



# Marine RNA Virus Quasispecies Are Distributed throughout the Oceans

Marli Vlok,<sup>a,b\*</sup>  Andrew S. Lang,<sup>c</sup>  Curtis A. Suttle<sup>a,b,d,e</sup>

<sup>a</sup>Department of Botany, University of British Columbia, Vancouver, British Columbia, Canada

<sup>b</sup>Institute for the Oceans and Fisheries, University of British Columbia, Vancouver, British Columbia, Canada

<sup>c</sup>Department of Biology, Memorial University of Newfoundland, St. John's, Newfoundland and Labrador, Canada

<sup>d</sup>Department of Earth, Ocean and Atmospheric Sciences, University of British Columbia, Vancouver, British Columbia, Canada

<sup>e</sup>Department of Microbiology and Immunology, University of British Columbia, Vancouver, British Columbia, Canada

**ABSTRACT** RNA viruses, particularly genetically diverse members of the *Picornavirales*, are widespread and abundant in the ocean. Gene surveys suggest that there are spatial and temporal patterns in the composition of RNA virus assemblages, but data on their diversity and genetic variability in different oceanographic settings are limited. Here, we show that specific RNA virus genomes have widespread geographic distributions and that the dominant genotypes are under purifying selection. Genomes from three previously unknown picorna-like viruses (BC-1, -2, and -3) assembled from a coastal site in British Columbia, Canada, as well as marine RNA viruses JP-A, JP-B, and *Heterosigma akashiwo* RNA virus exhibited different biogeographical patterns. Thus, biotic factors such as host specificity and viral life cycle, and not just abiotic processes such as dispersal, affect marine RNA virus distribution. Sequence differences relative to reference genomes imply that virus quasispecies are under purifying selection, with synonymous single-nucleotide variations dominating in genomes from geographically distinct regions resulting in conservation of amino acid sequences. Conversely, sequences from coastal South Africa that mapped to marine RNA virus JP-A exhibited more nonsynonymous mutations, probably representing amino acid changes that accumulated over a longer separation. This biogeographical analysis of marine RNA viruses demonstrates that purifying selection is occurring across oceanographic provinces. These data add to the spectrum of known marine RNA virus genomes, show the importance of dispersal and purifying selection for these viruses, and indicate that closely related RNA viruses are pathogens of eukaryotic microbes across oceans.

**IMPORTANCE** Very little is known about aquatic RNA virus populations and genome evolution. This is the first study that analyzes marine environmental RNA viral assemblages in an evolutionary and broad geographical context. This study contributes the largest marine RNA virus metagenomic data set to date, substantially increasing the sequencing space for RNA viruses and also providing a baseline for comparisons of marine RNA virus diversity. The new viruses discovered in this study are representative of the most abundant family of marine RNA viruses, the *Marnaviridae*, and expand our view of the diversity of this important group. Overall, our data and analyses provide a foundation for interpreting marine RNA virus diversity and evolution.

**KEYWORDS** *Picornavirales*, biogeography, marine RNA virus, purifying selection, quasispecies

Viruses shape the ecology and evolution of marine microbial communities. They are the most abundant biological entities in the ocean and are estimated to kill about 20% of its living biomass each day, thus affecting food web dynamics and biogeochemical cycling (1, 2). Most of these inferences are based on studies of double-

**Citation** 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>.

**Editor** Susannah Green Tringe, DOE Joint Genome Institute

**Copyright** © 2019 Vlok et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Curtis A. Suttle, [suttle@science.ubc.ca](mailto:suttle@science.ubc.ca).

\* Present address: Marli Vlok, Department of Biochemistry and Molecular Biology, University of British Columbia, Vancouver, British Columbia, Canada.

Solicited external reviewers: Valerian V. Dolja, Oregon State University; Yuji Tomaru, Japan Fisheries Research and Education Agency.

This paper was submitted via the [mSphereDirect™](https://mSphereDirect.com) pathway.

 RNA viruses take to the seas. @Viral\_Runner and @viroisphere2012 report that quasispecies are distributed throughout the oceans, with different viruses showing different biogeographical patterns.

**Received** 5 March 2019

**Accepted** 8 March 2019

**Published** 3 April 2019

stranded DNA (dsDNA) viruses, primarily phage. Yet, increasing evidence shows that RNA viruses are also important players in the ecology of marine ecosystems (3).

The first evidence that RNA viruses are pathogens of marine phytoplankton was the isolation of a picorna-like virus that infects the toxic, bloom-forming microalga *Heterosigma akashiwo* (4) and subsequent environmental sequencing that established the enormous diversity in seawater of viruses in the order *Picornavirales*. Targeted sequencing of picorna-like virus RNA-dependent RNA polymerase (RdRp) genes revealed that most sequences fell into a distinct phylogenetic cluster, which are likely viruses primarily infecting protists (5–9). An ensuing metagenomic survey of RNA viruses demonstrated that sequences associated with picorna-like viruses were the most abundant reads that could be assigned (10).

Not only are the abundance and diversity of viruses in the oceans high, but there are patterns in their distribution. Although identical or nearly identical viral genotypes have been documented in geographically and environmentally separate locations (6, 11–13), which is consistent with viral dispersal across vast distances (14), other data show that biogeographical patterns occur in some groups of marine DNA viruses (15). The global distributions of viruses are influenced by virus life cycle (i.e., lytic versus lysogenic/latent), replication rate, length of the latent period, burst size, and the presence of host-derived genes (16, 17) and atmospheric circulation (14). RNA viruses infecting protists are small, with lytic life cycles and large burst sizes and lacking auxiliary metabolic genes (18). For example, the diatom-infecting *Rhizosolenia setigera* RNA virus (RsRNAV) (19) and *Chaetoceros tenuissimus* RNA virus 01 (CtenRNAV01) (20) have different life cycles, which should affect their relative distributions. RsRNAV has a smaller burst size ( $10^3$  versus  $10^4$ ) and a longer infection cycle (2 days versus <24 h), so it should have a lower distribution potential.

Being obligate parasites, the success of virus genotypes (i.e., population growth) is tightly coupled with the virus hosts. Host traits, such as abundance, distribution, size, and physiological state, affect the biogeography of their viruses (16). Strain specificity will also potentially affect distribution, as a broader host range yields more potential hosts. RsRNAV, CtenRNAV01, *Heterocapsa circularisquama* RNA virus (HcRNAV) and *Heterosigma akashiwo* RNA virus (HaRNAV) infect widely distributed taxa, although the viruses are also strain specific (4, 19–21), which reduces the population size of potential host cells and, therefore, the virus distribution potential. Nonetheless, infectious HaRNAV in coastal sediments (22) and the presence of viruses JP-A and JP-B at multiple locations in the coastal waters of British Columbia (23) imply that some viruses are temporally and spatially widespread.

Virus-host coupling and infection specificity drive a coevolutionary arms race (24). RNA viruses are well adapted for this by having high mutation rates due to their error-prone RdRp, which results in approximately  $10^{-4}$  substitutions per nucleotide per replication cycle (25, 26). Thus, a single progenitor generates a distribution of genetically diverse genotypes, a quasispecies, that is clustered around the fittest sequence (27–29). Quasispecies are thought to provide virus populations with built-in flexibility to overcome host and environmental challenges (30–32). While quasispecies are well characterized in RNA viruses associated with higher plants and mammals (33–35), environmental quasispecies are largely unexplored. An analysis of quasispecies in assembled RNA virus genomes from an Antarctic lake suggested that the ecological setting was important in selecting the fittest genotype, as more single-nucleotide variants (SNVs) occurred in lake water than in microbial mats (36). Whether this variability was due to a greater diversity and turnover or because water from multiple locations mixes in the lake is unclear.

Selection ultimately determines which quasispecies genomes are most successful. Under neutral selection, a simple relationship exists between the rates at which mutations are generated and established within a population (37); deviations from neutral selection reveal evolutionary processes such as natural selection (38). For example, RNA viruses in plants that are transmitted by arthropod vectors exhibit lower rates of nonsynonymous ( $dN$ ) changes than synonymous changes ( $dS$ ), indicating

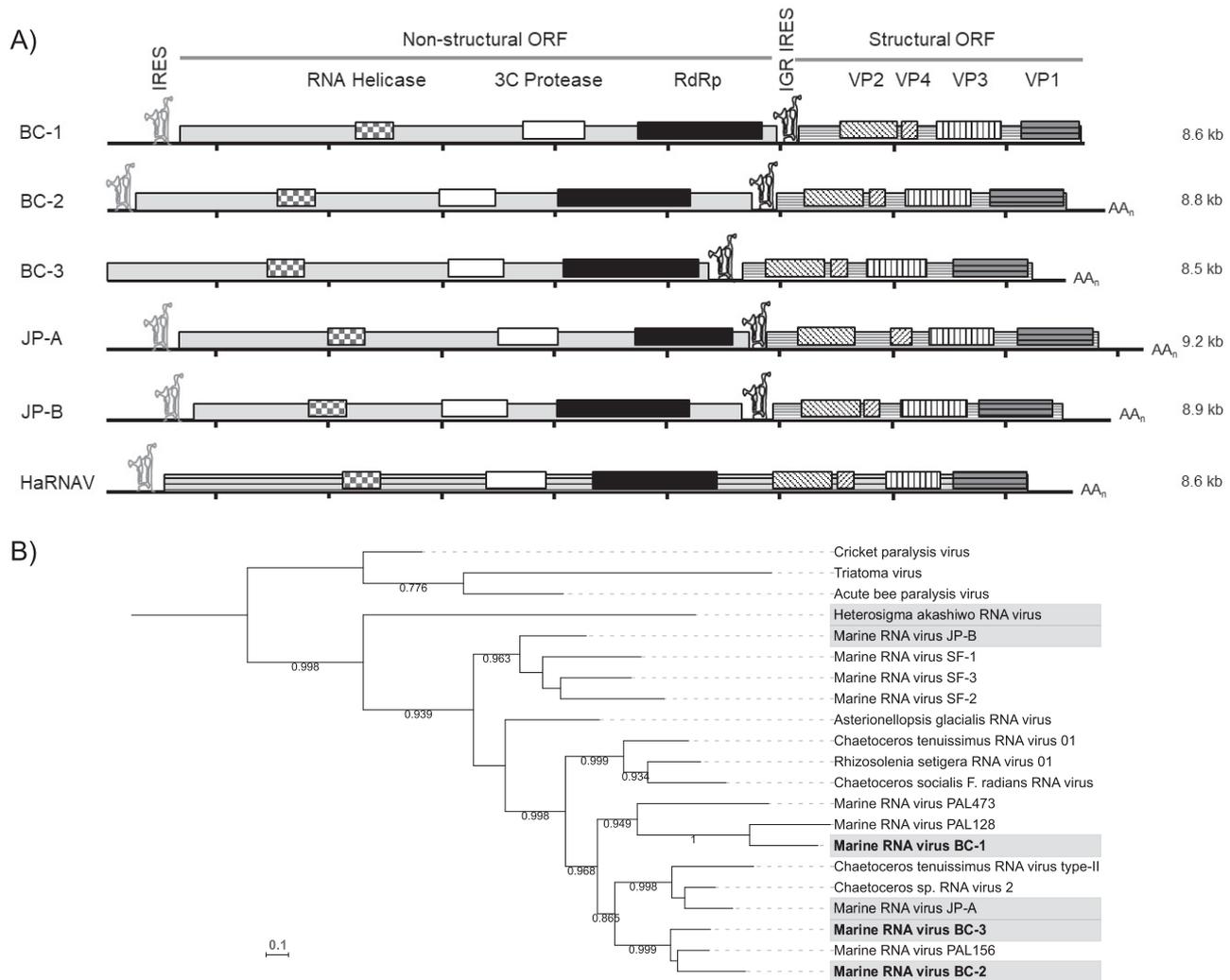
elevated purifying (or negative) selection pressure compared to that of viruses transmitted by other routes, probably due to the required interactions of virus capsids with insect cellular receptors (30). In contrast, influenza A viruses break periods of evolutionary stasis with intervals of positive selection, where  $dN$  exceeds  $dS$ , rapidly replacing circulating lineages with new ones, allowing the virus to escape immune pressure within a host population (39). Marine RNA viruses infecting protists have different obstacles than viruses infecting multicellular organisms. While purifying selection has been proposed for the RdRp of marine picorna-like viruses (40, 41), there is little understanding of the evolutionary pressures facing these viruses or how they respond to them. The known marine single-stranded RNA (ssRNA) viruses are important from both an ecological perspective and a larger evolutionary context. These viruses are part of an order (the *Picornavirales*) of viruses with ancient evolutionary origins and phylogenetic primacy among modern RNA viruses (42, 43). They are also abundant in seawater, with their numbers potentially rivaling those of DNA viruses (3).

Here, we examine six marine RNA viruses in the context of biogeography, evolutionary patterns, and quasispecies potential. HaRNAV, a cultivated virus, infects the toxic bloom-forming raphidophyte *Heterosigma akashiwo* (4), has a 9.1-kb monocistronic genome (44), and is the type species of the family *Marnaviridae* (45). JP-A and JP-B were assembled from metagenomic data, are dicistronic, have 9.2-kb and 8.8-kb genomes, respectively, and based on phylogenetic analysis of RdRp domain sequences, belong within a well-supported clade of marine picorna-like viruses (23); it is believed that these viruses infect protists (9). The other three virus genomes were assembled in this study from metagenomic data. The biogeographic patterns of these six viruses were analyzed across 15 locations, including the Arctic, the Antarctic, the west coasts of North and South America, the Gulf of Mexico, Hawaii, the southern coast of South Africa, two freshwater lakes, and reclaimed water. Based on analysis of SNVs, our results show that these viruses exist as quasispecies and display biogeographical patterns in their distribution.

## RESULTS

Three major findings stem from the results presented here. First, we assembled three previously unknown positive-sense ssRNA viruses from the coastal waters of British Columbia. Second, recruitment analysis of assembled metagenomic data collected across many distant marine and freshwater locations to these genomes, and three others from the same area, demonstrates that close relatives of these viruses occur worldwide. Finally, a detailed analysis of the sequence variation found across locations is consistent with biogeographic patterns and production of quasispecies within these viruses. These results are presented in detail below.

**Assembly of novel marine picorna-like viruses.** Three previously unknown and nearly complete picorna-like virus genomes (BC-1, BC-2, and BC-3) were assembled from metagenomic data collected at Jericho Pier in April 2014 (JP14) (Fig. 1). The viruses were dicistronic, with 8,638-nucleotide (nt) (BC-1), 8,843-nt (BC-2), and 8,496-nt (BC-3)-long genomes and GC contents of 41.2, 42.3, and 43.4%, and were assembled from 11,100, 31,880, and 156,011 reads with coverages of 128.65, 359.36, and 1,816.97 times, respectively. Seven well-conserved motifs were identified in each genome. The 3' end of marine RNA virus BC-1 and the 5' end of BC-3 were not complete, as judged by the lack of untranslated regions (UTRs) (Fig. 1). Poly(A) tails were identified at the 3' ends of the BC-2 and BC-3 genomes after UTRs of 315 and 273 nt, respectively, indicating that these regions were completely sequenced. The 5' UTRs of BC-1 and BC-2 were 634 and 247 nt, respectively and contained putative internal ribosome entry site (IRES) elements. The closest genetic relatives in the NCBI nonredundant database to the three novel viruses were all assembled genomes from metagenomic studies with no host information available. BLASTx analyses of the open reading frames (ORFs) indicated that both ORFs of BC-2 and -3 were most similar to marine RNA virus PAL 156 with E value scores of 0.0 and amino acid identities of 59% and 53% (nonstructural ORFs) and 59% and 60% (structural ORFs), respectively. The BC-1 virus exhibited the highest



**FIG 1** Three nearly complete viral genomes assembled from the Jericho Pier sample. (A) The genomes display the quintessential genome composition common to members of the order *Picornavirales*. The 5' ORFs contain the nonstructural protein domains (RNA helicase [checked rectangles], 3C protease [white rectangles], and RNA-dependent RNA polymerase [RdRp] [black rectangles]), and the 3' ORFs contain the four structural protein domains (striped rectangles) and are preceded by predicted intergenic region (IGR) IRES elements. The four capsid domains are organized similarly to those of HaRNAV, with VP2 at the 5' end, followed by VP4, VP3, and VP1. (B) A maximum likelihood phylogeny of the RdRp gene. The three dicistrovirus type species were used as an outgroup. The six genomes used in this analysis are highlighted in gray.

similarity to marine RNA virus PAL 128, with E values of 0.0 for both ORFs by BLASTx and amino acid identities of 51% and 45% for the nonstructural and structural ORFs, respectively. These viruses were used for all subsequent analyses, along with JP-A and JP-B, which also originated from Jericho Pier (6), and HaRNAV, which was isolated from the nearby Fraser River plume (4).

#### Recruitment of assembled metagenomic reads to marine RNA virus genomes.

The geographical distributions of the six viruses were explored using 17 RNA virus metagenomic data sets (15 generated in the present study and 2 from previous studies). Library size potentially affects the detection of rarer viral genomes, but there was no trend between the number of reads recruited to genomes and library size (see Fig. S1 in the supplemental material), suggesting that this was not an issue for our data. The recruitment analysis was not competitive, meaning that a single read could theoretically be recruited up to six times, once to each genome. However, on average, each read that was recruited from a metagenomic data set to a genome was recruited only 1.0 to 1.5 times. Data from Palmer Station and Nunavut 3 were exceptions (Fig. S1) to the observed trend. Overall, our analysis indicates that the same reads did not map to all six genomes in each analysis.

To assess the distribution of close relatives of the six marine RNA viruses, reads were recruited to their genomes from 17 marine and 7 fresh and reclaimed water RNA virus metagenomic data sets collected from widespread locations (Fig. 2; Fig. S2 and S3). The amino acid similarity and distribution of metagenomic reads across the genomes varied markedly among locations and for each virus. Even for data sets collected from the location from which the virus was identified, detection of reads mapping to the genomes was sporadic; however, for the viruses BC-1, BC-2, and BC-3, identical reads (Fig. 2A) were assembled from the two JP14 data sets. In contrast, the JP14 samples yielded very few reads matching HaRNAV (Fig. 2A and 3; Fig. S3O and S3P), while for JP-A and JP-B they ranged between 40 and 90% identity and were predominantly within the RdRp domain. In contrast, many reads from the JP13 sample were similar to those from HaRNAV (80 to 98% amino acid identity), while few matched JP-A, JP-B, or BC-3 (Fig. S3Q).

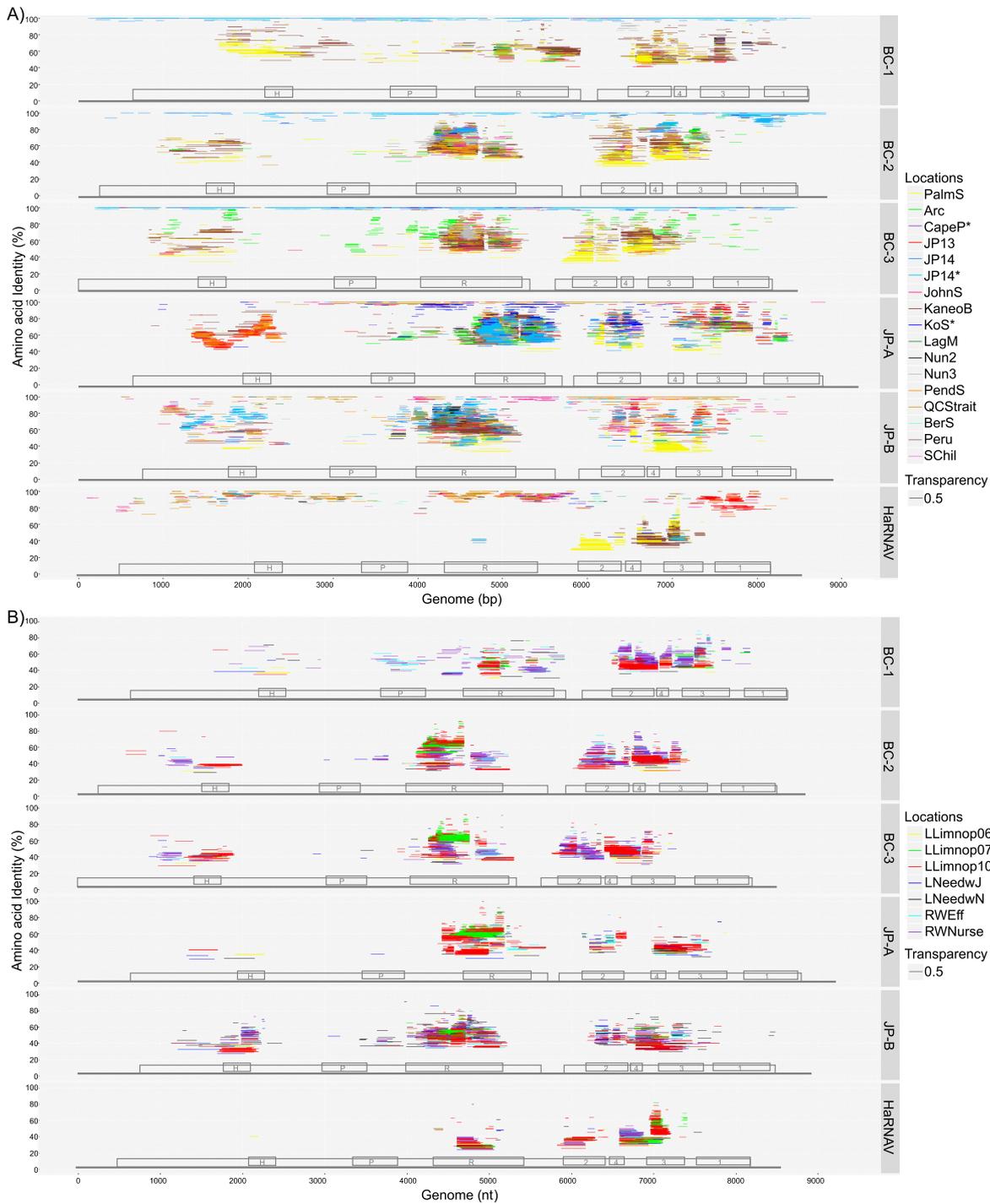
Nonetheless, deduced amino acid sequence with high levels of identity to the six viruses were detected at numerous locations (Fig. 2A; Fig. S2). For example, JP-A was detected at 95 to 100% identity in samples from Queen Charlotte Strait (QCStrait) and Johnstone Strait (JohnS) (Fig. 2A and 3; Fig. S2, S3F, and S3N). The coverage spanned more than 3 kb of the genome (3,000 to 7,800 nt from QCStrait and 3,000 to 6,000 nt from JohnS). Similarly, at Johnstone Strait, 2 kb of the JP-B genome was covered by reads with 100% amino acid identity. These locations also had reads corresponding to BC-2, BC-3, and HaRNAV, but these were fewer than for the JP viruses and more sporadically distributed across the genomes. Reads with 80 to 100% amino acid identities to HaRNAV were detected in the Peru, Nunavut, Bering Sea, Laguna Madre, and South Chile samples (Fig. 3). Like HaRNAV, reads mapping to JP-A occurred beyond coastal British Columbia, and 60 to 100% amino acid identities were detected in the Kenton-on-Sea (KoS) metagenome from South Africa (Fig. S3L).

Many reads in the JP14 sample mapped with high identity to BC-1, -2, and -3; however, there was wide variation in identity levels across the genomes (Fig. 2A; Fig. S2). The pattern is exemplified by the VP1 (capsid protein) domain region of BC-2, where many sequences that mapped to this small region differed by as much as 15% amino acid identity (Fig. S2). A similar pattern was observed for the HaRNAV genome in the JP13 sample, where a large number of sequences also mapped to the VP1 domain, although the range in amino acid sequence identities was even greater.

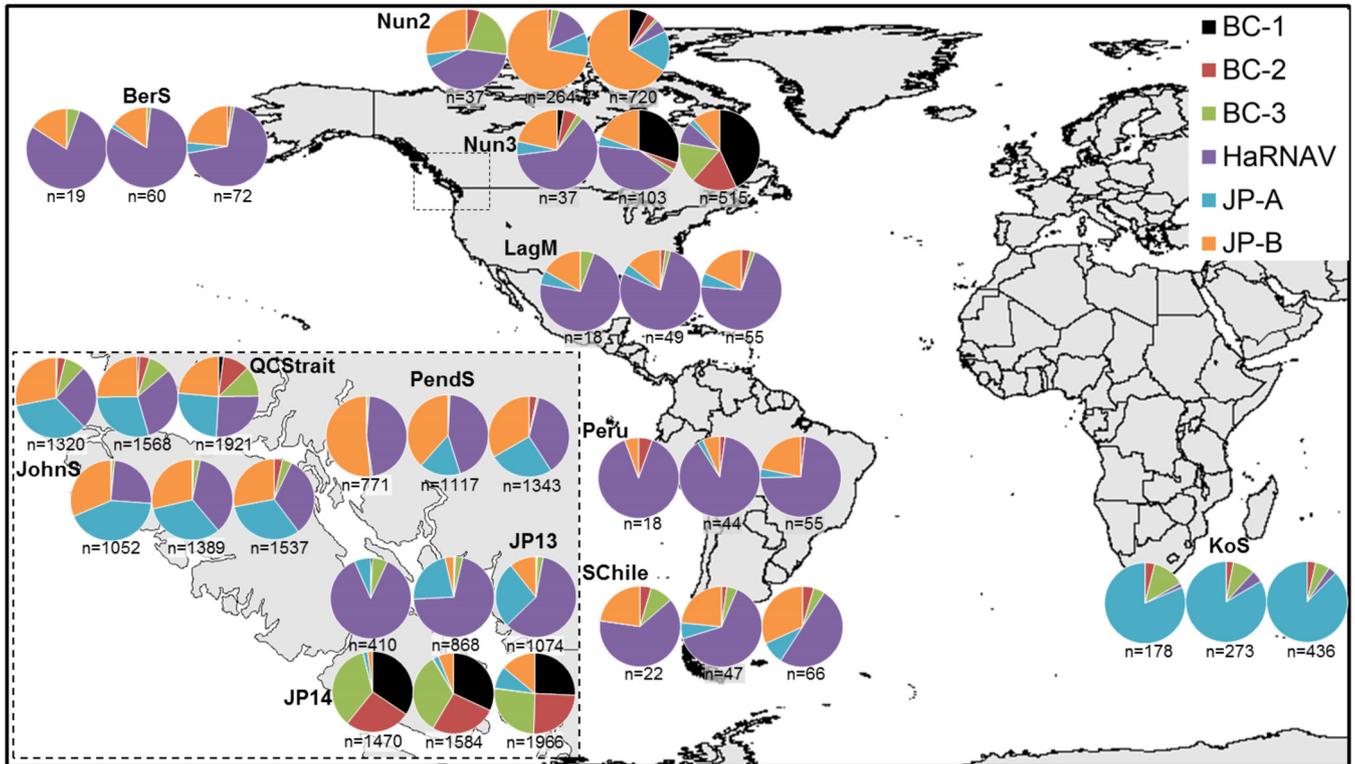
Across all locations, a large proportion of the reads showed low levels of amino acid sequence identity to the six viruses, ranging from about 20% to 80% across both structural and nonstructural domain regions, although many reads mapped to RdRp domains. BC-1 and HaRNAV were exceptions, and fewer environmental reads mapped to their genomes overall (see Fig. 5). Generally, of the seven recognizable domains, the protease and VP1 regions had the fewest identified environmental reads.

Many reads from the assembled metagenomic data obtained from the five lakes and two reclaimed water sites recruited to the six marine RNA virus genomes, some with identities as high as 90%, although most were between 30 and 60% (Fig. 2B). As in the marine recruitment analysis (Fig. 2A), most of the reads recruited to regions associated with the helicase, RdRp, and VP-2, -3, and -4 domains.

**Domain-specific nucleotide variation.** In order to infer selection pressures on the viral genomes, we examined nucleotide substitutions in six of the conserved domains for each of the viruses. Based on their consensus sequences, all of these domains were subject to purifying or negative selection (Fig. 4A). Values of  $dN/dS$  of  $<1.0$  for all data below the 75th percentile indicates that purifying selection was acting on these sequences. No difference was observed among the calculated means for the six domains (analysis of variance [ANOVA],  $P = 0.2507$ ), but based on the medians, all groups did not have the same distributions of  $dN/dS$  ratios (Kruskal-Wallis rank sum test,  $P = 0.0385$ ). Multiple pairwise comparisons showed that the VP1 domain had a statistically significant larger median  $dN/dS$  ratio distribution than the RdRp and VP2



**FIG 2** Sequences mapping to the six marine RNA viruses in the environmental viral metagenomic data sets. Fragments were recruited from marine RNA virus metagenomic data sets that were either generated during this study (Arctic [Arc], Cape Point [CapeP], Jericho Pier 13/14 [JP], Johnstone Strait [JohnS], Kenton-on-Sea [KoS], Laguna Madre [LagM], Nunavut 2/3 [Nun], Pendrell Sound [PendS], Queen Charlotte Strait [QCStrait], Bering Sea [BerS], Peru, and South Chile [SChil]) (A) or were from three freshwater data sets from Lake Limnopolar (LLimnop; Antarctica), Lake Needwood (LNeedw; MD, USA), and reclaimed water from Florida (RWEff and RWNurse; USA) (B). Metagenomic reads were recruited against each of the six virus genomes using tBLASTx with an E value of  $10^{-10}$ . The position of the colored bars on the y axis indicates the percent amino acid sequence identity to the virus, while the width of the bar covers the region of the genome where the similarity was observed. The bars are displayed at 0.5 transparency so that overlap among locations can be observed. Genome maps for each virus are superimposed on the plot with the domains labeled as follows: H, helicase; P, protease; R, RNA-dependent RNA polymerase; and 1, 2, 3, and 4, capsid VP1, VP2, VP3, and VP4, respectively.

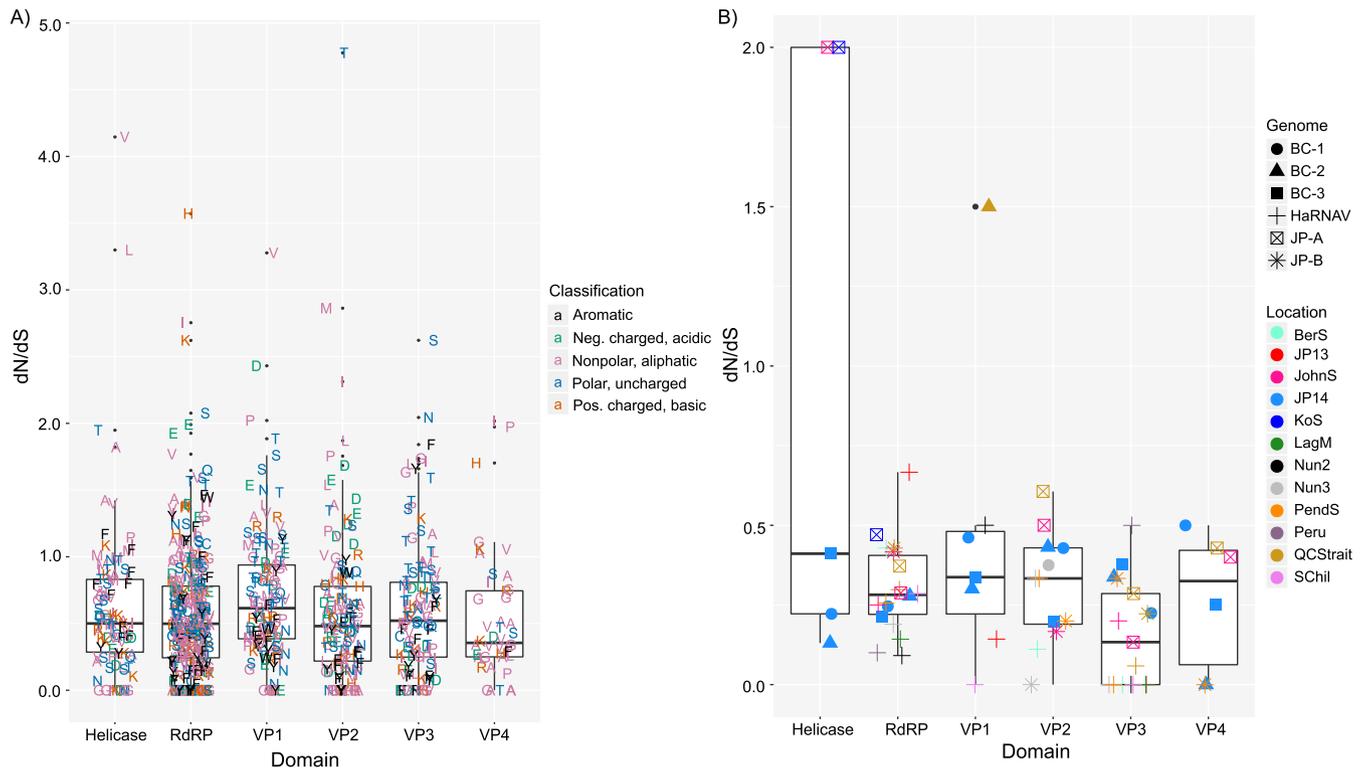


**FIG 3** Distribution of six marine RNA virus genomes. Charts represent the number of reads recruited to each genome at 95, 85, and 75% protein identity with the number of total reads recruited below each chart. Genomes are represented by colors for marine RNA viruses BC-1 (black), BC-2 (red), BC-3 (green), JP-A (blue), JP-B (orange), and *Heterosigma akashiwo* RNA virus (purple). Metagenomic data sets were selected based on high percentage amino acid identity read recruitment to the six viruses and include the Bering Sea (BerS), Nunavut 2/3 (Nun), Laguna Madre (LagM), Peru, South Chile (Schil), Kenton-on-Sea (KoS), Queen Charlotte Strait (QCStrait), Pendrell Sound (PendS), Johnstone Strait (JohnS), and Jericho Pier 13/14 (JP).

domains (Conover-Iman test;  $P = 0.0364$  and  $0.0442$ , respectively) (Table S2). Several codons were under neutral selection and a few more were under positive selection, indicating that the latter sites are more variable. All groups of amino acids (aromatic, acidic, basic, aliphatic, and uncharged) were subject to the same evolutionary pressure (Fig. S4).

Selection analyses of the metagenomic data that mapped to the six genomes reflected purifying selection acting on these viruses (Fig. 4B). There were no significant differences among the domains (ANOVA,  $P = 0.05343$ ; Kruskal-Wallis rank sum,  $P = 0.09858$ ). Calculated  $dN/dS$  ratios were very low apart from outliers for the VP1 domain of virus BC-2 in Queen Charlotte Strait and the helicase domain of JP-A in Johnstone Strait and Kenton-on-Sea.

Single-nucleotide-variant (SNV) analyses (Fig. 5) indicated that clouds of diverse variants for each of the six marine viruses existed in different samples. Because two sequence libraries were prepared from the Jericho Pier 14 sample, and most SNVs occurred in both libraries, this suggests that sequencing errors were not a significant source of the variation. Sequences mapping to viruses BC-1 and -2 were abundant in the Jericho Pier 14 sample, and many SNVs were distributed across the RdRp, VP2, and VP3 domains, at frequencies of 1 to 50% (Fig. 5A, B, and C). Most variants were synonymous, except in the VP2 domain, which had more nonsynonymous mutations. In contrast, SNVs in the same three domains of BC-3, as well as in its helicase domain, occurred at much lower frequencies (1 to 12%). Though fewer SNVs were observed across the domains of BC-1 and -2, these SNVs occurred more often within the quasispecies, resulting in a higher frequency of occurrence (Fig. S5A). The VP1 domain of BC-3 exhibited a similar trend of low-frequency mutations, except for three variants that occurred at considerably higher frequencies, two of which were nonsynonymous.

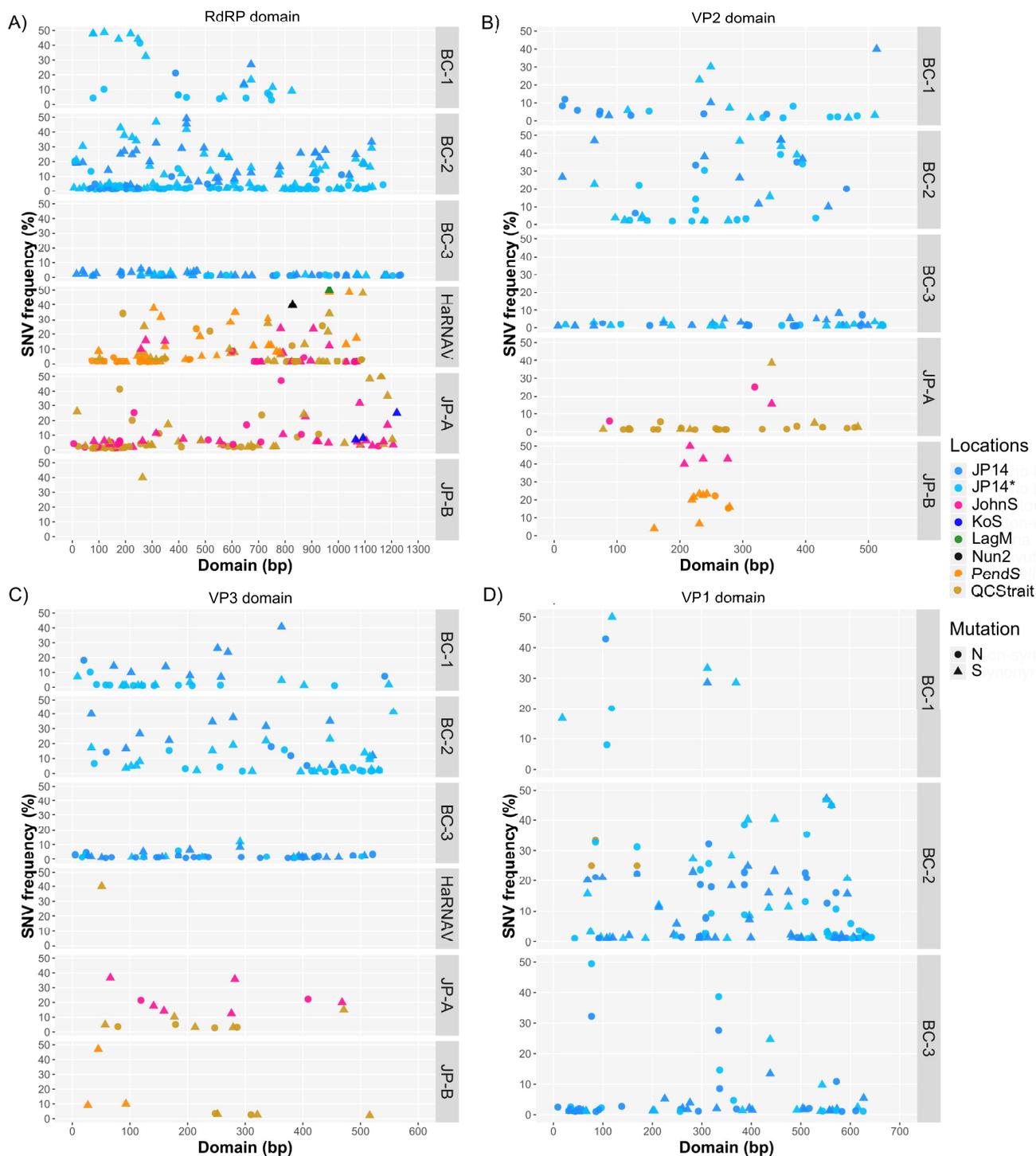


**FIG 4** Selection analyses of marine RNA virus domains in environmental data sets. (A) Maximum likelihood analysis of selection pressure based on the codons of marine RNA viruses BC-1, BC-2, BC-3, JP-A, JP-B, and HaRNAV. The substitutions were estimated using the joint maximum likelihood reconstructions of ancestral states under a Muse-Gaut model (86) of codon substitution and the Felsenstein 1981 model (87) of nucleotide substitutions. Amino acids were classified as aromatic (phenylalanine [F], tryptophan [W], tyrosine [Y]), negatively charged (-), acidic (aspartic acid [D], glutamic acid [E]), nonpolar, aliphatic (alanine [A], glycine [G], isoleucine [I], leucine [L], methionine [M], proline [P], valine [V]), polar, uncharged (asparagine [N], cysteine [C], glutamine [Q], serine [S], threonine [T]), and positively charged (+), basic (arginine [R], histidine [H], lysine [K]). (B) Selection pressure acting on the various domains of the six marine RNA viruses at various locations. Single-nucleotide variants (SNVs) were called using the quality-based variant detection tool (CLC Genomics Workbench) from alignments in which environmental sequences with  $\geq 90\%$  sequence identity were mapped to the genomes. The helicase domain was not analyzed due to insufficient data. Two additional virus domain-sample pairings were omitted due to insufficient data (VP1 of both BC-2 and -3 from the Nun2 and QCStrait samples, respectively).

Virus BC-1 had inadequate coverage in the VP1 domain to get a good profile of SNVs across the domain, but some high-frequency mutations toward the 5' end of the domain were observed. The VP1 domain of virus BC-2 had many SNVs distributed across the domain, at frequencies of 1 to 47%.

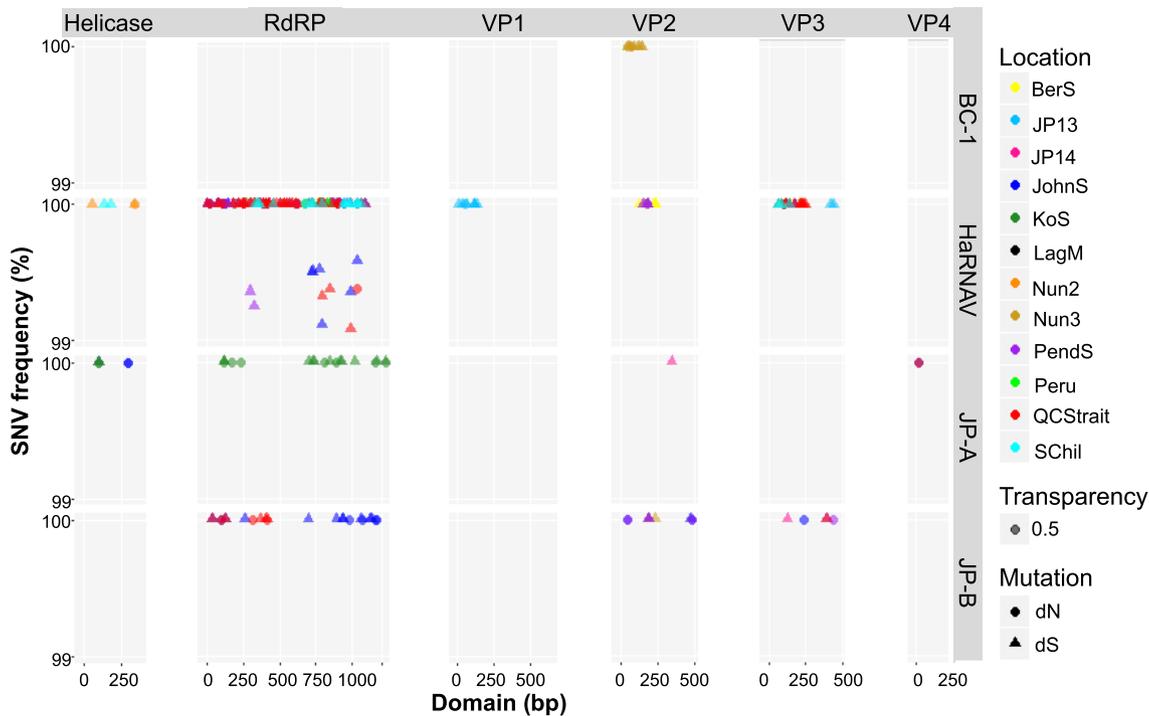
**Detection of established viral variants.** Single-nucleotide variants that were more frequent than 99% were abundant and detected at multiple locations (Fig. 6). These are variants from the original isolate/genome that have clearly become established in the current population and were most common in the RdRp domains. JP-B variants were detected in Queen Charlotte Strait (8 SNVs) and Johnstone Strait (15 SNVs), and variants of HaRNAV occurred at both of these sites (50 and 56 SNVs, respectively) and at Pendrell Sound (20 SNVs), South Chile (18 SNVs), Laguna Madre (28 SNVs), Bering Sea (10 SNVs), Jericho Pier 13 (10 SNVs), Nunavut 2/3 (33 and 43 SNVs, respectively), and Peru (7 SNVs). Note that the data for the latter four locations are partially obscured in Fig. 6. In contrast, high-frequency SNVs of JP-A were detected exclusively from Kenton-on-Sea (18 SNVs).

Unlike the RdRp domain, fewer high-frequency SNVs were associated with other domains, especially for the helicase, VP4, and VP1. SNVs that were concentrated in the same region of the VP3 domain of HaRNAV were identified in data from Bering Sea (3 SNVs), Jericho Pier 13 (2 SNVs), Johnstone Strait (6 SNVs), Laguna Madre (2 SNVs), Nunavut 2/3 (6 and 2 SNVs, respectively), Pendrell Sound (2 SNVs), Peru (3 SNVs), Queen Charlotte Strait (14 SNVs), and South Chile (4 SNVs).



**FIG 5** Detection of marine RNA virus quasispecies in environmental data sets. The distribution of single-nucleotide variants (SNVs) along the RdRp (A), VP2 (B), VP3 (C), and VP1 (D) domains of marine RNA viruses BC-1, BC-2, BC-3, HaRNAV, JP-A, and JP-B. Triangles (▲) indicate synonymous mutation variants, and circles (●) indicate nonsynonymous variants. The locations/samples where variants were detected are illustrated in different colors and include Jericho Pier (JP; light blue), Johnstone Strait (JohnS; pink), Kenton-on-Sea (KoS; royal blue), Laguna Madre (LagM; green), Nunavut 2 (Nun2; black), Pendrell Sound (PendS; gold), and Queen Charlotte Strait (QCStrait; brown).

Of the three new genomes, BC-1 was the only one in which high-frequency SNVs were identified; SNVs in the VP2 domain were detected from Nunavut 3 (11 SNVs). SNVs in the VP2 domain of HaRNAV were also detected at Nunavut 3 (4 SNVs), Bering Sea (9 SNVs), and Pendrell Sound (4 SNVs). Detection of SNVs in multiple domains of a single



**FIG 6** Established viral variants in different marine RNA virus communities. The distribution of high-frequency (>99%) single-nucleotide variants (SNVs) along the respective domains of the BC-1, HaRNAV, JP-A, and JP-B viruses is shown. Triangles (▲) indicate synonymous mutation variants, and circles (●) indicate nonsynonymous variants. The locations/samples where variants were detected are illustrated in different colors and include Bering Sea (BerS; yellow), Jericho Pier 13/14 (JP; blue/pink), Johnstone Strait (JohnS; royal blue), Kenton-on-Sea (KoS; green), Laguna Madre (LagM; black), Nunavut 2/3 (Nun; orange/brown), Pendrell Sound (PendS; purple), Peru (green), Queen Charlotte Strait (QCStrait; red), and South Chile (SChil; teal).

virus at specific locations supports the existence of biogeographical variations for the virus, with established genome variants existing in these locations.

## DISCUSSION

Metagenomic analysis uncovered sequences for three previously unknown marine picorna-like RNA viruses (BC-1, BC-2, BC-3). The distributions of these and three related viruses (JP-A, JP-B, HaRNAV) (23, 44) within global RNA virus metagenomic data sets revealed striking differences, even though the viruses originated from the same area. An analysis of sequence variation revealed that these viruses exist as quasispecies and that purifying selection is the predominant evolutionary process acting on their genomes. These observations and their implications are discussed in detail below.

**Discovery of three divergent marine picorna-like viruses.** Three nearly complete genomes (BC-1, -2, -3) were assembled from the Jericho Pier 14 sample (Fig. 1), adding to genomes for HaRNAV, JP-A, and JP-B that originated from water collected in the same area (23, 44). BC-2 and -3 are most genetically similar, and while BC-1 shares similarity with BC-2 and -3, it also has sequence similarity with JP-A. Among the six viruses, HaRNAV is the most distantly related to the others (Fig. 1; see Fig. S6 and S7 in the supplemental material), consistent with previous phylogenetic analyses of RdRp sequences that place HaRNAV as relatively divergent within marine picorna-like viruses (7). Most marine picorna-like virus isolates infect diatoms (*Bacillariophyta*), which are abundant, genetically diverse, and globally distributed. In contrast, the host of HaRNAV is *Heterosigma akashiwo*, a member of the *Raphidophyta*, a family comprised of relatively few and less diverse protists (46); HaRNAV is the only RNA virus isolated that infects a raphidophyte. Therefore, the greater genetic distance between HaRNAV and other marine picorna-like viruses probably reflects the lack of other viruses infecting raphidophytes.

Sequence divergence is different among the viral sequence domains. In particular, the protease is most divergent, whereas VP1 is the most divergent among the conserved structural proteins (VP1, VP2, and VP3). This is congruent with read recruitment (Fig. 2), in which fewer reads were detected for the VP1 and protease domains. VP1 can be involved in host recognition (47), and the protease modifies the host proteome during infection (48, 49). Thus, both proteins are likely tightly coupled with the host and more sequence divergence in these domains would be expected among different viruses.

Although all metagenomic data sets were interrogated for complete genomes, the three new genomes originated from a single sample, Jericho Pier 14. This may be because Jericho Pier 14 was the newest sample and the RNA genomes may have been less degraded than in the older samples (Table S1), which were stored at 4°C as virus concentrates prior to processing.

**Marine RNA virus biogeography.** The six viruses were selected for this study because they were all discovered in waters adjacent to Vancouver. Despite the large number of reads from Jericho Pier 14 that recruited to the three new genomes (BC-1, -2, and -3), few reads recruited to these genomes from a sample (Jericho Pier 13) collected a year earlier from the same location (Fig. 2A). In contrast, many more reads recruited to HaRNAV from the Jericho Pier 13 sample than from the Jericho Pier 14 sample. This shows that the relative abundances of these viruses are dynamic at the Jericho Pier site and is consistent with differences observed in the relative abundances of specific RdRp sequences at this location and nearby (7, 50).

Reads mapping to the viral genomes were distributed across wide geographical and environmental distances. Large portions of the JP-A and JP-B genomes were covered by reads from Queen Charlotte Strait and Johnstone Strait (Fig. 2A), whereas the coverage was much less for the BC-2, BC-3, and HaRNAV genomes. The variability in relative coverage of the viral genomes among metagenomic data sets from coastal British Columbia emphasizes the dynamic nature of the RNA viral communities across space and time. It shows that despite the oceanographic connectivity among these sites and the high dispersal potential of marine viruses, environmental and host-driven factors are critical in dictating the composition of marine RNA virus assemblages. This is consistent with the influences of host traits and environmental conditions on virus distribution (16) and metacommunity analysis of microbial communities in the Arctic Ocean (51). For example, HaRNAV infects a widely distributed host (*H. akashiwo*) that is sporadic in occurrence (52–56), explaining the widespread distribution of sequences matching this virus. However, HaRNAV exhibits host strain specificity (4), which explains why high-identity sequences mapping to large portions of the genome were not found across all samples. In contrast, JP-A, BC-2, and BC-3 show high similarity to viruses infecting diatoms in the genus *Cheatoceros* (Fig. 1B), which are much more abundant worldwide, especially in temperate waters. Therefore, it is not surprising that the sequence coverage for these viruses is typically much greater than that for HaRNAV.

The widespread presence of low-identity reads in coastal waters across locations indicates that distant relatives of these six viruses are widely distributed (Fig. 2). The large number of low-identity sequences mapping to the highly conserved RdRp domains is also consistent with the large amount of genetic diversity observed in gene surveys of RdRp sequences (5–8, 57). However, our data for full genomes show that there is also widespread diversity associated with the helicase and capsid domains (Fig. S2). Although the RdRp domain is generally the most well-conserved domain, in many samples there was discordance among the domains with respect to the number of sequences detected. For example, for the JP-A virus, many sequences from the JP13 sample mapped to the helicase region (Fig. S3Q), whereas few were associated with the RdRp domain. This could reflect different evolutionary pressures acting on the two domains or that recombination has occurred, both of which have been observed in picornaviruses (58). Regardless of the mechanism of diversification, a high degree of sequence richness was observed in viruses related to the six investigated here that

occurred across widely separated regions. In contrast to the other viruses, very few low-identity sequences mapped to HaRNAV, perhaps because this virus infects hosts that have relatively low species richness. Conversely, the high diversity of sequences related to those from viruses assembled from metagenomic data suggests that these sequences are from viruses that infect hosts with greater genetic diversity.

The freshwater metagenomes (Fig. 2B) provide context for interpreting the marine data sets. The same host species, and therefore the same viruses, are unlikely to occur in fresh and marine waters. For example, there are no known freshwater *Heterosigma* species, but there are freshwater raphidophytes. Therefore, the amino acid sequences with an inferred relatively low identity of <70% found in freshwater metagenomic data sets that match HaRNAV sequences likely represent distant relatives of this virus that infect other species of raphidophytes. Furthermore, read recruitment from freshwater streams ( $n = 14$  months) emptying into the marine water body from which the BC viruses were assembled suggests that these viruses are not of freshwater origin, as very few, if any, reads could be recruited (unpublished data).

**Purifying selection acts on the marine RNA virus quasispecies.** RNA virus replication results in “clouds” of sequence variants referred to as quasispecies upon which selection operates. In turn, selection is often viewed through the lens of purifying or negative selection, or the extent to which deleterious alleles are selectively removed from the population. Elevated purifying selection is commonly inferred from lower ratios of nonsynonymous ( $dN$ ) to synonymous changes ( $dS$ ) of nucleotides in a genome. Our data imply that purifying selection acts on both structural and nonstructural domains of the virus genomes, with overall  $dN/dS$  values of <1 (Fig. 4), consistent with postulated purifying selection in the RdRp domain (40). Purifying selection constrains the possible amino acid substitutions, so that most nucleotide changes conserve the amino acid sequence, as frequently observed in RNA viruses (59–61). Purifying selection is evident in the domain alignments, in which many amino acids that are conserved among the six genomes are encoded by alternate codons (Fig. S8). The capsid VP1 domain is an exception (Table S2), suggesting that this domain may be important in host recognition, as discussed earlier, and would therefore differ among viral taxa based on their hosts’ receptors.

The existence of quasispecies was validated for the six viruses by examining single-nucleotide variants (SNVs) in the sequence data from different locations. Not surprisingly, because they were assembled from the Jericho Pier 14 samples, BC-1, -2, and -3 had the greatest genome coverage, yet BC-3 recruited about four times more data than BC-1 and -2 but fewer SNVs for the RdRp, helicase, VP2, and VP3 domains (Fig. 5A, B, and C; Fig. S5A).

If a population of viruses contains an abundance of the fittest genotype surrounded by rarer quasispecies variants, most genotypes should resemble the original virus that replicated. This suggests that BC-3 stemmed from a lytic event that occurred shortly before sample collection, resulting in higher abundance and lower diversity. Similarly, an analysis of quasispecies in an Antarctic lake revealed that an abundant RNA virus genotype, APLV1, had few SNVs, with most occurring in <10% of the sequences, despite the virus being ecologically successful (36).

The patterns in the SNVs in the environmental data that mapped to JP-A and HaRNAV indicated that quasispecies-like populations occurred at locations beyond the original site of discovery (Fig. 5A). For example, SNVs were detected across the RdRp domain for HaRNAV and JP-A at Queen Charlotte Strait and Johnstone Strait and for HaRNAV at Pendrell Sound. Very few reads from Nunavut, Laguna Madre, or Kenton-on-Sea mapped to these genomes with sufficient similarity to call SNVs across the RdRp domain, but the detected variants were all synonymous and exhibited different frequencies, with the frequency of the Kenton-on-Sea variants being much lower than that of the Nunavut and Laguna Madre variants. While it is tempting to hypothesize that quasispecies are broadly distributed, there were too few matching sequences from Nunavut, Laguna Madre, and Kenton-on-Sea to conclusively state that they are indic-

ative of quasispecies. Realistically, it is more likely that these sequences correspond to similar viruses, some of which may not infect the same host.

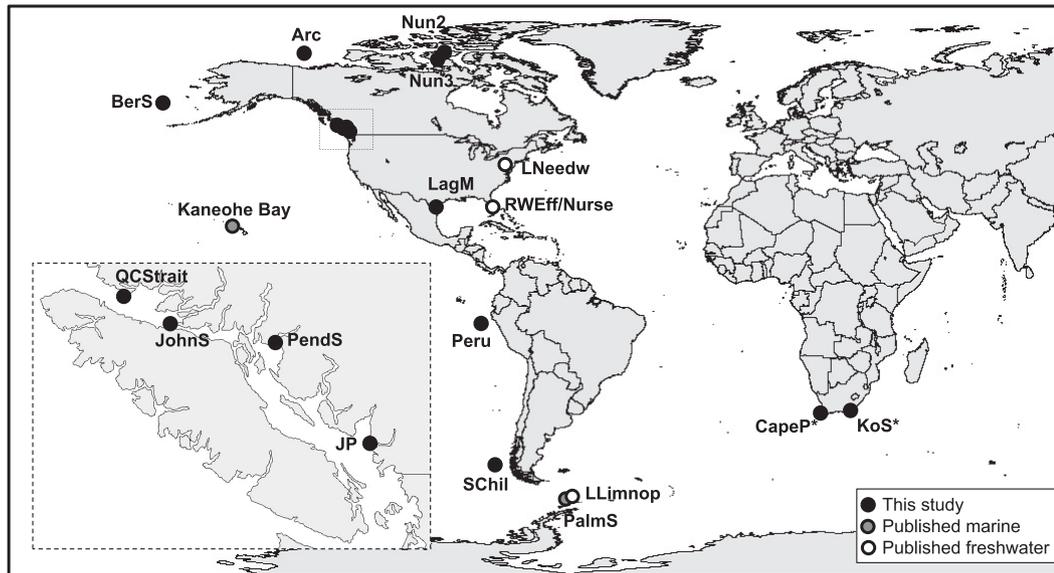
For the BC-3 virus, only the VP1 domain had enough variation to suggest a possible quasispecies (Fig. 5D). This domain is probably important for binding to the host receptor (47, 62, 63); thus, virus-host coevolution would affect selection on this domain. For the VP1 from BC-3, three of the four SNVs detected in both Jericho Pier 14 libraries coded for amino acid changes. Similarly, more low-frequency nonsynonymous SNVs were found in the VP1 domains of BC-1 and BC-2 (Fig. 5D). Both functional and structural studies suggest that RNA viruses tolerate restricted numbers and types of mutations (31). For example, during large-population passages of foot-and-mouth disease virus (FMDV), numerous nonsynonymous mutations occurred in the capsid protein, with 96% being on the protein surface (64). Without structural homology models for the proteins, it is not known if the nonsynonymous mutations correspond to sites located on the outside of the virus particle. Overall, the observed differences, along with the higher  $dN/dS$  ratio previously discussed, suggest that VP1 domains experience different degrees of evolutionary pressure in comparison to the other capsid proteins, perhaps due to the role of VP1 in host recognition.

**Established viral variants are widely dispersed.** Given the coevolutionary arms race between viruses and their hosts, and the competitive advantage of existing as quasispecies, some low-frequency variants should be selected for and become established as the fittest genotype. This would result in turnovers in the dominant viral genotypes, as frequently occurs in viruses of multicellular organisms (32, 65, 66). To investigate how HaRNAV, JP-A, and JP-B have changed since they were first sequenced, the high-frequency SNVs (>99%) detected in the environmental data sets were analyzed. High-frequency SNVs in the RdRp of HaRNAV indicate that many of the established genetic variants occur at geographically distant locations, including Bering Sea, Johnstone Strait, Queen Charlotte Strait, Pendrell Sound, Laguna Madre, Peru, South Chile, and Nunavut 2/3 (Fig. 6). Similarly, high-frequency variants of JP-B were also detected at geographically distinct locations (Johnstone Strait and Queen Charlotte Strait), but unlike HaRNAV, where location-specific sequences were frequently found for more than one domain, the domains and locations were for the most part distinct for JP-B. A similar discord was observed for JP-A, where only the RdRp domain was covered (Fig. 6). Despite inconsistent coverage across genomes, geographically distinct established variants were detected for multiple genotypes, indicating that turnover in the dominant genotype occurs.

Most SNVs in the HaRNAV and JP-B genomes were synonymous, indicating pressure to conserve amino acid sequences. In contrast, the SNVs in the RdRp of JP-A were more equally distributed between synonymous and nonsynonymous. The JP-A RdRp sequences originated from samples collected at Kenton-on-Sea, on the southeast coast of South Africa, which is very distant from the coast of southwestern Canada, where the JP-A genome originated. The large proportion of nonsynonymous substitutions suggests that the Kenton-on-Sea variant is evolutionarily relatively distant from the JP-A genome, emphasizing that marine picorna-like viruses are diverse and widespread in the world's oceans.

**Conclusions.** Assembly of marine RNA virus metagenomic data from a single metagenomic sample in British Columbia revealed three previously unknown viruses, BC-1, -2, and -3. Recruitment of sequences from 17 metagenomic data sets to these and other previously characterized marine RNA virus genomes uncovered biogeographic patterns and evidence for quasispecies and purifying selection in marine RNA viruses across distant locations.

The occurrence of genetically related viruses differed across oceans and hemispheres, implying that biotic factors such as virus life cycles and host traits, rather than solely abiotic factors, affected their distribution. In addition, differences in community composition across locations and years attest to the dynamic nature of marine RNA



**FIG 7** Sampling locations represented in this study. The samples represent multiple depths and time points from geographically distinct locations and ocean climate zones. Metagenomes from 14 locations were analyzed here (black) along with published marine data (gray) from Kaneohe Bay (KaneoB), Hawaii (57), and Palmer Station (PalmS), Antarctica (8), freshwater data sets (white) generated from samples from Lake Limnopolar (LLimnop), Antarctica (36), and Lake Needwood (LNeedw), USA (82), and reclaimed effluent and nursery water (RWEff and RWNurse) from Florida, USA (83). See the text for other location abbreviations.

virus assemblages and are consistent with the idea that these dynamics probably result from episodic infections (7).

Numerous sequences with low amino acid identities to the six viruses were detected, revealing a great richness of globally distributed, distantly related viruses. Single-nucleotide variation was also observed in structural and nonstructural protein domains, consistent with purifying selection; thus, despite the error-prone RdRp and its lack of proofreading, these sequences are constrained to conserve their function. However, the VP1 domain had a statistically higher  $dN/dS$  ratio than the other domains, the highest distance scores in pairwise similarity analyses, and a unique “cloud” of high-amino-acid-identity sequences. These observations are congruent with the conjecture of VP1 being important in host receptor binding (47), leading to greater amino acid sequence variation in response to host evolution. Of the three new viruses, BC-3 had the greatest coverage and the fewest low-frequency SNVs in the metagenomic data, consistent with a recent lytic event that proliferated the dominant genotype. Detection of high-frequency SNVs for HaRNAV, JP-A, and JP-B indicates that these sequences are related to those from the originally identified viruses but that due to purifying selection, most of the mutations are synonymous. The RdRp sequences of the JP-A-like virus detected from Kenton-on-Sea exhibited many nonsynonymous mutations, indicating that distant relatives of this virus are widely distributed. Clearly, both distant and close relatives of the viruses presented here are distributed across oceans and hemispheres, implying a balance between dispersal and selection within the quasispecies cloud generated during RNA virus replication.

## MATERIALS AND METHODS

**Sample collection and preparation.** Fourteen samples were processed from thirteen locations: four along the temperate southwest coast of Canada (Jericho Pier [JP], Johnstone Strait [JohnS], Pendrell Sound [PendS], and Queen Charlotte Strait [QCStrait]); one composite of 10 samples from the Canadian Arctic (Arc); two from the Nunavut coastal ocean (Nun2 and Nun3) representing the northern subpolar regions; one in the temperate/subpolar Bering Sea (BerS); one from Laguna Madre, representing the subtropical Gulf of Mexico (LagM); one from tropical coastal Peru (Peru); one from temperate southern Chile (Schil); and two from the coastal warm Indian Ocean Aghulas (Kenton-on-Sea [KoS]) and cold Southern Atlantic Benguela (Cape Town [CapeP]) currents of South Africa (Fig. 7). The samples were collected at various times (see Table S1 in the supplemental material), and JP was sampled in two

consecutive years, 2013 (JP13) and 2014 (JP14). The JP14 sample was split in two and concentrated using different methods.

Samples were collected using either Niskin bottles mounted on a CTD rosette (~150 to 200 liters) or by bucket from the surface (JP, KoS, CapeP) (~2 to 20 liters). Particulate matter was removed by pressure filtering (<17 kPa) the water through glass fiber filters (MFS GC50; nominal pore size, 1.2  $\mu\text{m}$ ) and polyvinylidene difluoride filters (Millipore GVWP; pore size, 0.22  $\mu\text{m}$ ), after which viruses were concentrated using either ultrafiltration (67) through a 30-kDa-cutoff cartridge (Amicon S1Y30; Millipore) or chemical flocculation (68) with resuspension in ascorbate-EDTA buffer. Virus concentrates were stored at 4°C in the dark until processed for sequencing.

**cDNA synthesis and metagenomic library preparation.** Viruses (70 ml) were further concentrated by ultracentrifugation (120 000  $\times g$  for 5 h at 8°C), and each pellet was resuspended overnight at 4°C in 400  $\mu\text{l}$  of supernatant liquid. Nucleic acids were extracted using the PureLink viral RNA/DNA mini kit (Thermo Fisher Scientific), and the DNA was removed using TURBO DNase (Thermo Fisher Scientific). Five replicates of cDNA were synthesized, with various concentrations of template RNA, by sequence-independent single-primer amplification (SISPA) using both random-hexamer (AD88, 5'-CCTGAATTCGG ATCCTCCNNNNN-3') and nonamer (SMA-p1, 5'-GACATGTATCCGGATGNNNNNNNN-3') primers (69, 70) and Superscript III reverse transcriptase (Thermo Fisher Scientific) per the manufacturer's recommendations.

Amplification of cDNA was carried out in four replicates using primers AD89 (5'-CCTGAATTCGGAT CCTCC-3') and SMA-P2 (5'-GACATGTATCCGGATGT-3') (69, 70) and Platinum *Taq* DNA polymerase (Thermo Fisher Scientific) per the manufacturer's protocol, with the addition of 5% (vol/vol) dimethyl sulfoxide (DMSO) and various concentrations of  $\text{MgCl}_2$ . All replicates were pooled, cleaned using the Genomic DNA Clean and Concentrator kit (Zymo), and sheared to approximately 300 bp using an S220 focused-ultrasonicator (Covaris).

Size selection (>200 bp) was done using Agencourt AMPure XP beads (Beckman Coulter), and libraries were prepared using NEXTFLEX-96 DNA barcodes (Bioo Scientific) and the NxSeq DNA sample prep kit (Lucigen) per the manufacturer's protocol. Pooled libraries were sequenced on an Illumina HiSeq 2000 (2  $\times$  100 bp) at the Biodiversity Research Centre, University of British Columbia, Vancouver, Canada.

**Assembly of metagenomic reads and genomic analysis.** Primer sequences were removed and quality trimmed (PHRED score of 30) using Trimmomatic (71). Paired reads were merged with PEAR (72) and combined with the reads that had no pairs, to produce the final data sets. Each metagenome was assembled separately using the default settings of the *de novo* assembly algorithm in the CLC Genomics Workbench v7.5 (CLCBio). The contigs and unassembled reads for each metagenome were combined and reassembled using the same parameters.

Open reading frames (ORFs) were identified using the CLC Genomics Workbench v7.5. ORFs smaller than 350 bp were not included in further analyses. The ORFs were compared to the Conserved Domain Database (CDD) (73) (October 2016) using BLASTp and the Pfam database (74) (December 2015) using HMMER v3.1b2 (75). Contig size, ORF architecture, and domain identity were used to identify nearly complete genomes. All predicted untranslated regions (UTRs) of assembled genomes were analyzed using IRESite (76) to identify potential internal ribosomal entry sites (IRES).

RdRp alignments were generated using MUSCLE v3.8.425 with the default parameters (77) and then manually refined with AliView version 1.17.1 (78). Amino acid model selection was conducted using the Smart Model Selection in PhyML (79). The maximum likelihood trees were constructed using the aLRT SH-like approach in PhyML 3.0 (80), the RtREV+G+I+F amino acid model with 100 bootstrap replicates. The tree was edited in iTOL v3 (81). Sequences used are deposited under accession numbers [NP944776](#), [BAE47143](#), [YP001429583](#), [AGZ83339](#), [AFM44930](#), [AHA44480](#), [YP009047193](#), [BAG30951](#), [BAE79742](#), [YP002647032](#), [YP009230124](#), [YP009230118](#), [YP009230120](#), [YP001429581](#), [YP009111336](#), [AYD68773](#), [AYD68775](#), [AYD68777](#), and [BAK40203](#).

**Metagenomic read recruitment.** Reads were recruited from viral metagenomic data sets (Fig. 7; Table S1), either generated in this study or previously published (8, 36, 57, 82, 83), to the three newly assembled RNA genomes ([MG584187](#), [MG584188](#), and [MG584189](#)) and JP-A ([NC\\_009757](#)), JP-B ([NC\\_009758](#)), and HaRNAV ([NC\\_005281](#)) using tBLASTx (E value,  $10^{-10}$ ). Conflicts were resolved based on the highest bit score and lowest E value. Translated reads were mapped onto the genomes based on percent amino acid identities and visualized with a 0.5 alpha setting to visualize overlapping reads using R 3.2.2 (84).

**Pairwise similarity and selection pressure analyses.** Predicted domain nucleotide and amino acid sequences were aligned using MUSCLE (77). Alignments were trimmed, by removing 5' and 3' ends that did not align, with AliView v1.17.1 (78). Pairwise sequence similarities were calculated and neighbor-joining trees were constructed using the CLC Genomics Workbench v7.5 and MEGA7 (85). The number of inferred synonymous (S) and nonsynonymous (N) substitutions for each codon were estimated using joint maximum likelihood reconstructions of ancestral states under a Muse-Gaut model of codon substitution (86) and the Felsenstein model of nucleotide substitutions (87), using HyPhy (88) and MEGA7. One-way analysis of variance (ANOVA), Kruskal-Wallis rank sum tests, and the Conover-Iman test, with Bonferroni corrections, were performed using base R 3.2.2 and the conover.test R package (89).

Metagenomic reads were mapped to the six genomes using the CLC Genomics Workbench v7.5 with a stringency of identity and read overlap of 90%. Single-nucleotide variation (SNV) analysis was conducted using the quality-based variant detection tool in CLC Genomics Workbench v7.5 (minimum read count of two).

**Accession number(s).** The newly assembled viral genomes are available from GenBank under the following accession numbers: [MG584187](https://doi.org/10.1093/genbank/MG584187), [MG584188](https://doi.org/10.1093/genbank/MG584188), and [MG584189](https://doi.org/10.1093/genbank/MG584189). Sequencing data were deposited in the BioSample database under the following numbers: [SAMN08391118](https://doi.org/10.1093/biosample/SAMN08391118) to [SAMN08391132](https://doi.org/10.1093/biosample/SAMN08391132).

## SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/mSphereDirect.00157-19>.

**FIG S1**, EPS file, 0.1 MB.

**FIG S2**, EPS file, 0.1 MB.

**FIG S3**, EPS file, 0.9 MB.

**FIG S4**, EPS file, 0.1 MB.

**FIG S5**, EPS file, 0.1 MB.

**FIG S6**, EPS file, 0.2 MB.

**FIG S7**, EPS file, 0.2 MB.

**FIG S8**, EPS file, 0.4 MB.

**TABLE S1**, DOCX file, 0.02 MB.

**TABLE S2**, DOCX file, 0.01 MB.

## ACKNOWLEDGMENTS

We thank Adrian Jones (University of British Columbia, Canada), who provided advice on the statistical analysis, and H el ene Sanfa on (University of British Columbia and Agriculture and Agri-Food Canada), who provided feedback on the manuscript. We thank Amy M. Chan, Cindy Short, Steven M. Short, Caroline Ch enard, and Martin H. Villet for enabling the work through advice, logistics, and sample collection and the crews of the *CCGS Vector*, the *CCGS Louis St. Laurent*, and the *ARSV Laurence M. Gould*. We thank Valerian V. Dolja and Yuji Tomaru for critically reviewing the manuscript.

This project was supported by funding to CAS from Discovery (RGPIN-2015-05896) and Ship-Time Grants from the Natural Science and Engineering Research Council of Canada, the Canada Foundation for Innovation, and the British Columbia Knowledge Development Fund. A.S.L. was supported by a salary-based sabbatical research grant from Memorial University.

We declare no conflict of interest.

## REFERENCES

- Suttle CA. 2007. Marine viruses—major players in the global ecosystem. *Nat Rev Microbiol* 5:801–812. <https://doi.org/10.1038/nrmicro1750>.
- Rohwer F, Thurber RV. 2009. Viruses manipulate the marine environment. *Nature* 459:207–212. <https://doi.org/10.1038/nature08060>.
- Steward GF, Culley AI, Mueller JA, Wood-Charlson EM, Belcaid M, Poisson G. 2013. Are we missing half of the viruses in the ocean? *ISME J* 7:672–679. <https://doi.org/10.1038/ismej.2012.121>.
- Tai V, Lawrence JE, Lang AS, Chan AM, Culley AI, Suttle CA. 2003. Characterization of HaRNAV, a single-stranded RNA virus causing lysis of *Heterosigma akashiwo* (Raphidophyceae). *J Phycol* 39:343–352. <https://doi.org/10.1046/j.1529-8817.2003.01162.x>.
- Culley AI, Lang AS, Suttle CA. 2003. High diversity of unknown picorna-like viruses in the sea. *Nature* 424:1054–1057. <https://doi.org/10.1038/nature01886>.
- 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>.
- 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>.
- Miranda JA, Culley AI, Schvarcz CR, Steward GF. 2016. RNA viruses as major contributors to Antarctic viroplankton. *Environ Microbiol* 18:3714–3727. <https://doi.org/10.1111/1462-2920.13291>.
- Lang AS, Rise ML, Culley AI, Steward GF. 2009. RNA viruses in the sea. *FEMS Microbiol Rev* 33:295–323. <https://doi.org/10.1111/j.1574-6976.2008.00132.x>.
- Culley AI, Lang AS, Suttle CA. 2006. Metagenomic analysis of coastal RNA virus communities. *Science* 312:1795–1798. <https://doi.org/10.1126/science.1127404>.
- Short SM, Suttle CA. 2002. Sequence analysis of marine virus communities reveals that groups of related algal viruses are widely distributed in nature. *Appl Environ Microbiol* 68:1290–1296. <https://doi.org/10.1128/AEM.68.3.1290-1296.2002>.
- Labont e JM, Swan BK, Poulos B, Luo H, Koren S, Hallam SJ, Sullivan MB, Woyke T, Wommack KE, Stepanauskas R. 2015. Single-cell genomics-based analysis of virus-host interactions in marine surface bacterioplankton. *ISME J* 9:2386–2399. <https://doi.org/10.1038/ismej.2015.48>.
- 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>.
- Reche I, D'Orta G, Mladenov N, Winget DM, Suttle CA. 2018. Deposition rates of viruses and bacteria above the atmospheric boundary layer. *ISME J* 12:1154–1162. <https://doi.org/10.1038/s41396-017-0042-4>.
- Hanson CA, Marston MF, Martiny JHB. 2016. Biogeographic variation in host range phenotypes and taxonomic composition of marine cyanophage isolates. *Front Microbiol* 7:983.
- Chow C-E, Suttle CA. 2015. Biogeography of viruses in the sea. *Annu Rev Virol* 2:41–66. <https://doi.org/10.1146/annurev-virology-031413-085540>.
- Suttle CA. 2016. Environmental microbiology: viral diversity on the global stage. *Nat Microbiol* 1:16205. <https://doi.org/10.1038/nmicrobiol.2016.205>.
- King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ. 2011. *Virus taxonomy: ninth report of the International Committee on Taxonomy of Viruses*. Elsevier Academic Press, San Diego, CA.
- Nagasaki K, Tomaru Y, Katanozaka N, Shirai Y, Nishida K, Itakura S, Yamaguchi M. 2004. Isolation and characterization of a novel single-

- stranded RNA virus infecting the bloom-forming diatom *Rhizosolenia setigera*. *Appl Environ Microbiol* 70:704–711. <https://doi.org/10.1128/AEM.70.2.704-711.2004>.
20. Shirai Y, Tomaru Y, Takao Y, Suzuki H, Nagumo T, Nagasaki K. 2008. Isolation and characterization of a single-stranded RNA virus infecting the marine planktonic diatom *Chaetoceros tenuissimus* Meunier. *Appl Environ Microbiol* 74:4022–4027. <https://doi.org/10.1128/AEM.00509-08>.
  21. Nagasaki K, Shirai Y, Takao Y, Mizumoto H, Nishida K, Tomaru Y. 2005. Comparison of genome sequences of single-stranded RNA viruses infecting the bivalve-killing dinoflagellate *Heterocapsa circularisquama*. *Appl Environ Microbiol* 71:8888–8894. <https://doi.org/10.1128/AEM.71.12.8888-8894.2005>.
  22. Lawrence JE, Suttle CA. 2004. Effect of viral infection on sinking rates of *Heterosigma akashiwo* and its implications for bloom termination. *Aquat Microb Ecol* 37:1–7. <https://doi.org/10.3354/ame037001>.
  23. Culley AI, Lang AS, Suttle CA. 2007. The complete genomes of three viruses assembled from shotgun libraries of marine RNA virus communities. *Virology* 369:4–6. <https://doi.org/10.1016/j.virol.2007.04.011>.
  24. Weitz J, Hartman H, Levin S. 2005. Coevolutionary arms races between bacteria and bacteriophages. *Proc Natl Acad Sci U S A* 102:9535–9540. <https://doi.org/10.1073/pnas.0504062102>.
  25. Milo R, Jorgensen P, Moran U, Weber G, Springer M. 2010. BioNumbers—the database of key numbers in molecular and cell biology. *Nucleic Acids Res* 38:D750–D753. <https://doi.org/10.1093/nar/gkp889>.
  26. Holmes EC. 2009. The evolution and emergence of RNA viruses. Oxford University Press, New York, NY.
  27. 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>.
  28. Holmes EC. 2010. The RNA virus quasispecies: fact or fiction? *J Mol Biol* 400:271–273. <https://doi.org/10.1016/j.jmb.2010.05.032>.
  29. Ojosnegros S, Beerenwinkel N, Antal T, Nowak MA, Escarmis C, Domingo E. 2010. Competition-colonization dynamics in an RNA virus. *Proc Natl Acad Sci U S A* 107:2108–2112. <https://doi.org/10.1073/pnas.0909787107>.
  30. Chare ER, Holmes EC. 2004. Selection pressures in the capsid genes of plant RNA viruses reflect mode of transmission. *J Gen Virol* 85:3149–3157. <https://doi.org/10.1099/vir.0.80134-0>.
  31. Domingo E, Holland JJ. 1997. RNA virus mutations and fitness for survival. *Annu Rev Microbiol* 51:151–178. <https://doi.org/10.1146/annurev.micro.51.1.151>.
  32. Moya A, Holmes EC, González-Candelas F. 2004. The population genetics and evolutionary epidemiology of RNA viruses. *Nat Rev Microbiol* 2:279–288. <https://doi.org/10.1038/nrmicro863>.
  33. Lauring AS, Andino R. 2010. Quasispecies theory and the behavior of RNA viruses. *PLoS Pathog* 6:e1001005. <https://doi.org/10.1371/journal.ppat.1001005>.
  34. Elena SF, Sanjuán R. 2007. Virus evolution: insights from an experimental approach. *Annu Rev Ecol Syst* 38:27–52. <https://doi.org/10.1146/annurev.ecolsys.38.091206.095637>.
  35. Schneider WL, Roossinck MJ. 2001. Genetic diversity in RNA virus quasispecies is controlled by host-virus interactions. *J Virol* 75:6566–6571. <https://doi.org/10.1128/JVI.75.14.6566-6571.2001>.
  36. López-Bueno A, Rastrojo A, Peiró R, Arenas M, Alcami A. 2015. Ecological connectivity shapes quasispecies structure of RNA viruses in an Antarctic lake. *Mol Ecol* 24:4812–4825. <https://doi.org/10.1111/mec.13321>.
  37. Graur D, Li W-H. 2000. Fundamentals of molecular evolution. Sinauer Associates, Inc., Sunderland, MA.
  38. Duffy S, Shackleton LA, Holmes EC. 2008. Rates of evolutionary change in viruses: patterns and determinants. *Nat Rev Genet* 9:267–276. <https://doi.org/10.1038/nrg2323>.
  39. Wolf YI, Viboud C, Holmes EC, Koonin EV, Lipman DJ. 2006. Long intervals of stasis punctuated by bursts of positive selection in the seasonal evolution of influenza A virus. *Biol Direct* 1:34. <https://doi.org/10.1186/1745-6150-1-34>.
  40. Greenspan G, Geiger D, Gotch F, Bower M, Patterson S, Nelson M, Gazzard B, Stebbing J. 2004. Recombination does not occur in newly identified diverged oceanic picornaviruses. *J Mol Evol* 58:359–360. <https://doi.org/10.1007/s00239-003-2574-7>.
  41. Lang AS, Suttle CA. 2008. Marnaviruses, p 280–285. In Mahy BWJ, van Regenmortel M (ed), *Encyclopedia of virology*. Elsevier, Oxford, United Kingdom.
  42. Wolf YI, Kazlauskas D, Iranzo J, Lucía-Sanz A, Kuhn JH, Krupovic M, Dolja VV, Koonin EV. 2018. Origins and evolution of the global RNA virome. *mBio* 9:e02329-18. <https://doi.org/10.1128/mBio.02329-18>.
  43. Koonin EV, Dolja VV, Krupovic M. 2015. Origins and evolution of viruses of eukaryotes: the ultimate modularity. *Virology* 479–480:2–25. <https://doi.org/10.1016/j.virol.2015.02.039>.
  44. Lang AS, Culley AI, Suttle CA. 2004. Genome sequence and characterization of a virus (HaRNAV) related to picorna-like viruses that infects the marine toxic bloom-forming alga *Heterosigma akashiwo*. *Virology* 320:206–217. <https://doi.org/10.1016/j.virol.2003.10.015>.
  45. Culley AI, Lang AS, Suttle CA. 2011. Taxonomic structure of the family Marnaviridae, p 850–854. In King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ (ed), *Virus taxonomy: ninth report of the International Committee on Taxonomy of Viruses*. Elsevier, Oxford, United Kingdom.
  46. Horiguchi T. 2017. Raphidophyceae (Raphidophyta), p 305–330. In Archibald JM, Simpson AGB, Slamovits CH (ed), *Handbook of the protists*. Springer International Publishing, Cham, Switzerland.
  47. Rossmann MG. 1994. Viral cell recognition and entry. *Protein Sci* 3:1712–1725. <https://doi.org/10.1002/pro.5560031010>.
  48. Laitinen OH, Svedin E, Kapell S, Nurminen A, Hytönen VP, Flodström-Tullberg M. 2016. Enteroviral proteases: structure, host interactions and pathogenicity. *Rev Med Virol* 26:251–267. <https://doi.org/10.1002/rmv.1883>.
  49. Jagdeo JM, Dufour A, Klein T, Solis N, Kleinfeld O, Kizhakkedathu J, Luo H, Overall CM, Jan E. 2018. N-terminomics TAILS identifies host cell substrates of poliovirus and coxsackievirus B3 3C proteinases that modulate virus infection. *J Virol* 92:e02211-17. <https://doi.org/10.1128/JVI.02211-17>.
  50. Gustavsen JA. 2016. Drivers of viral diversity and community compositional change over spatial and temporal scales in coastal British Columbia. PhD dissertation. The University of British Columbia, Vancouver, British Columbia.
  51. Winter C, Matthews B, Suttle CA. 2013. Effects of environmental variation and spatial distance on Bacteria, Archaea and viruses in sub-polar and arctic waters. *ISME J* 7:1507–1518. <https://doi.org/10.1038/ismej.2013.56>.
  52. Chang FJ, Anderson C, Boustead NC. 1990. First record of a *Heterosigma* (Raphidophyceae) bloom with associated mortality of cage-reared salmon in Big Glory Bay, New Zealand. *N Z J Mar Freshw Res* 24:461–469. <https://doi.org/10.1080/00288330.1990.9516437>.
  53. Hongo T. 1993. Overview on bloom dynamics and physiological ecology of *Heterosigma akashiwo*, p 33–41. In Smayda T, Shimizu Y (ed), *Toxic phytoplankton blooms in the sea*. Elsevier, Amsterdam, The Netherlands.
  54. Taylor FJ. 1993. Current problems with harmful phytoplankton blooms in British Columbia waters, p 699–703. In Smayda T, Shimizu Y (ed), *Toxic phytoplankton blooms in the sea*. Elsevier, Amsterdam, The Netherlands.
  55. Band-Schmidt CJ, Morquecho L, Hernández-Becerril DU, Reyes-Salinas A, Bravo-Sierra E. 2004. Raphidophyceans on the coasts of Mexico. *Hydrobiologia* 515:79–89. <https://doi.org/10.1023/B:HYDR.0000027320.00977.8b>.
  56. Band-Schmidt CJ, Martínez-López A, Bustillos-Guzmán JJ, Carreón-Palau L, Morquecho L, Olguín-Monroy NO, Zenteno-Savín T, Mendoza-Flores A, González-Acosta B, Hernández-Sandoval FH, Tomas C. 2012. Morphology, biochemistry, and growth of raphidophyte strains from the Gulf of California. *Hydrobiologia* 693:81–97. <https://doi.org/10.1007/s10750-012-1088-y>.
  57. 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>.
  58. Simmonds P. 2006. Recombination and selection in the evolution of picornaviruses and other mammalian positive-stranded RNA viruses. *J Virol* 80:11124–11140. <https://doi.org/10.1128/JVI.01076-06>.
  59. Holmes EC. 2003. Patterns of intra- and interhost nonsynonymous variation reveal strong purifying selection in dengue virus. *J Virol* 77:11296–11298. <https://doi.org/10.1128/JVI.77.20.11296-11298.2003>.
  60. Edwards CTT, Holmes EC, Pybus OG, Wilson DJ, Viscidi RP, Abrams EJ, Phillips RE, Drummond AJ. 2006. Evolution of the human immunodeficiency virus envelope gene is dominated by purifying selection. *Genetics* 174:1441–1453. <https://doi.org/10.1534/genetics.105.052019>.
  61. Pybus OG, Rambaut A, Belshaw R, Freckleton RP, Drummond AJ, Holmes EC. 2007. Phylogenetic evidence for deleterious mutation load in RNA viruses and its contribution to viral evolution. *Mol Biol Evol* 24:845–852. <https://doi.org/10.1093/molbev/msm001>.
  62. Rossmann MG, Johnson JE. 1989. Icosahedral RNA virus structure. *Annu Rev Biochem* 58:533–573. <https://doi.org/10.1146/annurev.bi.58.070189.002533>.
  63. Rossmann MG, He Y, Kuhn RJ. 2002. Picornavirus-receptor interactions.

- Trends Microbiol 10:324–331. [https://doi.org/10.1016/S0966-842X\(02\)02383-1](https://doi.org/10.1016/S0966-842X(02)02383-1).
64. Escarmis C, Dávila M, Charpentier N, Bracho A, Moya A, Domingo E. 1996. Genetic lesions associated with Muller's Ratchet in an RNA virus. *J Mol Biol* 264:255–267. <https://doi.org/10.1006/jmbi.1996.0639>.
  65. Hoenen T, Safronetz D, Groseth A, Wollenberg KR, Koita OA, Diarra B, Fall IS, Haidara FC, Diallo F, Sanogo M, Sarro YS, Kone A, Togo AC, Traore A, Kodio M, Dosseh A, Rosenke K, de Wit E, Feldmann F, Ebihara H, Munster VJ, Zoon KC, Feldmann H, Sow S. 2015. Mutation rate and genotype variation of Ebola virus from Mali case sequences. *Science* 348:117–119. <https://doi.org/10.1126/science.aaa5646>.
  66. Bourlière M, Barberin JM, Rotily M, Guagliardo V, Portal I, Lecomte L, Benali S, Boustière C, Perrier H, Jullien M, Lambot G, Loyer R, LeBars O, Daniel R, Khiri H, Halfon P. 2002. Epidemiological changes in hepatitis C virus genotypes in France: evidence in intravenous drug users. *J Viral Hepat* 9:62–70. <https://doi.org/10.1046/j.1365-2893.2002.00319.x>.
  67. 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.
  68. John SG, Mendez CB, Deng L, Poulos B, Kauffman AK, Kern S, Brum J, Polz MF, Boyle EA, Sullivan MB. 2011. A simple and efficient method for concentration of ocean viruses by chemical flocculation. *Environ Microbiol Rep* 3:195–202. <https://doi.org/10.1111/j.1758-2229.2010.00208.x>.
  69. Márquez LM, Redman RS, Rodriguez RJ, Roossinck MJ. 2007. A virus in a fungus in a plant: three-way symbiosis required for thermal tolerance. *Science* 315:513–515. <https://doi.org/10.1126/science.1136237>.
  70. Pan X, Durrett RE, Zhu H, Tanaka Y, Li Y, Zi X, Marjani SL, Euskirchen G, Ma C, Lamotte RH, Park IH, Snyder MP, Mason CE, Weissman SM. 2013. Two methods for full-length RNA sequencing for low quantities of cells and single cells. *Proc Natl Acad Sci U S A* 110:594–599. <https://doi.org/10.1073/pnas.1217322109>.
  71. Bolger AM, Lohse M, Usadel B. 2014. Genome analysis Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
  72. Zhang J, Kobert K, Flouri T, Stamatakis A. 2014. Genome analysis PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* 30:614–620. <https://doi.org/10.1093/bioinformatics/btt593>.
  73. Marchler-Bauer A, Zheng C, Chitsaz F, Derbyshire MK, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Lu S, Marchler GH, Song JS, Thanki N, Yamashita RA, Zhang D, Bryant SH. 2013. CDD: conserved domains and protein three-dimensional structure. *Nucleic Acids Res* 41:D348–D352.
  74. Finn RD, Coghill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A. 2016. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* 44:279–285. <https://doi.org/10.1093/nar/gkv1344>.
  75. Finn RD, Clements J, Arndt W, Miller BL, Wheeler TJ, Schreiber F, Bateman A, Eddy SR. 2015. HMMER web server: 2015 update. *Nucleic Acids Res* 43:30–38. <https://doi.org/10.1093/nar/gkv397>.
  76. Mokrejš M, Mašek T, Vopálenký V, Hlubuček P, Delbos P, Pospíšek M. 2010. IRESite—a tool for the examination of viral and cellular internal ribosome entry sites. *Nucleic Acids Res* 38:D131–D136. <https://doi.org/10.1093/nar/gkp981>.
  77. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797. <https://doi.org/10.1093/nar/gkh340>.
  78. 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>.
  79. Lefort V, Longueville J-E, Gascuel O. 2017. SMS: smart model selection in PhyML. *Mol Biol Evol* 34:2422–2424. <https://doi.org/10.1093/molbev/msx149>.
  80. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 59:307–321. <https://doi.org/10.1093/sysbio/syq010>.
  81. Letunic I, Bork P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* 44:W242–W245. <https://doi.org/10.1093/nar/gkw290>.
  82. Djikeng A, Kuzmickas R, Anderson NG, Spiro DJ. 2009. Metagenomic analysis of RNA viruses in a fresh water lake. *PLoS One* 4:e7264. <https://doi.org/10.1371/journal.pone.0007264>.
  83. Rosario K, Nilsson C, Lim YW, Ruan Y, Breitbart M. 2009. Metagenomic analysis of viruses in reclaimed water. *Environ Microbiol* 11:2806–2820. <https://doi.org/10.1111/j.1462-2920.2009.01964.x>.
  84. R Core Team. 2015. R: a language and environment for statistical computing. Foundation for Statistical Computing, Vienna, Austria. <https://www.r-project.org>.
  85. Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* 33:1870–1874. <https://doi.org/10.1093/molbev/msw054>.
  86. Muse SV, Gaut BS. 1994. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. *Mol Biol Evol* 11:715–724. <https://doi.org/10.1093/oxfordjournals.molbev.a040152>.
  87. Felsenstein J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol* 17:368–376. <https://doi.org/10.1007/BF01734359>.
  88. Pond SLK, Frost SDW, Muse SV. 2005. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21:676–679. <https://doi.org/10.1093/bioinformatics/bti079>.
  89. Dinno A. 2017. conover.test: Conover-Iman test of multiple comparisons using rank sums. <http://cran.r-project.org/package=conover.test>.