followed in coastal water samples collected every 2 weeks for 13 months and showed that the genetic composition of viral assemblages is highly dynamic. Data for a suite of environmental variables and data from high-throughput sequencing of marker genes were used to place the dynamics of these viral groups in the context of their putative hosts and the environment. Within the viral assemblages, clusters of phylogenetically related viruses showed temporal dynamics of dominance whereby different clusters represented >50% of the relative abundance of the assemblages at different times (Fig. 1). While some members of these clusters were persistent, others were ephemeral. Changes in the composition of the viral assemblages were related to changes in the composition of the cellular communities and were consistent with the concept that there is a “seed bank” of viruses that dominate ephemerally when suitable hosts are present. Moreover, within groups of closely related viruses (clusters), the rank order of the viruses fluctuated, but the relative abundances of the

FIG 4 Marine marnavirus-like viral cluster A OTUs compared to OTUs classified as raphidophytes over time. (A) (Top) Diversity of all 18S rRNA OTUs at each time point. (Bottom) Diversity of OTUs (97% amino acid similarity) classified as raphidophytic over time. (B) (Top) Diversity of all marnavirus-like virus OTUs at each time point. (Bottom) Diversity of marnavirus-like virus cluster A OTUs (95% amino acid similarity) over time. Vertical gray lines indicate boundaries between seasons. X indicates missing or removed samples.

clusters in the overall assemblage did not shift with these fluctuations. Our results show that the temporal dynamics of viral assemblages have a distinct phylogenetic signal, with most genotypes being shuffled within several viral clusters. The implications of these results for our understanding of viral ecology and host-virus interactions are discussed below.

Phylogenetic clusters of viruses were dominated by close relatives whose relative abundances shifted over time. The concept of a seed bank model for viruses (44) is that high local viral diversity is maintained by continuously shuffling the rank order abundance of viral taxa, a process supported by a number of observations (11, 46–48). Yet viral assemblages are often dominated by a few successful taxa, while other taxa are rare and are maintained in the “bank” (34). In our study, only a few viral taxa (OTUs) within each phylogenetic cluster were frequently found at high relative abundances, while most were ephemeral; thus, this bank structure occurs not just at the overall level of the assemblage but also on a finer scale, within the phylogenetic clusters. Within both MLV and T4-like viral assemblages, there was persistently high taxonomic richness (Fig. S2 in the supplemental material); however, our data showed no evidence that this richness was maintained because of shuffling of the dominant

FIG 5 T4-like virus cluster I compared to bacterial OTUs classified as cyanobacteria over time. (A) (Top) Diversity of all 16S rRNA OTUs at each time point. (Bottom) Diversity of bacterial OTUs (97% amino acid similarity) classified as cyanobacteria over time. (B) (Top) Diversity of all T4-like virus OTUs. (Bottom) Diversity of T4-like virus cluster I OTUs (95% amino acid similarity) over time. Vertical gray lines indicate boundaries between seasons. X indicates missing or removed samples.

viral taxa within these assemblages. Rather, a few taxa persisted at high abundances, while a few other, closely related taxa would wax and wane in relative abundance (Fig. S10 and S11). The implication of these observations is that the seed bank model offers an explanation for the shifts in abundance that occur in relatively rare viral taxa but does not explain why specific taxa dominate for extended periods.

Regarding the continued production of viral genotypes, it has been suggested that the most abundant OTUs persist through time in a geographic region, while transient viruses are dispersed from other areas (49). Yet in our study, some viral OTUs were persistent, and thus continually successful, while others were ephemeral and frequently were closely related to the persistent OTUs (Fig. S10 and S11), suggesting that they did not arrive by dispersal. Some viral clusters (T4-like clusters C and D and MVL clusters B, C, D, F, and G) contained no persistent OTUs (Fig. S10 and S11), although these were typically clusters with relatively few OTUs. There are numerous other OTU-based studies reporting assemblages that contain both persistent and ephemeral viruses over time scales of weeks to years, including large double-stranded DNA viruses of

FIG 6 Results of Mantel tests comparing community similarity matrices and distance matrices of environmental data.

phytoplankton in a freshwater lake (50, 51), in a Norwegian inlet (52), and at a coastal site (53), as well as marine T4-like virus assemblages (11, 12). In contrast, over time scales of days, T4-like viral assemblages were relatively stable, and overall changes in both the bacterial and T4-like viral assemblages were slow (54). Collectively, these data demonstrate that across a wide range of systems and on time scales from weeks to years, some viruses persist for long periods, whereas most are ephemeral. This pattern appears consistent for both RNA and DNA viruses infecting microeukaryotic hosts, as well as for bacteriophages, suggesting that different environmental conditions may affect the virulence of particular viruses such that some viruses are more infectious to hosts or to specific strains of the hosts. Alternatively, the patterns could be explained by a patchy distribution of hosts, which could affect the abundances of different hosts at one time point. These dynamics related to the persistence of the viral genotypes could be regulated through ecological processes (e.g., through habitat filtering, where closely related species can persist in a particular environment [55]).

Changes in viral dominance were tied to the host community and environmental conditions. The differences in the distributions of the relative abundances of the OTUs over time suggested different lifestyles within the viral assemblages. The OTU compositions in both the T4-like and MVL viruses changed over time, even though the overall viral diversity in both assemblages stayed relatively constant (Fig. 2 and 3; Fig. S2). The phylogenetic clusters showed three different distributions of their relative abundances during the time series: either one large peak, multiple peaks, or no peaks (constant diversity). These different distributions could represent different viral lifestyles, where the clusters with large or multiple peaks could represent r-selected organisms, that is, organisms with fast growth/high abundance, while clusters with a constant presence

(no peaks) could represent K-selected organisms, which have slow growth/low abundance (56). Thus, viruses that are likely physiologically similar can have different lifestyles when observed at a 2-week scale, suggesting that different lifestyles can be supported by the hosts and the environmental conditions.

Examining viruses with respect to potential hosts and their environments is crucial for interrogating the Royal Family framework (35), whereby closely related viruses change in rank abundance over time, because this shuffling could reflect strain specificity or changes in environmental conditions. The observation that the diversity of viruses related to those infecting the raphidophyte *Heterosigma akashiwo* (viral cluster A) increased sharply following a bloom of raphidophytes, whose diversity also increased (Fig. 4), is consistent with viral infection affecting the diversity of the host population, which, in turn, has effects on virus diversity. We also saw a small increase in raphidophyte abundance and diversity in February but no concomitant increase in the abundance or diversity of viruses in the MVL viral cluster A. There are several possible explanations for the lack of an increase in MVL viral cluster A, including that these raphidophytes were resistant to viruses in cluster A, that the number of susceptible hosts was too low to allow for a detectable increase in viruses, or that other viruses, such as the DNA virus HaV (57), or protist grazing (58) suppressed the raphidophyte populations and prevented the replication of MVL cluster A viruses. Similar patterns were detected for cyanobacteria and T4-like myoviral cluster I, which includes isolates from *Prochlorococcus* phage PSSM2 and *Synechococcus* phage SSM2 (Fig. 5); here, increases in the diversity of T4-like myoviral cluster I followed increased diversity of the cyanobacteria. These types of patterns have been observed in other studies in which viral OTUs had strong time-lagged correlations with bacterial OTUs (59). Also, in a mesocosm experiment with *Emiliania huxleyi*, a peak in host abundance was followed 4 days later by a peak in *Emiliania huxleyi* virus (EhV) abundance (60). In our study, continuous shifts in the fine-scale structure of viral assemblages, but stability at the coarse scale (where the fine scale is represented by OTUs and the coarse scale by phylogenetic clusters), is consistent with viral strain-level diversity preserving host strain-level diversity (19, 49, 61). This structure ties into the Royal Family framework (35) by demonstrating how the members of the “family” play a role in the abundance patterns of the hosts.

Alignment of sequences from marnavirus-like cluster A (Fig. S12), the cluster that includes the viral isolate HaRNAV, revealed that in the “palm” section of catalytic site C of the RdRp protein (45), the persistent OTUs contained the amino acid aspartate (D) while the ephemeral OTUs contained glutamate (E). These ephemeral sequences point to a population that may be infecting a less-abundant host, while OTU_2, the most abundant OTU and the only OTU present at every time point, retained the amino acid aspartate. This suggests that the diversity within a phylogenetic cluster could arise from a large-scale infection event in which a single infection creates a genetic cloud of viruses (62). Evidence for such strain-level diversity has been observed in marine RNA viral metagenomic data sampled at a global scale (28), where viral single nucleotide variants (SNVs) are restricted to specific geographic regions. The SNVs also reveal that viral genomes are under purifying selection, since synonymous variants are more prevalent than nonsynonymous variants, suggesting that the latter are less successful and are removed from the population. Alternatively, predation of hosts by viruses with strain-level heterogeneity could maintain constant diversity in the host populations (6). Furthermore, within-strain dynamics and seasonal repeatability in assemblage structure are seen in viral assemblages from a 5-year viral metagenomic time series in coastal California waters (63). The within-strain dynamics are observed using polymorphic positions in the assembled metagenomes which exemplify a Red Queen dynamic, where viruses maintain strain-level diversity to “outrun” hosts (63). Thus, our observations give credence to the idea that strain-level heterogeneity is prevalent in viral populations in marine environments.

In both the MVL and T4-like assemblages, the overall seasonal stability of the phylogenetic clusters (Fig. 1B and D) and the changes in the composition and relative

abundance of OTUs within clusters between time points (Fig. S10 and S11) likely reflected the observation that viral and microbial communities can be stable at the genus level and up but are dynamic at the strain level for viruses and at the species level for cellular microbes (49). Furthermore, taxonomically related species often have similar niches and ecology (30, 31). Similarly, within cyanophages and viruses infecting eukaryotic phytoplankton, there are phylogenetic groups that are associated with the hosts that they infect (26, 64), in accord with the concept that the hosts represent specific niches. Moreover, similarity in marker gene sequences reflects gene content in cyanophages (26) and in double-stranded DNA viruses infecting photosynthetic protists (27). These marker genes include genes that encode proteins involved in cellular metabolic processes such as photosynthesis (65), thus suggesting that environmental conditions can also affect the phylogenetic structure within viral assemblages. Previously, a study at the Bermuda Atlantic Time-series Study (BATS) site used deep sequencing of the viral gene marker *phoH* to resolve phylogenetically distinct clusters of viral assemblages (34). The assemblage composition differed between fall and winter, and these differences were attributed largely to a phylogenetic cluster containing cyanobacterium-infecting viruses and were thought to reflect differences in water stratification (34). Thus, based on our results and existing evidence, it seems that the phylogenetic structure within a viral assemblage can be influenced directly by changes in the host communities and indirectly by environmental changes. In turn, this implies that the most closely related viruses are adapted to infecting closely related hosts and to similar environmental conditions; thus, small changes in host communities or environmental conditions may have cascading effects leading to successional changes within phylogenetic clusters of viruses. Moreover, the linkages implied between viral phylogeny, host community structure, and environmental conditions provide a mechanism for the persistence of viral taxa within a phylogenetic cluster when the host community and environmental conditions are stable.

Caveats for interpretation of the results. Although we adopted a commonly used approach for extracting nucleic acids from filters, which included incubating with lysozyme, followed by SDS and several freeze-thaw cycles (53), there are a number of alternative approaches in the literature, including using commercially available kits (66) or adding a step to disrupt the cells by bead beating (67). Indeed, for unicellular eukaryotes, different extraction methods engender different biases for the community represented and the relative abundances of different taxa (68). Given that the taxonomic composition of the samples varied seasonally, and alternate methods may bias the results in different ways, it is appropriate to use only one method; hence, to examine changes in the eukaryotic microbial community, we adopted the same approach previously used at this location (53). Given that the taxonomic profile of the cellular communities can be affected by the methodology used, the taxonomic data must be viewed as representative, but not absolute.

In addition to potential biases introduced by different methods of extracting nucleic acids, there is a rich debate on the best way to access the microbial composition of a sample from extracted nucleic acids. For cellular microbes, it is generally accepted that deep sequencing of PCR-amplified fragments from the rRNA genes is a powerful approach for assessing taxonomic composition, although there is much discussion about the specific fragments to target. For viruses, PCR amplification of targeted marker genes is also a well-established and powerful approach, but in contrast to the rRNA genes for cellular microbes, the marker genes represent only a small subset of the total viral diversity. Again, there is a rich literature on the marker genes to target and the specific primers to use; regardless, a single primer set will only yield information about a small slice of the total viral diversity in a natural sample. However, despite the challenges of using PCR amplicons for assessing diversity with viral gene markers (37, 69), and the use of PCR in general (70), high-throughput sequencing of amplicons is an excellent approach for examining viral population structures. Although rare and putative OTUs can result from PCR errors (71) or sampling anomalies (72), this

was not likely a factor in this study, given that almost all OTUs were seen multiple times in different samples, indicating that they were not spurious. In addition, a number of bioinformatic protocols were put in place to reduce the possibility of spurious OTUs (see Materials and Methods). Hence, although the relative abundance of sequences for any OTU is semiquantitative (71), PCR-based approaches remain powerful for comparing taxonomic richness and diversity among samples. The use of amplicon sequence variants (ASVs) was considered for this analysis; however, given the grouping of the data into phylogenetic clusters and taxonomic groups for comparison, the use of OTUs was deemed the most appropriate approach for our analyses.

Conclusions. Phylogenetically related clusters of OTUs showed temporal patterns. Different phylogenetic clusters had the highest relative abundance in the assemblages at different times during this study. Furthermore, within these clusters, some viruses persisted while others were ephemeral, providing evidence of the operation of a seed bank in these assemblages. Viral assemblages displayed evidence of time-lagged dynamics relative to communities of potential hosts. For example, increases in the relative abundances of sequences in MVL viral cluster A, which contains a viral isolate (HaRNAV) that infects a raphidophyte, were correlated in a lagged fashion with increases in the relative abundances of sequences assigned to raphidophytes. Similarly, increases in the relative abundances of sequences assigned to the T4-like virus cluster I, which contains the cyanophage isolates PSSM2 and SSM2, were correlated with increases in sequences assigned to cyanobacteria. MVL and T4-like viral assemblages differed in diversity, evenness, and the proportion of OTUs that persisted over time.

Previously, the link between the assemblage dominance, persistence, and phylogeny of virus-host communities has largely been overlooked. Other studies have found that many viral OTUs are ephemeral and that a few are persistent. Here, we expand this understanding by showing that in two different viral assemblages, most ephemeral viral OTUs were closely related to a persistent viral OTU and that over time, the assemblages were dominated by different phylogenetic clusters of viruses that comprise both ephemeral and persistent OTUs.

These results provide strong support for the concept of the Royal Family in viral ecology and add the layer of relatedness of viruses. This relatedness is crucial when one is examining OTUs that are shuffled in rank abundance over time, since the replacement of an OTU that is the most abundant at one time point (rank 1) by an OTU from within the same cluster at the next time point offers very different ecological information from replacement by an OTU from a different cluster. The dominance of different viral clusters over time was seen in this study and was correlated with changes in environmental conditions and/or putative hosts. Moreover, these dynamics occurred in both bacteriophages and RNA viruses infecting microbial eukaryotes, suggesting that this structure is a conserved property of natural viral assemblages.

MATERIALS AND METHODS

Sample collection. Seawater samples were collected from Jericho Pier (49°16'36.73"N, 123°12'05.41"W) in British Columbia, Canada. Jericho Pier is adjacent to the shoreline in a well-mixed location, with mixed semidiurnal tides and significant freshwater influence from the Fraser River. To obtain a representative sample of water and enough material for viral extraction, 60 liters of water was collected from the 1-m depth at the daytime high tide every 2 weeks between June 2010 and July 2011 (33 samples). Salinity and temperature were measured using a YSI (Yellow Springs, OH, USA) probe. For all samples, the water was prefiltered through 65- μ m-pore-size Nitex mesh and filtered sequentially through a GC50 glass-fiber filter (diameter, 142 mm; nominal pore size, 1.2 μ m; Advantec MFS, Dublin, CA, USA) and a 0.22- μ m-pore-size polyvinylidene filter (Millipore, Bedford, MA, USA). Viral-size particles in the filtrate were concentrated to ~500 ml (viral concentrate) using tangential flow ultrafiltration with a 30-kDa molecular weight (MW) Prep/Scale Spiral Wound TFF-6 cartridge (Millipore) (73).

Nutrients. Phosphate, silicate, and nitrate-plus-nitrite concentrations were determined in duplicate for each time point by collecting 15-ml seawater samples, which were filtered through 0.45- μ m-pore-size HA filters (Millipore) and stored at -20°C until processing on an air-segmented continuous-flow AutoAnalyzer 3 instrument (Bran+Luebbe, Norderstedt, Germany). Chlorophyll *a* (Chl *a*) concentrations were determined in triplicate by filtering 100 ml of seawater through 0.45- μ m-pore-size HA filters (Millipore), storing the filters in the dark at -20°C until acetone extraction, and then quantifying the chlorophyll fluorometrically (74).

TABLE 1 PCR parameters used in this study

Marker gene	Target	Primer names	PCR conditions						Reference(s)
			Initialization	Denaturation	Annealing temp (°C)	Extension	No. of cycles	Final extension	
<i>gp23</i>	T4-like viruses	MZIA1bis, MZIA6	94°C for 90 s	94°C for 45 s	50	72°C for 45 s	35	5 min at 72°C	42
<i>RdRp</i>	Marnavirus-like viruses	MPL-2F, MPL-2R	94°C for 75 s	94°C for 45 s	43	72°C for 60 s	40	9 min at 72°C	43
18S rRNA gene	Eukaryotes	Euk1209f, Uni1392r	94°C for 75 s	94°C for 1 min	65 touchdown for 10 cycles, followed by 55	72°C for 60 s	10 + 20	9 min at 72°C	84
16S rRNA gene	Bacteria	341F, 907R	94°C for 75 s	94°C for 1 min	64 for 12 cycles, followed by 54	72°C for 60 s	12 + 25	10 min at 72°C	85, 86

Enumeration of bacteria and viruses. Duplicate 980- μ l samples for viral and bacterial abundances were collected in cryovials at each time point, fixed with a final concentration of 0.5% electron microscopy (EM)-grade glutaraldehyde, frozen in liquid nitrogen, and stored at -80°C until processing by flow cytometry (75).

Briefly, viral samples were diluted 1:10 to 1:10,000 in sterile filtered (pore size, 0.1 μm) $1\times$ Tris-EDTA (TE), stained with SYBR green I (Invitrogen, Waltham, MA, USA) at a final concentration of 0.5×10^{-4} of the commercial stock, heated for 10 min at 80°C , and then cooled in the dark for 5 min before processing. Bacterial samples were diluted up to 1:1,000 in sterile filtered (pore size, 0.1 μm) $1\times$ TE, stained with SYBR green I (Invitrogen) at a final concentration of 0.5×10^{-4} of the commercial stock, and incubated in the dark for 15 min before processing. All samples were run for 1 min on a FACSCalibur flow cytometer (Becton, Dickinson, Franklin Lakes, NJ, USA) at medium and high flow rates for the viral and bacterial samples, respectively. Event rates were kept between 100 and 1,000 events s^{-1} , and green fluorescence and side scatter detectors were used. Data were processed and gated using CellQuest software (Becton, Dickinson).

Extraction of viral nucleic acids. The viral concentrate was filtered twice through 0.22- μm -pore-size Durapore polyvinylidene difluoride (PVDF) filters (Millipore) in a sterile Sterivex filter unit (Millipore). Viral-size particles in the filtrate were pelleted by ultracentrifugation (Beckman-Coulter, Brea, CA, USA) in an SW40 rotor at $108,000\times g$ 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 1 U μl^{-1} DNase with a final concentration of 5 mM MgCl_2 for 3 h at room temperature. Nucleic acids were extracted using a QIAamp MinElute virus spin kit (Qiagen, Hilden, Germany) according to the manufacturer's directions.

PCR amplification of the marker gene for T4-like viruses. Throughout this report, we use the term "T4-like viruses" to refer to a broad group of phages that encode several evolutionarily related core structural proteins in common with the genus *Tequatrovirus*, including the *gp23* gene, encoding the major capsid protein (24, 76). These viruses infect bacteria, including cyanobacteria.

To target the marine T4-like virus capsid protein gene (*gp23*), PCR was performed on the concentrated samples as in the work of Filée et al. (24). Briefly, each reaction mixture (final volume, 50 μl) consisted of 2 μl extracted viral DNA (see the preceding section), $1\times$ (final concentration) PCR buffer (Invitrogen, Carlsbad, CA, USA), 1.5 mM MgCl_2 , 0.2 mM each deoxynucleoside triphosphate (Bioline, London, UK), 40 pmol of MZIA1bis and 40 pmol of MZIA6, and 1 U Platinum *Taq* DNA polymerase (Invitrogen). Amplification was performed using the program conditions presented in Table 1.

The sequences of PCR primers used in this study are given in Table 2.

PCR amplification of the marker gene for marnaviruses. The second group of viruses targeted was marnavirus-like (MVL) viruses, belonging to the expansive family of marine RNA viruses *Marnaviridae* in the order *Picornavirales* (29). These viruses infect eukaryotic phytoplankton and likely heterotrophic protists. They were targeted by amplifying part of the RNA-dependent RNA polymerase

TABLE 2 Sequences of PCR primers used in this study

Primer name	Primer sequence (5'-3')	Reference
MZIA1bis	GATATTGIGGIGTTCAGCCATGA	24
MZIA6	CGCGGTTGATTCCAGCATGATTC	24
Mpl.sc2F	ITWGCIGGIGATTWCA	37
Mpl.sc2R	CKYTTCARRAAWTCAGCATC	37
Euk1209f	CAGGTCTGTGATGCC	77
Uni1392r	ACGGGCGGTGTGTRC	77
341F	CCTACGGGIGGICGA	78
907R	CCGTCAATTCCTTTAGTTT	79

(*RdRp*) gene (37). Briefly, half of each viral extract was used to synthesize cDNA. To remove DNA, the extracted viral pellets were digested with amplification-grade DNase I (Invitrogen). The reaction was terminated by adding 2.5 mM EDTA (final concentration) and incubating for 10 min at 65°C. 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 (37). Each reaction mixture (final volume, 50 μl) consisted of 50 ng of cDNA, 1 $\times$ (final concentration) PCR buffer, 2 mM MgCl_2 , 0.2 mM each deoxynucleoside triphosphate, 1 μM each primer, and 1 U Platinum *Taq* DNA polymerase. The reaction was run in a PCR Express thermocycler (Hybaid, Ashford, UK) under the program conditions presented in Table 1. Products were run on a 0.5 $\times$ Tris-borate-EDTA (TBE) 1% low-melt gel, excised, and extracted using a Zymoclean Gel DNA recovery kit (Zymo, Irvine, CA, USA) as per the manufacturer with a final elution step of 2 $\times$ 10 μl EB buffer (Qiagen).

Filtration and extraction of marine bacteria and eukaryotes. One liter of 65- μm -pore-size Nitex mesh-filtered water from each 60-liter seawater sample was filtered through a 0.22- μm -pore-size Durapore PVDF 47-mm filter (Millipore) in a sterile Sterivex filter unit (Millipore). The filter was either stored at -20°C until extraction or immediately extracted as follows (53). Briefly, filters were aseptically cut into small pieces and were incubated with lysozyme (Sigma-Aldrich, St. Louis, MO, USA) at a final concentration of 1 mg ml^{-1} for 2 h at 37°C. Sodium dodecyl sulfate was added at a final concentration of 0.1% (wt/vol), and each filter was put through three freeze-thaw cycles. Proteinase K (Qiagen) was then added to a final concentration of 100 $\mu\text{g ml}^{-1}$ and incubated for 1 h at 55°C. DNA was sequentially extracted using equal volumes of phenol-chloroform-isoamyl alcohol (25:24:1) and chloroform-isoamyl alcohol (24:1). DNA was precipitated by adding NaCl to a final concentration of 0.3 M and adding twice the extract volume of ethanol. Samples were incubated at -20°C for 1 h and were then centrifuged for 1 h at 20,000 $\times g$ and 4°C. Extracts were washed with 70% ethanol and were resuspended in 50 μl EB buffer.

PCR amplification of bacterial and eukaryotic ribosomal sequences. PCRs targeting eukaryotes were performed using primers Euk1209f and Uni1392r (77). These primers target positions 1423 to 1641 of the 18S rRNA and include the variable region V8. Each reaction mixture (final volume, 50 μl) consisted of 2 μl DNA template, 1 $\times$ (final concentration) PCR buffer, 1.5 mM MgCl_2 , 0.2 mM each deoxynucleoside triphosphate, 0.3 μM each primer, and 2.5 U Platinum *Taq* DNA polymerase. The reactions were run in a PCR Express thermocycler under the program conditions set forth in Table 1.

PCRs targeting bacteria used primers 341F (78) and 907R (79). These primers target the V3-to-V5 regions of the 16S rRNA. PCRs were run under the following conditions: each reaction mixture (final volume, 50 μl) consisted of 2 μl DNA template, 1 $\times$ (final concentration) PCR buffer, 1.5 mM MgCl_2 , 0.2 mM each deoxynucleoside triphosphate, 0.4 μM each primer, and 1 U Platinum *Taq* DNA polymerase. The reactions were run in a PCR Express thermocycler under the program conditions set forth in Table 1.

Sequencing library preparation. (i) Construction. PCR products not requiring gel excision (T4-like viruses, rRNA amplicons) were purified after PCR using AMPure XP beads (Beckman Coulter) at a bead/product ratio of 1.2:1. Cleaned products were resuspended in 30 μl EB buffer (Qiagen) and quantified using the PicoGreen dsDNA assay (Invitrogen) with lambda DNA (Invitrogen) as a standard. Sample concentrations were read using the iQ5 and CFX96 Touch systems (both from Bio-Rad, Hercules, CA, USA). Pooled libraries were constructed using one each of the four different amplicons with an equal molarity for each amplicon; the total input material of each pool was between 700 and 900 ng. Pooled amplicons were concentrated using AMPure XP beads (Beckman Coulter) at a bead/product ratio of 1.2:1. NxSeq DNA sample prep kit 2 (Lucigen, Middleton, WI, USA) was used as per manufacturer's directions with either NEXTflex 48 bar codes (BioO, Austin, USA), NEXTflex 96 HT bar codes (BioO), or TruSeq adapters (IDT, Coralville, Iowa). Libraries were purified using AMPure XP beads (Beckman Coulter) at a bead/library ratio of 0.9:1.

(ii) Quantification and quality control of libraries. Libraries were analyzed for small fragments (primer dimers and adapter dimers) using a 2100 Bioanalyzer system (Agilent, Santa Clara, CA, USA) with the High Sensitivity DNA kit (Agilent). The concentrations of libraries were determined using the PicoGreen dsDNA assay as described above. The libraries were quantified and checked for amplifiable adapters using Library Quantification DNA standards 1 to 6 (Kapa Biosystems, Wilmington, MA, USA) with the SsoFast EvaGreen quantitative PCR (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 1:1,000, 1:5,000, and 1:10,000 dilutions of the libraries run in triplicate on iQ5 and CFX96 Touch qPCR machines. Cycling parameters were as follows: 95°C for 30 s; 35 cycles of 95°C for 5 s and 60°C for 30 s; and melt curve generation from 65°C to 95°C in 0.5°C steps (10 s/step). Quantification from both the PicoGreen and qPCR assays was used to determine final pooling of all libraries before sequencing. Libraries were sequenced using 2 $\times$ 250 bp MiSeq sequencing (Illumina, San Diego, CA, USA) at the Génome Québec Innovation Centre at McGill University (Montreal, QC, Canada) and 2 $\times$ 300 bp PE MiSeq sequencing (Illumina) at the UBC Pharmaceutical Sciences Sequencing Centre (Vancouver, BC, Canada) and at UCLA's GenoSeq Core (Los Angeles, CA, USA).

Sequence processing. Sequencing libraries were processed as follows. The libraries were either split by the sequencing center using CASAVA (Illumina) or with the MiSeq Reporter (Illumina), and sequence quality was initially examined using FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Contaminating sequencing adapters were removed using Trimmomatic, version 0.32 (80), and the quality of the libraries was further examined using fastx_quality (81). Libraries were further split into individual amplicons (i.e., 18S rRNA, 16S rRNA, *gp23*, and *RdRp*), and then, if the expected overlap of the

paired-end reads was ≥ 40 bp, the paired reads were merged using PEAR (82). Since there were many fewer reads for the 16S rRNA reverse primer, only the forward reads were analyzed. All sequences were then quality trimmed using Trimmomatic with the default quality settings.

16S and 18S rRNA sequences were aligned to the Silva 119 database (83) using align.seqs in mothur, version 1.33.3 (84). Sequences that aligned were retained and were checked for chimeras using USEARCH, version 8.0.1517 (85), with the Gold reference database. Unique, nonchimeric sequences were clustered at 97% similarity and were assigned to taxa using mothur (Wang-type algorithm) and the Silva 119 taxonomy (83).

For viruses, sequences were queried using BLAST against databases containing the gene markers of interest from isolates and environmental surveys (see Supplemental Methods for details on databases). Sequences with matches with an E value below 10^{-3} were kept, chimera-checked using USEARCH with the *de novo* and reference methods (85), and then translated using FragGeneScan, version 1.20 (86). Since sequence similarity cutoffs for OTU clustering are not well-defined for viruses, the relationship between the numbers of OTUs over a range of similarities (at the protein level) from 50% to 100% was examined for both MVL and T4-like viruses. The rise in OTUs was linear for similarities from 50% to 95%, but above 95%, the slope was nonlinear; thus, USEARCH (85) was used to cluster sequences with 95% similarity into an OTU and was used to build OTU tables. Similar approaches for determining sequence similarity cutoffs have been used in other studies (27, 69, 87–89). vegan (90) was used to generate rarefaction curves and normalize the sequences; only samples in which $>2,000$ quality-trimmed reads were retained were analyzed. Viral OTUs or clusters are said to dominate when they individually or collectively constitute $>50\%$ of the relative abundance at a given time point. A viral OTU is defined as persistent if it is present in $>90\%$ of samples and as ephemeral if it occurs in $<20\%$ of samples.

Data analysis, multivariate statistics, and phylogenetic analysis. Environmental parameters with missing data were mean imputed to fill in missing values. Day length data were retrieved using the R package geosphere (91). adonis was used as implemented in vegan (90) to test whether assemblage and community matrices showed seasonal differences. Bray-Curtis distance matrices were constructed from the normalized OTU abundance tables. Mantel tests as implemented in vegan (90) were performed by comparing the assemblage and community distance matrices to each other and to distance matrices of environmental parameters.

NCBI Conserved Domains Database (CDD) domain alignments for *RdRp* and *gp23* were retrieved and used as hidden Markov models with HMMER (92) to align the translated OTUs using Clustal Omega (93). Sequences for *gp23* (11, 24, 94–99) and *RdRp* (25, 37) from marine isolates of *Marnaviridae* and T4-like viruses, as well as environmental surveys, were aligned with OTUs using Clustal Omega (93). The environmental sequences retrieved from GenBank were clustered at the same percentages as the OTUs, and actual sequences (not consensus sequences) were used to represent the unique sequences on the phylogenetic trees. The full list of the GenBank accession numbers used is included in Supplemental Methods.

Alignments were viewed and manually curated with AliView (100). Automated trimming of the alignments was done using trimAl (101). Initial phylogenetic trees were built with FastTree (102), and the final maximum-likelihood trees were generated using RAxML (103) with 1,000 bootstraps. For the *RdRp* tree, the VT substitution model was used with the gamma model of heterogeneity (PROTGAMMA), and for the *gp23* tree, JTT was used with the gamma model of heterogeneity (PROTGAMMA). Models were chosen using ProtTest (104). Viral clusters were chosen using the phylogenetic tree by choosing distinct groups with good support as described by Goldsmith et al. (34). Faith's phylogenetic diversity (105) was calculated as implemented in Picante (106). The R package ggtree (107) was used for visualizing and annotating trees, and all other plots were created using R (108).

Data availability. The raw sequence files have been deposited in NCBI BioProject under accession number PRJNA406940. All scripts used for processing the data are available at <https://doi.org/10.5281/zenodo.2554772>.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 2.6 MB.

ACKNOWLEDGMENTS

Members of the Suttle lab and the Beaty Biodiversity Center 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.M. 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 J.A.G., Discovery Research and Ship-time Grants to C.A.S., and fellowships to J.A.G. 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.

We declare no conflict of interest.

REFERENCES

- Lynch MDJ, Neufeld JD. 2015. Ecology and exploration of the rare biosphere. *Nat Rev Microbiol* 13:217–229. <https://doi.org/10.1038/nrmicro3400>.
- DuRand MD, Olson RJ, Chisholm SW. 2001. Phytoplankton population dynamics at the Bermuda Atlantic Time-series station in the Sargasso Sea. *Deep Sea Res Part II Top Stud Oceanogr* 48:1983–2003. [https://doi.org/10.1016/S0967-0645\(00\)00166-1](https://doi.org/10.1016/S0967-0645(00)00166-1).
- 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. *Limnol Oceanogr* 50:1687–1696. <https://doi.org/10.4319/lo.2005.50.5.1687>.
- Fuhrman JA, Hewson I, Schwalbach M, Steele J, Brown M, Naeem S. 2006. Annually reoccurring bacterial communities are predictable from ocean conditions. *Proc Natl Acad Sci U S A* 103:13104–13109. <https://doi.org/10.1073/pnas.0602399103>.
- 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. *Environ Microbiol* 11:3132–3139. <https://doi.org/10.1111/j.1462-2920.2009.02017.x>.
- Rodríguez-Valera F, Martín-Cuadrado A-B, Rodríguez-Brito B, Pašić L, Thingstad TF, Rohwer F, Mira A. 2009. Explaining microbial population genomics through phage predation. *Nat Rev Microbiol* 7:828–836. <https://doi.org/10.1038/nrmicro2235>.
- Bergh O, Borsheim KY, Bratbak G, Heldal M. 1989. High abundance of viruses found in aquatic environments. *Nature* 340:467–468. <https://doi.org/10.1038/340467a0>.
- Jiang SC, Paul JH. 1994. Seasonal and diel abundance of viruses and occurrence of lysogeny/bacteriocinogeny in the marine environment. *Mar Ecol Prog Ser* 104:163–172. <https://doi.org/10.3354/meps104163>.
- Winget DM, Helton RR, Williamson KE, Bench SR, Williamson SJ, Wommack KE. 2011. Repeating patterns of virioplankton production within an estuarine ecosystem. *Proc Natl Acad Sci U S A* 108:11506–11511. <https://doi.org/10.1073/pnas.1101907108>.
- Parsons RJ, Breitbart M, Lomas MW, Carlson CA. 2012. Ocean time-series reveals recurring seasonal patterns of virioplankton dynamics in the northwestern Sargasso Sea. *ISME J* 6:273–284. <https://doi.org/10.1038/ismej.2011.101>.
- Chow C-ET, Fuhrman JA. 2012. Seasonality and monthly dynamics of marine myovirus communities. *Environ Microbiol* 14:2171–2183. <https://doi.org/10.1111/j.1462-2920.2012.02744.x>.
- Pagarete A, Chow C-ET, Johannessen T, Fuhrman JA, Thingstad TF, Sandaa RA. 2013. Strong seasonality and interannual recurrence in marine myovirus communities. *Appl Environ Microbiol* 79:6253–6259. <https://doi.org/10.1128/AEM.01075-13>.
- Ahlgren NA, Perelman JN, Yeh Y-C, Fuhrman JA. 2019. Multi-year dynamics of fine-scale marine cyanobacterial populations are more strongly explained by phage interactions than abiotic, bottom-up factors. *Environ Microbiol* 21:2948–2963. <https://doi.org/10.1111/1462-2920.14687>.
- Waterbury JB, Valois FW. 1993. Resistance to co-occurring phages enables marine *Synechococcus* communities to coexist with cyanophages abundant in seawater. *Appl Environ Microbiol* 59:3393–3399. <https://doi.org/10.1128/AEM.59.10.3393-3399.1993>.
- Suttle CA, Chan AM. 1994. Dynamics and distribution of cyanophages and their effect on marine *Synechococcus* spp. *Appl Environ Microbiol* 60:3167–3174. <https://doi.org/10.1128/AEM.60.9.3167-3174.1994>.
- Marston MF, Taylor S, Sme N, Parsons RJ, Noyes TJE, Martiny JHB. 2013. Marine cyanophages exhibit local and regional biogeography. *Environ Microbiol* 15:1452–1463. <https://doi.org/10.1111/1462-2920.12062>.
- Sandaa R-A, Storesund J, Olesin E, Lund PM, Larsen A, Bratbak G, Ray J. 2018. Seasonality drives microbial community structure, shaping both eukaryotic and prokaryotic host-viral relationships in an Arctic marine ecosystem. *Viruses* 10:715. <https://doi.org/10.3390/v10120715>.
- Clasen J, Hanson C, Ibrahim Y, Weihe C, Marston M, Martiny J. 2013. Diversity and temporal dynamics of Southern California coastal marine cyanophage isolates. *Aquat Microb Ecol* 69:17–31. <https://doi.org/10.3354/ame01613>.
- Needham DM, Sachdeva R, Fuhrman JA. 2017. Ecological dynamics and co-occurrence among marine phytoplankton, bacteria and myoviruses shows microdiversity matters. *ISME J* 11:1614–1629. <https://doi.org/10.1038/ismej.2017.29>.
- Arkhipova K, Skvortsov T, Quinn JP, McGrath JW, Allen CC, Dutilh BE, McElarney Y, Kulakov LA. 2018. Temporal dynamics of uncultured viruses: a new dimension in viral diversity. *ISME J* 12:199–211. <https://doi.org/10.1038/ismej.2017.157>.
- Woese CR, Fox GE. 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A* 74:5088–5090. <https://doi.org/10.1073/pnas.74.11.5088>.
- Woese CR, Kandler O, Wheelis ML. 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci U S A* 87:4576–4579. <https://doi.org/10.1073/pnas.87.12.4576>.
- Chen F, Suttle CA. 1995. Amplification of DNA polymerase gene fragments from viruses infecting microalgae. *Appl Environ Microbiol* 61:1274–1278. <https://doi.org/10.1128/AEM.61.4.1274-1278.1995>.
- Filée J, Tétart F, Suttle CA, Krisch HM. 2005. Marine T4-type bacteriophages, a ubiquitous component of the dark matter of the biosphere. *Proc Natl Acad Sci U S A* 102:12471–12476. <https://doi.org/10.1073/pnas.0503404102>.
- Culley AI, Lang A, Suttle CA. 2003. High diversity of unknown picorna-like viruses in the sea. *Nature* 424:1054–1057. <https://doi.org/10.1038/nature01886>.
- Finke J, Winget D, Chan A, Suttle C. 2017. Variation in the genetic repertoire of viruses infecting *Micromonas pusilla* reflects horizontal gene transfer and links to their environmental distribution. *Viruses* 9:116. <https://doi.org/10.3390/v9050116>.
- Finke JF, Suttle CA. 2019. The environment and cyanophage diversity: insights from environmental sequencing of DNA polymerase. *Front Microbiol* 10:167. <https://doi.org/10.3389/fmicb.2019.00167>.
- Vlok M, Lang AS, Suttle CA. 2019. Marine RNA virus quasispecies are distributed throughout the oceans. *mSphere* 4:e00157-19. <https://doi.org/10.1128/mSphereDirect.00157-19>.
- Vlok M, Lang AS, Suttle CA. 2019. Application of a sequence-based taxonomic classification method to uncultured and unclassified marine single-stranded RNA viruses in the order *Picornavirales*. *Virus Evol* 5:vez056. <https://doi.org/10.1093/ve/vez056>.
- Harvey P, Purvis A. 1991. Comparative methods for explaining adaptations. *Nature* 351:619–624. <https://doi.org/10.1038/351619a0>.
- Srivastava DS, Cadotte MW, MacDonald AAM, Marushia RG, Mirotchnick N. 2012. Phylogenetic diversity and the functioning of ecosystems. *Ecol Lett* 15:637–648. <https://doi.org/10.1111/j.1461-0248.2012.01795.x>.
- Horner-Devine MC, Bohannan BJM. 2006. Phylogenetic clustering and overdispersion in bacterial communities. *Ecology* 87(7 Suppl): S100–S108. [https://doi.org/10.1890/0012-9658\(2006\)87\[100:PCAOIB\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2006)87[100:PCAOIB]2.0.CO;2).
- Lennon JT, Aanderud ZT, Lehmkuhl BK, Schoolmaster DR. 2012. Mapping the niche space of soil microorganisms using taxonomy and traits. *Ecology* 93:1867–1879. <https://doi.org/10.1890/11-1745.1>.
- Goldsmith DB, Parsons RJ, Beyene D, Salamon P, Breitbart M. 2015. Deep sequencing of the viral *phoH* gene reveals temporal variation, depth-specific composition, and persistent dominance of the same viral *phoH* genes in the Sargasso Sea. *PeerJ* 3:e997. <https://doi.org/10.7717/peerj.997>.
- Breitbart M, Bonnain C, Malki K, Sawaya NA. 2018. Phage puppet masters of the marine microbial realm. *Nat Microbiol* 3:754–766. <https://doi.org/10.1038/s41564-018-0166-y>.
- Culley AI, Lang A, Suttle CA. 2006. Metagenomic analysis of coastal RNA virus communities. *Science* 312:1795–1798. <https://doi.org/10.1126/science.1127404>.

37. Culley AI, Steward GF. 2007. New genera of RNA viruses in subtropical seawater, inferred from polymerase gene sequences. *Appl Environ Microbiol* 73:5937–5944. <https://doi.org/10.1128/AEM.01065-07>.
38. Tomaru Y, Toyoda K, Kimura K. 2015. Marine diatom viruses and their hosts: resistance mechanisms and population dynamics. *PLoS One* 10:e0121014. <https://doi.org/10.1371/journal.pone.0121014>.
39. Culley AI, Mueller JA, Belcaid M, Wood-Charlson EM, Poisson G, Steward GF. 2014. The characterization of RNA viruses in tropical seawater using targeted PCR and metagenomics. *mBio* 5:e01210-14. <https://doi.org/10.1128/mBio.01210-14>.
40. Miranda JA, Culley AI, Schvarcz CR, Steward GF. 2016. RNA viruses as major contributors to Antarctic viroplankton: RNA viruses in the Antarctic. *Environ Microbiol* 18:3714–3727. <https://doi.org/10.1111/1462-2920.13291>.
41. Allen LZ, McCrow JP, Ininbergs K, Dupont CL, Badger JH, Hoffman JM, Ekman M, Allen AE, Bergman B, Venter JC. 2017. The Baltic Sea virome: diversity and transcriptional activity of DNA and RNA viruses. *mSystems* 2:e00125-16. <https://doi.org/10.1128/mSystems.00125-16>.
42. Hewson I, Bistolas KS, Button JB, Jackson EW. 2018. Occurrence and seasonal dynamics of RNA viral genotypes in three contrasting temperate lakes. *PLoS One* 13:e0194419. <https://doi.org/10.1371/journal.pone.0194419>.
43. Urayama S-I, Takaki Y, Nishi S, Yoshida-Takashima Y, Deguchi S, Takai K, Nunoura T. 2018. Unveiling the RNA virosphere associated with marine microorganisms. *Mol Ecol Resour* 18:1444–1455. <https://doi.org/10.1111/1755-0998.12936>.
44. Breitbart M, Rohwer F. 2005. Here a virus, there a virus, everywhere the same virus? *Trends Microbiol* 13:278–284. <https://doi.org/10.1016/j.tim.2005.04.003>.
45. Te Velthuis AJW. 2014. Common and unique features of viral RNA-dependent polymerases. *Cell Mol Life Sci* 71:4403–4420. <https://doi.org/10.1007/s00018-014-1695-z>.
46. Short CM, Rusanova O, Short SM. 2011. Quantification of virus genes provides evidence for seed-bank populations of phycodnaviruses in Lake Ontario, Canada. *ISME J* 5:810–821. <https://doi.org/10.1038/ismej.2010.183>.
47. Zhong X, Jacquet S. 2014. Differing assemblage composition and dynamics in T4-like myophages of two neighbouring sub-alpine lakes. *Freshw Biol* 59:1577–1595. <https://doi.org/10.1111/fwb.12365>.
48. Brum JR, Ignacio-Espinoza JC, Roux S, Doulcier G, Acinas SG, Alberti A, Chaffron S, Cruaud C, de Vargas C, Gasol JM, Gorsky G, Gregory AC, Guidi L, Hingamp P, Iudicone D, Not F, Ogata H, Pesant S, Poulos BT, Schwenck SM, Speich S, Dimier C, Kandels-Lewis S, Picheral M, Searson S, Tara Oceans Coordinators; Bork P, Bowler C, Sunagawa S, Wincker P, Karsenti E, Sullivan MB. 2015. Patterns and ecological drivers of ocean viral communities. *Science* 348:1261498. <https://doi.org/10.1126/science.1261498>.
49. Rodriguez-Brito B, Li L, Wegley L, Furlan M, Angly F, Breitbart M, Buchanan J, Desnues C, Dinsdale E, Edwards R, Felts B, Haynes M, Liu H, Lipson D, Mahaffy J, Martin-Cuadrado AB, Mira A, Nulton J, Pašić L, Rayhawk S, Rodriguez-Mueller J, Rodriguez-Valera F, Salamon P, Srinagesh S, Thingstad TF, Tran T, Thurber RV, Willner D, Youle M, Rohwer F. 2010. Viral and microbial community dynamics in four aquatic environments. *ISME J* 4:739–751. <https://doi.org/10.1038/ismej.2010.1>.
50. Short SM, Short CM. 2009. Quantitative PCR reveals transient and persistent algal viruses in Lake Ontario, Canada. *Environ Microbiol* 11:2639–2648. <https://doi.org/10.1111/j.1462-2920.2009.01988.x>.
51. Rozon RM, Short SM. 2013. Complex seasonality observed amongst diverse phytoplankton viruses in the Bay of Quinte, an embayment of Lake Ontario. *Freshw Biol* 58:2648–2663. <https://doi.org/10.1111/fwb.12241>.
52. Gran-Stadniczenko S, Krabberød KA, Sandaa R-A, Yau S, Egge E, Edvardsen B. 2019. Seasonal dynamics of algae-infecting viruses and their inferred interactions with protists. *Viruses* 11:1043. <https://doi.org/10.3390/v11111043>.
53. Short SM, Suttle CA. 2003. Temporal dynamics of natural communities of marine algal viruses and eukaryotes. *Aquat Microb Ecol* 32:107–119. <https://doi.org/10.3354/ame032107>.
54. Needham DM, Chow C-ET, Cram JA, Sachdeva R, Parada A, Fuhrman JA. 2013. Short-term observations of marine bacterial and viral communities: patterns, connections and resilience. *ISME J* 7:1274–1285. <https://doi.org/10.1038/ismej.2013.19>.
55. Koepfel AF, Wu M. 2014. Species matter: the role of competition in the assembly of congeneric bacteria. *ISME J* 8:531–540. <https://doi.org/10.1038/ismej.2013.180>.
56. Suttle CA. 2007. Marine viruses—major players in the global ecosystem. *Nat Rev Microbiol* 5:801–812. <https://doi.org/10.1038/nrmicro1750>.
57. Nagasaki K, Yamaguchi M. 1997. Isolation of a virus infectious to the harmful bloom causing microalga *Heterosigma akashiwo* (Raphidophyceae). *Aquat Microb Ecol* 13:135–140. <https://doi.org/10.3354/ame013135>.
58. Harvey EL, Menden-Deuer S. 2012. Predator-induced fleeing behaviors in phytoplankton: a new mechanism for harmful algal bloom formation? *PLoS One* 7:e46438. <https://doi.org/10.1371/journal.pone.0046438>.
59. Chow C-ET, Kim DY, Sachdeva R, Caron DA, Fuhrman JA. 2014. Top-down controls on bacterial community structure: microbial network analysis of bacteria, T4-like viruses and protists. *ISME J* 8:816–829. <https://doi.org/10.1038/ismej.2013.199>.
60. Schroeder DC, Oke J, Hall M, Malin G, Wilson WH. 2003. Virus succession observed during an *Emiliania huxleyi* bloom. *Appl Environ Microbiol* 69:2484–2490. <https://doi.org/10.1128/aem.69.5.2484-2490.2003>.
61. Emerson JB, Thomas BC, Andrade K, Heidelberg KB, Banfield JF. 2013. New approaches indicate constant viral diversity despite shifts in assemblage structure in an Australian hypersaline lake. *Appl Environ Microbiol* 79:6755–6764. <https://doi.org/10.1128/AEM.01946-13>.
62. Domingo E, Sheldon J, Perales C. 2012. Viral quasispecies evolution. *Microbiol Mol Biol Rev* 76:159–216. <https://doi.org/10.1128/MMBR.05023-11>.
63. Ignacio-Espinoza JC, Ahlgren NA, Fuhrman JA. 2020. Long-term stability and Red Queen-like strain dynamics in marine viruses. *Nat Microbiol* 5:265–271. <https://doi.org/10.1038/s41564-019-0628-x>.
64. Chénard C, Suttle CA. 2008. Phylogenetic diversity of sequences of cyanophage photosynthetic gene *psbA* in marine and freshwaters. *Appl Environ Microbiol* 74:5317–5324. <https://doi.org/10.1128/AEM.02480-07>.
65. Bragg JG, Chisholm SW. 2008. Modeling the fitness consequences of a cyanophage-encoded photosynthesis gene. *PLoS One* 3:e3550. <https://doi.org/10.1371/journal.pone.0003550>.
66. de Vargas C, Audic S, Henry N, Decelle J, Mahé F, Logares R, Lara E, Berney C, Le Bescot N, Probert I, Carmichael M, Poulain J, Romac S, Colin S, Aury J-M, Bittner L, Chaffron S, Dunthorn M, Engelen S, Flegontova O, Guidi L, Horák A, Jaillon O, Lima-Mendez G, Lukeš J, Malviya S, Morard R, Mout R, Scalco E, Siano R, Vincent F, Zingone A, Dimier C, Picheral M, Searson S, Kandels-Lewis S, Tara Oceans Coordinators; Acinas SG, Bork P, Bowler C, Gorsky G, Grimsley N, Hingamp P, Iudicone D, Not F, Ogata H, Pesant S, Raes J, Sieracki ME, Speich S, Stemann L, et al. 2015. Eukaryotic plankton diversity in the sunlit ocean. *Science* 348:1261605. <https://doi.org/10.1126/science.1261605>.
67. Pichard SL, Paul JH. 1991. Detection of gene expression in genetically engineered microorganisms and natural phytoplankton populations in the marine environment by mRNA analysis. *Appl Environ Microbiol* 57:1721–1727. <https://doi.org/10.1128/AEM.57.6.1721-1727.1991>.
68. Koid A, Nelson WC, Mraz A, Heidelberg KB. 2012. Comparative analysis of eukaryotic marine microbial assemblages from 18S rRNA gene and gene transcript clone libraries by using different methods of extraction. *Appl Environ Microbiol* 78:3958–3965. <https://doi.org/10.1128/AEM.06941-11>.
69. Gustavsen JA, Winget DM, Tian X, Suttle CA. 2014. High temporal and spatial diversity in marine RNA viruses implies that they have an important role in mortality and structuring plankton communities. *Front Microbiol* 5:703. <https://doi.org/10.3389/fmicb.2014.00703>.
70. Lee CK, Herbold CW, Polson SW, Wommack KE, Williamson SJ, McDonald IR, Cary SC. 2012. Groundtruthing next-gen sequencing for microbial ecology—biases and errors in community structure estimates from PCR amplicon pyrosequencing. *PLoS One* 7:e44224. <https://doi.org/10.1371/journal.pone.0044224>.
71. Pinto AJ, Raskin L. 2012. PCR biases distort bacterial and archaeal community structure in pyrosequencing datasets. *PLoS One* 7:e43093. <https://doi.org/10.1371/journal.pone.0043093>.
72. Shade A, Jones SE, Caporaso JG, Handelsman J, Knight R, Fierer N, Gilbert JA. 2014. Conditionally rare taxa disproportionately contribute to temporal changes in microbial diversity. *mBio* 5:e01371-14. <https://doi.org/10.1128/mBio.01371-14>.
73. Suttle CA, Chan AM, Cottrell MT. 1991. Use of ultrafiltration to isolate viruses from seawater which are pathogens of marine phytoplankton. *Appl Environ Microbiol* 57:721–726. <https://doi.org/10.1128/AEM.57.3.721-726.1991>.
74. Parsons T, Maita Y, Lalli C. 1984. A manual of chemical and biological methods for seawater analysis. Pergamon Press, New York, NY.

75. Brussaard CPD. 2004. Optimization of procedures for counting viruses by flow cytometry. *Appl Environ Microbiol* 70:1506–1513. <https://doi.org/10.1128/aem.70.3.1506-1513.2004>.
76. Comeau AM, Krisch HM. 2008. The capsid of the T4 phage superfamily: the evolution, diversity, and structure of some of the most prevalent proteins in the biosphere. *Mol Biol Evol* 25:1321–1332. <https://doi.org/10.1093/molbev/msn080>.
77. Diez B, Pedros-Alio C, Marsh TL, Massana R. 2001. Application of denaturing gradient gel electrophoresis (DGGE) to study the diversity of marine picoeukaryotic assemblages and comparison of DGGE with other molecular techniques. *Appl Environ Microbiol* 67:2942–2951. <https://doi.org/10.1128/AEM.67.7.2942-2951.2001>.
78. Baker G, Smith J, Cowan D. 2003. Review and re-analysis of domain-specific 16S primers. *J Microbiol Methods* 55:541–555. <https://doi.org/10.1016/j.mimet.2003.08.009>.
79. Muyzer G, Teske A, Wirsén CO, Jannasch HW. 1995. Phylogenetic relationships of *Thiomicrospira* species and their identification in deep-sea hydrothermal vent samples by denaturing gradient gel electrophoresis of 16S rDNA fragments. *Arch Microbiol* 164:165–172. <https://doi.org/10.1007/BF02529967>.
80. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
81. Gordon A. 2014. Fastx_toolkit. https://github.com/agordon/fastx_toolkit.
82. Zhang J, Kobert K, Flouri T, Stamatakis A. 2014. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* 30:614–620. <https://doi.org/10.1093/bioinformatics/btt593>.
83. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glockner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 41: D590–D596. <https://doi.org/10.1093/nar/gks1219>.
84. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 75:7537–7541. <https://doi.org/10.1128/AEM.01541-09>.
85. Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26:2460–2461. <https://doi.org/10.1093/bioinformatics/btq461>.
86. Rho M, Tang H, Ye Y. 2010. FragGeneScan: predicting genes in short and error-prone reads. *Nucleic Acids Res* 38:e191. <https://doi.org/10.1093/nar/gkq747>.
87. Montalvo-Proano J, Buerger P, Weynberg KD, van Oppen MJH. 2017. A PCR-based assay targeting the major capsid protein gene of a dinornalike ssRNA virus that infects coral photosymbionts. *Front Microbiol* 8:1665. <https://doi.org/10.3389/fmicb.2017.01665>.
88. Hou W, Wang S, Briggs BR, Li G, Xie W, Dong H. 2018. High diversity of myocyanophage in various aquatic environments revealed by high-throughput sequencing of major capsid protein gene with a new set of primers. *Front Microbiol* 9:887. <https://doi.org/10.3389/fmicb.2018.00887>.
89. Buerger P, Weynberg K, Wood-Charlson E, Sato Y, Willis B, van Oppen MJH. 2019. Novel T4 bacteriophages associated with black band disease in corals. *Environ Microbiol* 21:1969–1979. <https://doi.org/10.1111/1462-2920.14432>.
90. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Wagner H. 2015. vegan: community ecology package. <https://CRAN.R-project.org/package=vegan>.
91. Hijmans RJ. 2015. geosphere: spherical trigonometry. <https://CRAN.R-project.org/package=geosphere>.
92. Johnson LS, Eddy SR, Portugaly E. 2010. Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics* 11:431. <https://doi.org/10.1186/1471-2105-11-431>.
93. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, Thompson JD, Higgins DG. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7:539. <https://doi.org/10.1038/msb.2011.75>.
94. Jia Z, Ishihara R, Nakajima Y, Asakawa S, Kimura M. 2007. Molecular characterization of T4-type bacteriophages in a rice field. *Environ Microbiol* 9:1091–1096. <https://doi.org/10.1111/j.1462-2920.2006.01207.x>.
95. López-Bueno A, Tamames J, Velázquez D, Moya A, Quesada A, Alcámi A. 2009. High diversity of the viral community from an Antarctic lake. *Science* 326:858–861. <https://doi.org/10.1126/science.1179287>.
96. Comeau AM, Arbiol C, Krisch HM. 2010. Gene network visualization and quantitative synteny analysis of more than 300 marine T4-like phage scaffolds from the GOS metagenome. *Mol Biol Evol* 27:1935–1944. <https://doi.org/10.1093/molbev/msq076>.
97. Liu J, Wang G, Wang Q, Liu J, Jin J, Liu X. 2012. Phylogenetic diversity and assemblage of major capsid genes (g23) of T4-type bacteriophages in paddy field soils during rice growth season in Northeast China. *Soil Sci Plant Nutr* 58:435–444. <https://doi.org/10.1080/00380768.2012.703610>.
98. Bellas CM, Anesio AM. 2013. High diversity and potential origins of T4-type bacteriophages on the surface of Arctic glaciers. *Extremophiles* 17:861–870. <https://doi.org/10.1007/s00792-013-0569-x>.
99. Butina TV, Belykh OI, Potapov SA, Sorokovikova EG. 2013. Diversity of the major capsid genes (g23) of T4-like bacteriophages in the eutrophic Lake Kotokel in East Siberia, Russia. *Arch Microbiol* 195:513–520. <https://doi.org/10.1007/s00203-013-0884-8>.
100. Larsson A. 2014. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* 30:3276–3278. <https://doi.org/10.1093/bioinformatics/btu531>.
101. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972–1973. <https://doi.org/10.1093/bioinformatics/btp348>.
102. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490. <https://doi.org/10.1371/journal.pone.0009490>.
103. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>.
104. Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* 27:1164–1165. <https://doi.org/10.1093/bioinformatics/btr088>.
105. Faith DP. 1992. Conservation evaluation and phylogenetic diversity. *Biol Conserv* 61:1–10. [https://doi.org/10.1016/0006-3207\(92\)91201-3](https://doi.org/10.1016/0006-3207(92)91201-3).
106. 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. <https://doi.org/10.1093/bioinformatics/btq166>.
107. 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 Ecol Evol* 8:28–36. <https://doi.org/10.1111/2041-210X.12628>.
108. R Core Team. 2014. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.