

Previously unknown evolutionary groups dominate the ssDNA gokushoviruses in oxic and anoxic waters of a coastal marine environment

Jessica M. Labonté^{1†}, Steven J. Hallam^{1,2,3} and Curtis A. Suttle^{1,2,4,5*}

¹ Department of Microbiology and Immunology, University of British Columbia, Vancouver, BC, Canada, ² Canadian Institute for Advanced Research, Toronto, ON, Canada, ³ Graduate Program in Bioinformatics, University of British Columbia, Vancouver, BC, Canada, ⁴ Department of Earth, Ocean and Atmospheric Sciences, University of British Columbia, Vancouver, BC, Canada, ⁵ Department of Botany, University of British Columbia, Vancouver, BC, Canada

OPEN ACCESS

Edited by:

Katrine L. Whiteson,
University of California, Irvine, USA

Reviewed by:

Andrew David Millard,
University of Warwick, UK
Karyna Rosario,
University of South Florida, USA

*Correspondence:

Curtis A. Suttle,
Department of Earth, Ocean
and Atmospheric Sciences, University
of British Columbia, 2178-2207 Main
Mall, Vancouver, BC V6T 1Z4,
Canada
suttle@science.ubc.ca

†Present address:

Jessica M. Labonté,
Bigelow Laboratory for Ocean
Sciences, East Boothbay, ME, USA

Specialty section:

This article was submitted to Virology,
a section of the journal *Frontiers in
Microbiology*

Received: 27 October 2014

Accepted: 29 March 2015

Published: 22 April 2015

Citation:

Labonté JM, Hallam SJ and Suttle CA
(2015) Previously unknown
evolutionary groups dominate
the ssDNA gokushoviruses in oxic
and anoxic waters of a coastal marine
environment.
Front. Microbiol. 6:315.
doi: 10.3389/fmicb.2015.00315

Metagenomic studies have revealed that ssDNA phages from the family *Microviridae* subfamily *Gokushovirinae* are widespread in aquatic ecosystems. It is hypothesized that gokushoviruses occupy specialized niches, resulting in differences among genotypes traversing water column gradients. Here, we use degenerate primers that amplify a fragment of the gene encoding the major capsid protein to examine the diversity of gokushoviruses in Saanich Inlet (SI), a seasonally anoxic fjord on the coast of Vancouver Island, BC, Canada. Amplicon sequencing of samples from the mixed oxic surface (10 m) and deeper anoxic (200 m) layers indicated a diverse assemblage of gokushoviruses, with greater richness at 10 m than 200 m. A comparison of amplicon sequences with sequences selected on the basis of RFLP patterns from eight surface samples collected over a 1-year period revealed that gokushovirus diversity was higher in spring and summer during stratification and lower in fall and winter after deep-water renewal, consistent with seasonal variability within gokushovirus populations. Our results provide persuasive evidence that, while specific gokushovirus genotypes may have a narrow host range, hosts for gokushoviruses in SI consist of a wide range of bacterial taxa. Indeed, phylogenetic analysis of clustered amplicons revealed at least five new phylogenetic groups of previously unknown sequences, with the most abundant group associated with viruses infecting SUP05, a ubiquitous and abundant member of marine oxygen minimum zones. Relatives of SUP05 dominate the anoxic SI waters where they drive coupled carbon, nitrogen, and sulfur transformations along the redoxline; thus, gokushoviruses are likely important mortality agents of these bacteria with concomitant influences on biogeochemical cycling in marine oxygen minimum zones.

Keywords: *Gokushovirinae*, host range, viral diversity, SUP05, oxygen minimum zones, PCR sequencing

Introduction

Bacteriophages belonging to the *Microviridae* family consist of a ~30 nm icosahedral capsid containing a positive-sense ssDNA molecule of 4.5–6.1 kb (King et al., 2012). Replication requires a minimum of two coat proteins (VP1 and VP2), a scaffolding protein (VP3), a replication protein (VP4), and a DNA packaging protein (VP5). Based on the phylogeny of the major capsid protein (MCP or VP1) from isolates, the *Microviridae* family is divided into two groups (Brentlinger et al., 2002); the *Microvirus* genus contains phages like phiX174 and G4 that infect *Escherichia coli* (Godson et al., 1978), while the *Gokushovirinae* subfamily includes those infecting parasitic bacteria such as *Chlamydia* [Chp1 (Storey et al., 1989), Chp2 (Liu et al., 2000; Everson et al., 2002), Chp3 (Garner et al., 2004)], *Bdellovibrio* [(phiMH2K; Brentlinger et al., 2002)], and *Spiroplasma* [(SpV4; Chipman et al., 1998)]. While it is commonly thought that *Microviridae* phages are strictly lytic (Liu et al., 2000; Garner et al., 2004; Salim et al., 2008), an *in silico* study found sequences with similar genome organization to gokushoviruses associated with *Bacteroidetes* from the human gut and mouth, which suggests that these phages can be temperate (Krupovic and Forterre, 2011). The temperate *Microviridae* phages are phylogenetically distinct from gokushoviruses, and have been assigned to the *Alpavirinae*, a proposed new sub-family within the *Microviridae* (Krupovic and Forterre, 2011). Recently, 81 microvirus genomes (including 42 gokushoviruses) were assembled from various environmental metagenomic data, identifying a new group, the *Pichovirinae*, which harbored a different genome organization of the conserved genes, indicating that microviruses display great diversity and may play an important role in many ecosystems (Roux et al., 2012b).

Marine gokushoviruses were first revealed in viral metagenomic data from the Strait of Georgia (SOG), Gulf of Mexico (GOM), and Sargasso Sea (SAR; Angly et al., 2006), and are among the most commonly recovered sequences from ssDNA phages in marine metagenomic data (Rosario and Breitbart, 2011). They were particularly abundant in the SAR, where 6% of the sequences were similar to the phage Chp1 that infects *Chlamydia psittaci* (Angly et al., 2006). This abundance of ssDNA sequences allowed for the assembly of two environmental *Gokushovirinae* genomes, with the help of PCR amplification (Tucker et al., 2011). A survey in the Atlantic Ocean showed a different depth distribution of these two genomes, consistent with alternative host-infection patterns. Sequences belonging to gokushoviruses have also been found in marine (Labonté and Suttle, 2013b; McDaniel et al., 2014) and fresh waters (López-Bueno et al., 2009; Roux et al., 2012a), stromatolites (Desnues et al., 2008), confined aquifers (Smith et al., 2013), and pelagic sediments (Yoshida et al., 2013). Based on these observations, degenerate primers designed to amplify fragments of the genes encoding the replication initiator (ORF4 or Rep; Tucker et al., 2011) and MCP (Labonté and Suttle, 2013a; Hopkins et al., 2014) were used to examine the distribution and diversity of gokushoviruses in marine ecosystems. Phylogenetic analyses of the amplified fragments revealed that ssDNA phages have different geographic distributions (Labonté and Suttle, 2013a), and

that the genetic distance of gokushovirus sequences increased with geographic distance (Tucker et al., 2011). Most of the hosts of environmental gokushoviruses are unknown, but it is hypothesized that they occupy specialized niches, and that specific gokushovirus genotypes have limited geographic range (Angly et al., 2006; Tucker et al., 2011; Labonté and Suttle, 2013a).

Saanich Inlet (SI) is a steep-sided fjord with restricted circulation due to a shallow glacial sill located at the entrance. During spring and summer, high primary productivity in surface waters combined with limited basin circulation contribute to the formation of deep-water anoxia (Anderson, 1973). The anoxic zone is characterized by accumulation of CH₄, NH₃, and H₂S (Anderson, 1973; Lilley et al., 1982; Ward et al., 1989). Typically, in late summer oxygenated nutrient-rich water from Haro Strait (connecting SI to the SOG) cascades over the sill, mixing the oxic, and anoxic waters from top to bottom (Anderson, 1973). Recently, single-cell amplified genomic data (SAGs) from uncultured SUP05 bacteria from marine oxygen minimum zones revealed identical *Microviridae* sequences in 8 of 127 SAGs, suggesting a recent infection event (Roux et al., 2014).

Here, rather than looking at seasonal changes (Labonté and Suttle, 2013a), the diversity of gokushoviruses was examined in the mixed oxic surface (10 m) and deeper anoxic (200 m) layers of SI to better understanding their dynamics and roles in environments with contrasting levels of oxygen. We used degenerate primers for the MCP in combination with amplicon sequencing using 454 technology to reveal a diverse assemblage of gokushoviruses with greater richness at 10 m than 200 m. The results provide persuasive evidence that gokushoviruses likely infect a wide range of hosts, and may be important mortality agents of the SUP05 clade of gamma proteobacteria, an important taxonomic group involved in carbon, nitrogen, and sulfur cycling in marine oxygen minimum zones.

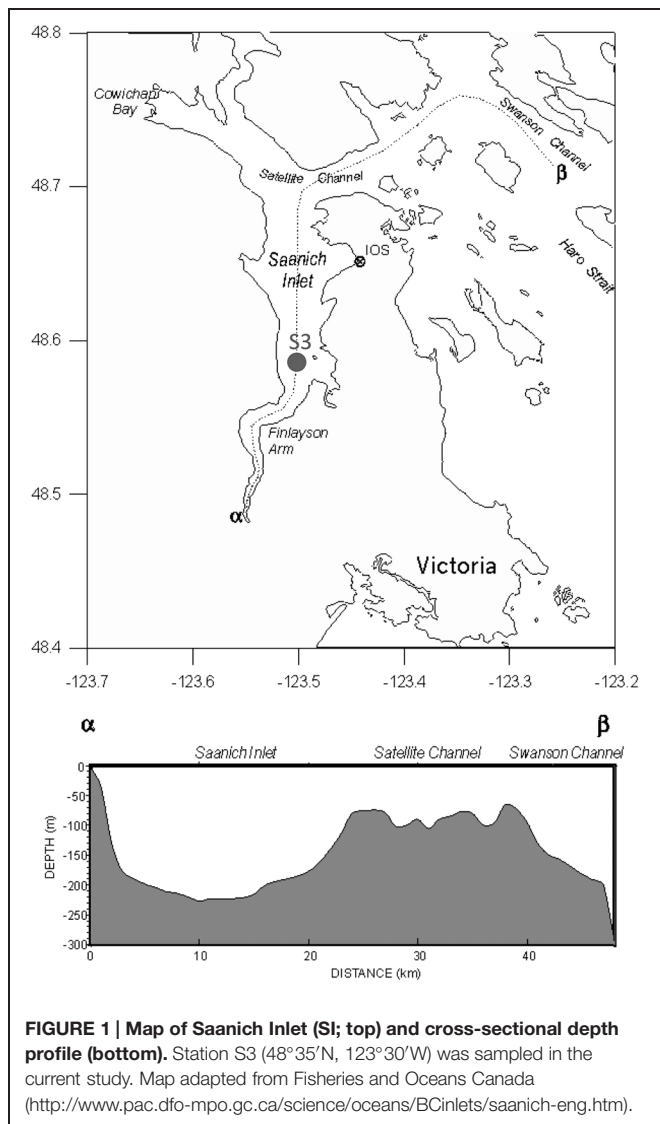
Materials and Methods

Preparation of Marine Samples

On a monthly basis, ~20 L of water from Station S3 in SI (**Figure 1**) were filtered to remove cells using 0.22- μ m pore-size SterivexTM filter units (Millipore). The viruses were concentrated from the filtrate by tangential flow filtration using a TFF 30-kDa cartridge (Millipore) to a final volume of ~250 mL, and stored at 4°C until used following the procedure outlined in (Suttle et al., 1991). Amplicons were sequenced from two composite samples of virus concentrates comprised of samples collected from 10 m and 200 m, respectively, during April 2007, and February, March, April, June, August, and November (200 m), or December (10 m) 2008.

ssDNA Purification, PCR Amplification, and Sequencing

For each virus concentrate, viral ssDNA was extracted using a QIAprep Spin M13 kit (Qiagen), according to the manufacturer's protocol. To create double-stranded DNA, 5 μ L of the ssDNA



preparation was subjected to multiple displacement amplification (MDA; Repli-g Mini kit), and purified using a QIAamp DNA Mini kit. Denaturation of dsDNA was limited during MDA by adding the stop solution N1 immediately after the denaturation solution D1. The purified MDA DNA was resuspended in 100 μ L of TE, and 10 μ L was used as template in each PCR reaction mixture consisting of Taq DNA polymerase assay buffer [20 mM Tris-HCl (pH 8.4), 50 mM KCl], 1.5 mM MgCl₂, 125 μ M of each deoxyribonucleoside triphosphate, 1 μ M of each primer amplifying a fragment of the MCP gene [MicroVP1-F1 (CGN GCN TAY AAY TTR ATH), VP1-F2 (AGN GCN TAY AAY TTR CTN), MicroVP1-R1 (NCG YTC YTG RTA NCC RAA), and Micro-VP1-R2 (NCT YTC YTG RTA NCC RAA)] and 2.5 U of Platinum[®] Taq DNA polymerase (Invitrogen). Forward and reverse primer pairs were used to reduce the degeneracy within a single primer. Negative controls contained all reagents except DNA template. The samples were denatured at 94°C for 3 min, followed by 35 cycles of denaturation at 94°C

for 30 s, annealing at 50°C for 30 s, and elongation at 72°C for 50 s, with a final elongation step of 72°C for 5 min. The expected PCR product was \sim 800 bp in length. PCR amplicons were purified with a MinElute PCR purification kit (Qiagen), pooled into mixes from the 10 or 200 m depth and concentrated using a Millipore YM-30 Microcon centrifugal filter to a final volume of \sim 50 μ L; a total of 500 ng of DNA amplicons from each pooled 10 and 200 m sample were sent for pyrosequencing using Roche 454 FLX instrumentation with Titanium chemistry at the Broad Institute at the Massachusetts Institute of Technology.

Sequence Binning and Clustering

The sequences were screened for quality and length. Reads were removed if they contained one or more ambiguous bases (Ns), were shorter than 200 nucleotides, or did not match the priming site at the proximal end. Reads were binned based on the primer sequence, which was subsequently trimmed. Sequences arising from both forward primers (F) and both reverse primers (R) were combined resulting in F and R bins. Sequence errors can occur throughout the workflow, including an error rate of 1 per 10^6 – 10^7 bp (Dean et al., 2001) for WGA, and homopolymers, insertions and deletions of about 0.1% per base for pyrosequencing (Margulies et al., 2005; Huse et al., 2010; Quince et al., 2011); however, the impact of these errors were minimized by clustering the reads into operational taxonomic units (OTUs) at 95% identity using CD-hit (Li and Godzik, 2006). Clustering at 95% also recruited most singletons into an OTU and allowed the data to be compared with the OTUs from a previous seasonal study of gokushoviruses in SI (Labonté and Suttle, 2013a).

Operational taxonomic units were queried in a BLAST search analysis (NCBI BLAST 2.2.2) using an *e*-value cut-off of 10^{-5} against a manually curated database derived from environmental sequences and sequences from gokushovirus isolates composed of all the *Microviridae* genomes available in GenBank (as of November 23, 2014), assembled genomes from Roux et al. (2012b), environmental MCP amplicon sequences from Labonté and Suttle (2013a) and Hopkins et al. (2014). OTUs that did not have a significant hit to sequenced gokushoviruses were removed from the phylogenetic analysis.

Diversity and Species Richness Calculations

For each primer bin, a rank-abundance distribution of phenotypes was generated and subsequently fitted to a power-law function using non-linear regression. For each primer bin, rarefaction species richness curves, and diversity indices were calculated using the Vegan Ecological Diversity package in R (R Development Core Team, 2011). Total estimated richness (S_p) was calculated following Chao's equation (Chao, 1987). The Shannon-Weaver (H') diversity index was calculated as in Hill (1973) on a subsample of 700 reads.

Phylogenetic Analyses

Nucleotide OTUs were aligned with other environmental sequences (Roux et al., 2012b; Yoshida et al., 2013; Hopkins et al., 2014) using MAFFT (Katoh et al., 2002) with the E-INS-I

parameters. We worked with the nucleotide sequences because of problems such as homopolymers associated with the 454 platform made it difficult to accurately infer the correct amino-acid sequences. The alignment of the end product (R primers) was trimmed to get the conserved regions only and phylogenetic analysis was performed with phyML (Guindon et al., 2010) under the HKY85 substitution model with an invgamma distribution with approximate likelihood ratio test (aLRT). Trees were viewed in FigTree¹.

Nucleotide Sequence Accession

Raw sequences, OTU sequences, alignments, and trees are publicly available on Dryad².

Results and Discussion

The results from our analyses showed that the oxic and anoxic waters of SI are home to diverse gokushoviruses that comprise at least five previously unknown phylogenetic groups, composed

¹<http://tree.bio.ed.ac.uk/software/figtree/>

²<http://dx.doi.org/10.5061/dryad.7gg25>

of many numerically dominant OTUs. The most abundant group is associated with viruses that infect SUP05, a group of sulfur-oxidizing bacteria that are ubiquitous and abundant players in marine oxygen minimum zones. These results and their interpretations are detailed below.

Amplicon Sequencing of the MCP From Gokushoviruses

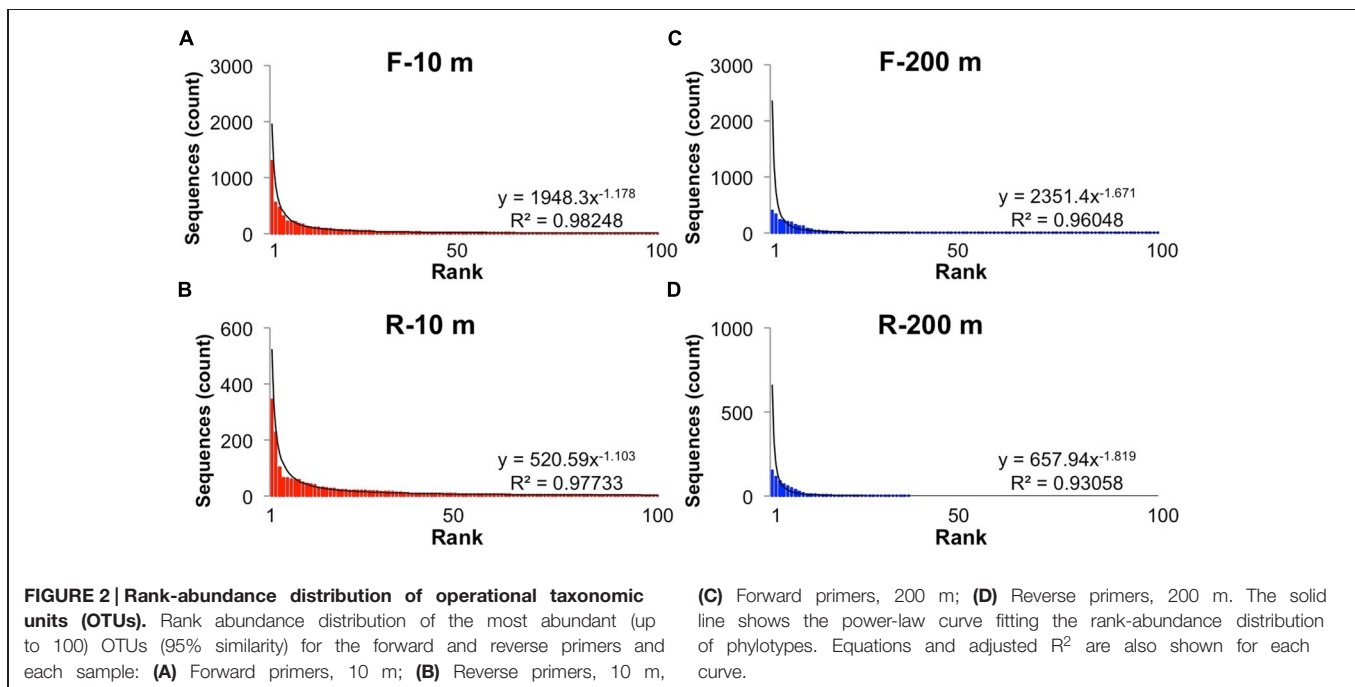
Amplicon sequences of the MCP from two pooled mixes (10 and 200 m) generated 7195 (F) and 2135 (R) good reads (no N, exact primer match, no chimeras) for the 10-m bin and 2687 (F) and 710 (R) good reads for the 200-m bin (**Table 1**). Quality controls using gel electrophoresis and DNA quantification, indicated that the yields from PCR amplification of the 200-m samples were consistently lower than for the 10-m samples (data not shown), which can explain the lower number of MCP amplicons obtained for the 200-m samples.

Reads longer than 200 bp from each bin were clustered into OTUs with more than 95% sequence similarity, and ranked to show the relative abundance of each gokushovirus taxon in our dataset (**Figure 2**). The rank-abundance plots display a similar trend for each primer bin, with a few dominant genotypes and a

TABLE 1 | Number of reads and OTUs after clustering at 95% similarity and OTUs sharing similarity with known *Microviridae* phages.

	Primer	Number of good reads	Number of OTUs	Microviridae OTUs	Most abundant % reads	Singletons % reads	Doublets % reads
10 m	F	7195	504	446	17.9	2.9	2.1
	R	2135	228	213	16.2	4.3	3
200 m	F	2687	117	80	14.8	1.6	0.8
	R	710	36	25	21.5	1.7	1.1

Clustering was performed on sequences that were >200 bp. The number of reads represents the number of reads without ambiguous base.



long tail of doubletons and singletons. The rank abundance distribution of genotypes was approximated by a power-law function with R^2 -values >0.95 . In contrast, there were fewer OTUs in the 200-m samples, and the distribution was less well described by a power-law function. Phages and their hosts usually follow a power-law rank-abundance distribution (Edwards and Rohwer, 2005; Hoffmann et al., 2007; Suttle, 2007). Possible explanations for environmental phage genotypes following a power-law distribution (Edwards and Rohwer, 2005) are that multiple viruses compete for the same hosts or that each virus is specific for one host but the hosts compete for resources. In the first scenario, the most abundant viruses are the most successful at finding and infecting their hosts; hence, every round of infection produces more of that virus. In the second situation, the host that gets more nutrients divides more rapidly resulting in more available hosts for the viruses. In this case, the most abundant viruses are the ones that infect the most abundant bacteria.

The overall sequence similarity within primer bins was compared by clustering the sequences using multiple similarity thresholds (Figure 3). Both F and R sequences were more similar to each other at 200 m than at 10 m. This was particularly true for the R sequences (3' end) for which a decrease in the similarity threshold from 95 to 90% resulted in 41.7% fewer OTUs at 200 m, indicating that many of the sequences were similar and suggesting recent infection events. The 3' end of the PCR product is more conserved, which can explain the higher similarity among OTUs in the reverse primer bins. For the other samples, the decrease in OTUs was steeper when the similarity threshold was below 80%, indicating the sequences were more diverse.

Rarefaction curves did not plateau (Figure 4), indicating that the sequencing depth was inadequate to capture the entire gokushovirus community. Based on the number and frequency of each OTU, the total richness was estimated to be between 355 and 504 OTUs at 10 m, and between 49 and 189 OTUs at 200 m

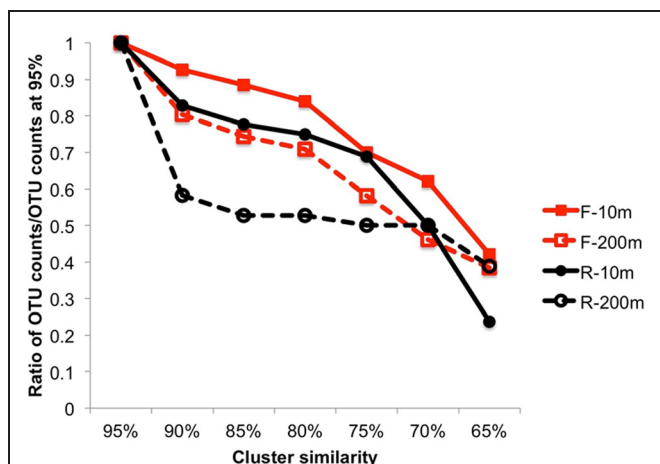


FIGURE 3 | Cluster analysis from 95 to 65% similarity showing the higher similarity within the end of the amplicon sequences and a lower similarity within the beginning of the amplicon sequences. The ratio is expressed as the number of OTUs/number of OTUs at 95%.

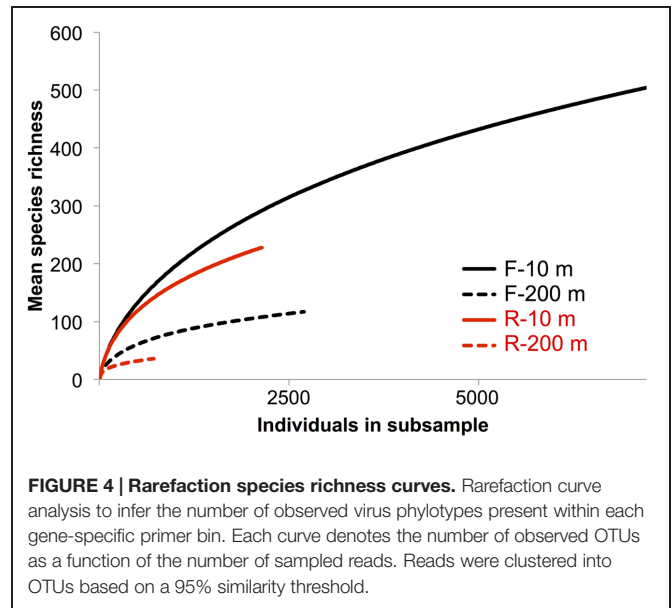


FIGURE 4 | Rarefaction species richness curves. Rarefaction curve analysis to infer the number of observed virus phylotypes present within each gene-specific primer bin. Each curve denotes the number of observed OTUs as a function of the number of sampled reads. Reads were clustered into OTUs based on a 95% similarity threshold.

(Table 2). Therefore, between 65 and 62% of total gokushovirus amplicon richness in SI (Table 2) was estimated to be captured using the F primers at 10 and 200 m, respectively. Since most reads were not singletons (e.g., only 2.9% of singletons for the F primers), the richness that was not captured likely comprised rare genotypes. Usually, environmental samples are dominated by relatively few genotypes and many more low-abundance ones that account for most of the genetic diversity (Hoffmann et al., 2007; Suttle, 2007). Moreover, abundant and rare genotypes can be temporally and spatially dynamic with a rare genotype being dominant in a different environment or when conditions change (Sogin et al., 2006; Huse et al., 2010; Gobet et al., 2012; Shade et al., 2014).

Care must be taken in interpreting diversity estimates calculated from the data. First, evenness can be influenced by the use of MDA, which can unevenly amplify the initial template (Dean et al., 2001), be biased toward small circular ssDNA molecules and form chimeras (Binga et al., 2008; Polson et al., 2011). Second, the PCR products were pooled from multiple months, which could have affected the evenness observed. For example, if one time point was dominated by a single genotype, it may be highly represented in the pooled sample, even if it was in relatively low abundance at other times. In contrast, a sample with many genotypes at similar abundance will have a lower

TABLE 2 | Observed and estimated richness and diversity indices.

	Primer	Observed richness	Estimated richness (S_{chao})	Shannon diversity index
10 m	F	504	786±54	3.93
	R	228	355±36	3.87
200 m	F	117	189±31	3.03
	R	36	49±10	2.50

The richness is expressed in numbers of OTUs defined at 95% identity.

concentration of each genotype that will be further diluted in the mix. Despite these caveats, and even though temporal variation is integrated across samples, it is reasonable to assume that the observed differences in diversity estimates between the pooled 10 m and 200 m samples is valid. Moreover, because temporal variation is integrated across dates, differences in overall richness are unlikely to be the result of among sample variation; hence, the 4–10-fold higher richness observed in the 10-m samples relative to the 200-m samples (Table 2) is likely real. Consistent with this observation, Shannon indices are also higher for the 10-m samples (4.21 and 4.05) than for the 200-m samples (3.18 and 2.54; Table 2), reinforcing observations of lower prokaryotic diversity in the anoxic versus oxic waters of SI, with a Shannon index of 0.87 vs. 1.15 for archaea and 0.39 vs. 3.15 at 215 m and 10 m, respectively, (Zaikova et al., 2010).

Seasonal Variation of Gokushoviruses in the Saanich Inlet

In a previous study Labonté and Suttle (2013a) used RFLP analysis to select MCP amplicons for sequencing from SI, SOG, and the GOM. For SI, RFLP analysis of 180 clones from PCR products spanning nine samples resolved 19 unique bands. More bands were observed in spring and summer, when the bacterial abundance is higher and the water column becomes increasingly stratified, while fewer bands were observed in the fall and winter, when bacterial abundance is lower after deep water renewal.

Amplicon 454 sequencing recovered 15 of the 19 sequences from SI associated with the RFLPs (Table 3). All four sequences associated with a specific RFLP that were not recovered in the 454 data were found only once, and consequently may have been absent from the 454 data. Since both methods used different DNA preparations, the absence of these sequences may have resulted from MDA or PCR biases. It is also possible that the 454 sequencing was not deep enough, as alluded to in the rarefaction curves that did not plateau (Figure 4). A lack of sequencing depth could also explain why most of the RFLP sequences were only found with either the F or R primer, but not both. Nonetheless, the richness recovered was much higher using 454 sequencing than by RFLP analysis.

In contrast to the SI results, none of the 13 sequences from the GOM and only two out of 12 sequences from the SOG (SOG3-31 and SOG4-29) that were associated with RFLPs, were recovered in the 454-sequencing data. The SOG sample comprised a mixture of 85 virus concentrates from the SOG and surrounding inlets, including SI. The fact that only two sequences were recovered in the 454 data suggests that gokushovirus sequences display a high degree of endemicity. In contrast, studies on the portal protein from myoviruses (Short and Suttle, 2005; Sullivan et al., 2008), and DNA polymerase B from podoviruses (Breitbart et al., 2004; Chen et al., 2009; Labonté et al., 2009; Huang et al., 2010) and phycodnaviruses (Short and Suttle, 2002) have recovered identical or nearly identical sequences from very different environments.

TABLE 3 | Recovery of RFLP Saanich sequences within the amplicon deep-sequencing database.

RFLP sequence	Apr 08	Jan 09	Mar 09	May 09	Jul 09	Aug 09	Nov 09	OTU	Frequency of OTU	Frequency (%)	Sample
SI-01	█										
SI-02			█					DMER9	1/7195	0.01	F-10 m
SI-03					█						
SI-04				█							
SI-05	█							DBSZ3	1/7195	0.01	F-10 m
SI-06	█							D5JJE	2/2135	0.09	R-10 m
SI-07	█	█	█					ENNJ4	60/2135	2.81	R-10 m
SI-07	█	█	█			█		DCFCF	1/2135	0.05	F-10 m
SI-08/SI-09				█	█			DV3J0	1/7195	0.01	F-10 m
SI-10	█		█	█				D0YTA	20/7195	0.28	F-10 m
SI-11	█							EW8A2	57/2135	2.67	R-10 m
SI-12							█	C3HEQ	118/7195	1.64	F-10 m
SI-13	120							DNX08	64/2135	3.00	R-10 m
SI-13	120							F9ET8	49/2687	1.82	F-200 m
SI-14			█					BOBYV	1/7195	0.01	F1-10 m
SI-16				█				D008I	1/7195	0.01	F1-10 m
SI-17	120										
SI-18		█						EKHXK	5/7195	0.07	F-10 m
SI-18		█						D3RZ4	6/2135	0.28	R-10 m
SI-SOG-19					█			C4GME	3/2135	0.14	R-10 m
SI-SOG-19					█			DSP87	4/7195	0.06	F-10 m

A total of 15 out of 19 sequences (shown in gray) were sequenced again with amplicon deep sequencing. The seasonal variability of each of these OTUs is represented by dark boxes in the sample months from which they were sequenced. SI, Saanich Inlet; SOG, Strait of Georgia. SI-08 and SI-09 are combined as they were >95% similar. The SI sample depth is indicated if not 10 m.

Phylogenetic Relationships Among Gokushovirus MCP Sequences from Saanich Inlet

The phylogenetic relationships among MCP sequences from SI were assessed in relation to MCP sequences from isolates and of other environmental PCR amplicons. The sequenced isolates included phages infecting the parasitic bacteria *Chlamydia* sp., *Bdellovibrio bacteriovorus*, and *Spiroplasma melliferum*; of these only MH2K that infects *B. bacteriovorus* has close marine relatives (**Figure 5**). Marine *Bdellovibrio*-and-like-organisms (BALOs) are commonly found in marine environments and parasitize *Vibrio* sp. (Martin, 2002); hence, viruses similar to MH2K could be infecting marine BALOs. In addition to isolates, phage sequences infecting uncultivated SUP05 from SI obtained using single-cell genomics were also included (Roux et al., 2014), as were MCP sequences from other studies that used metagenomic (Roux et al., 2012b; Yoshida et al., 2013) or targeted amplification approaches (Hopkins et al., 2014).

The majority of the sequences fell within supported ($\geq 90\%$ bootstrap support; ≥ 4 sequences) phylogenetic groups with four or more sequences (**Figure 5**), although sequences also fell outside these clades. Gokushovirus genomes assembled from metagenomics or single-cell genomics were representative of the SI-5, ENV-2, ENV-9, and WSB-2 clades. Many groups comprised sequences that were location specific (red boxes for the SI groups, dark gray for other environments), indicating that gokushoviruses infect endemic bacteria, and congruent with previous studies that suggest a biogeographic separation of gokushoviruses coupled to specific hosts (Labonté et al., 2009; Tucker et al., 2011). However, many groups contained sequences from multiple environments (light gray boxes on **Figure 5**), suggesting that some gokushoviruses infect widely distributed hosts. In contrast to the 43 unique sequences that were recovered from 77 RFLP patterns observed among 400 analyzed clones (Labonté and Suttle, 2013a), high-throughput sequencing allowed the discovery of five previously unknown groups of gokushoviruses.

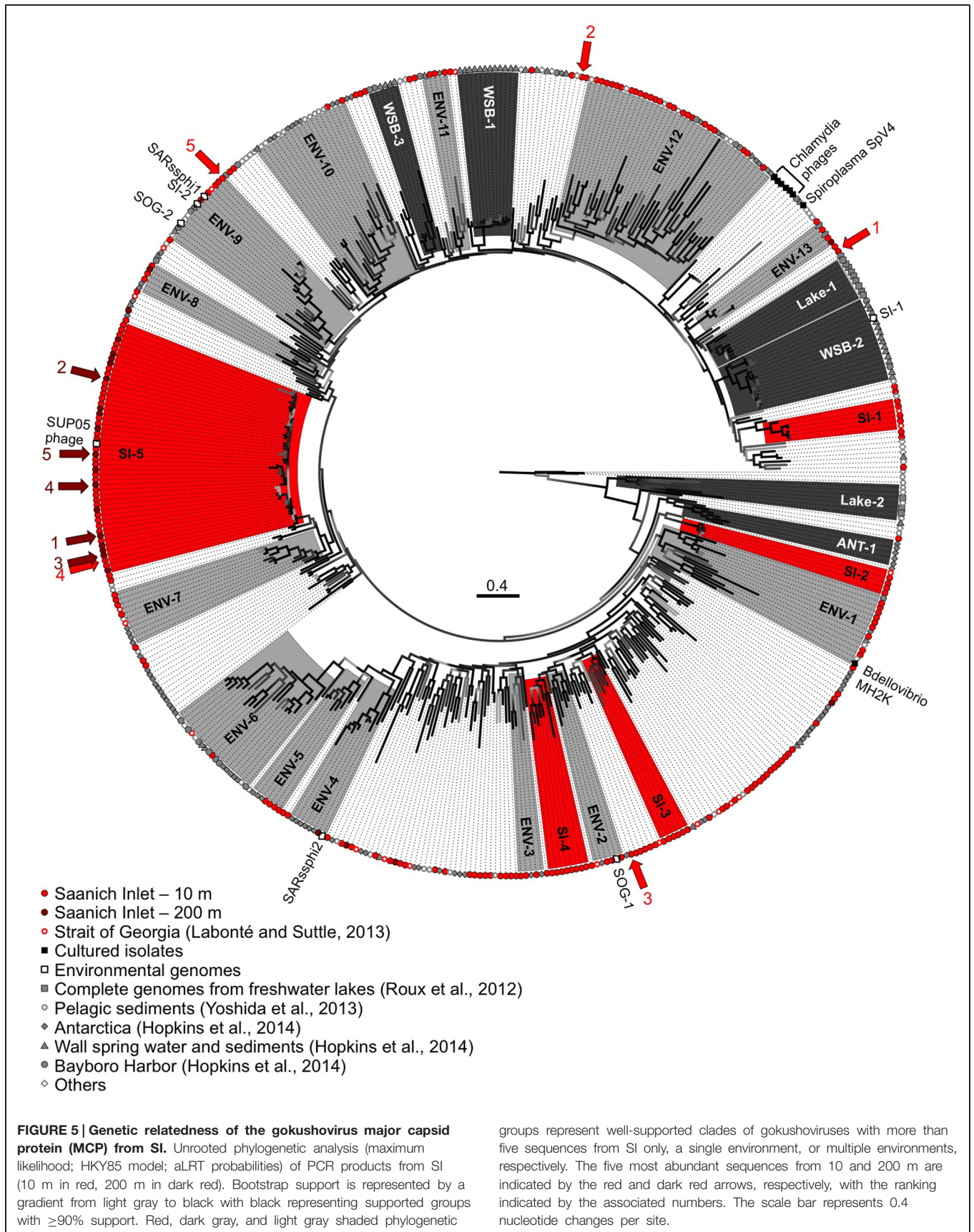
Phylogenetic analysis of the 3' end of the gene encoding the MCP (**Figure 5**) revealed that group SI-5 contained most of the sequences from 200 m, as well as some from 10 m that shared $>90\%$ nucleotide identity, but none from other locations. Although all the sequences could not be aligned with the 5' end of the sequences, members of the SI-5 clade shared $>80\%$ pairwise identity with the 5' end supporting the close phylogenetic relationship among these sequences. The five most common sequences from 200 m, and one of the five most common ones from 10 m, fell within the SI-5 clade (**Figure 5**). Based on the RFLP pattern, sequence SI-13 from an April 2007 anoxic-zone sample from 120 m belonged to the SI-5 group, and was also the fourth most abundant sequence in the 10 m deep-sequencing data. Some of the temporal changes in the gokushovirus phylogenetic relationships, such as the presence of RFLP sequence SI-13 in the anoxic zone and 10 m sample in April 2007 likely resulted from changes in the bacterial community. These results agree with a metagenomic study of four aquatic ecosystems, in which the dominant viral taxa persisted over time, while the relative abundances of rare ones constantly changed (Rodríguez-Brito et al.,

2010). It has been hypothesized that microbial and viral taxa continuously replace each other in a 'kill-the-winner' manner, maintaining stable metabolic potential and species composition (Rodríguez-Brito et al., 2010). The bacterial community of SI is dynamic and varies based on changing levels of oxygen-deficiency in the water column throughout the year (Zaikova et al., 2010). As SUP05 dominates in the anoxic zone of SI (Zaikova et al., 2010), Group SI-5 sequences are likely from viruses that infect SUP05. In contrast, viruses infecting more ephemeral taxa likely belong to phylogenetic groups with fewer representative OTUs.

The short genetic distance among sequences in Group SI-5 compared to other clades is consistent with a recent infection event. The presence of sequences from both 10 and 200 m suggests that the event occurred in the fall, during deep-water renewal. Interestingly, gokushovirus sequences associated with single-cell genomes from uncultured SUP05 bacteria, found in marine oxygen minimum zones (Roux et al., 2014), fell within the SI-5 gokushoviruses. SUP05 is the most abundant group of bacteria in the anoxic layer of SI (Zaikova et al., 2010), and it seems likely that SI-5 gokushoviruses infect, and are important agents of mortality for SUP05 bacteria and their relatives.

In contrast, the most common sequences from the 10-m sample fell within unsupported groups containing less than five genetically distinct OTUs. An abundant sequence suggests the occurrence of a recent lytic event or an abundant host. Since many sequences did not fall into any phylogenetic group (**Figure 5**), gokushoviruses likely infect a wide range of bacterial taxa, as suggested with other virus groups, where a wider genetic diversity implies a wider diversity of hosts organisms that are infected (Filée et al., 2005; Clasen and Suttle, 2009). Also, the lower richness of gokushoviruses at 200 m than at 10 m in SI (**Figure 4**; **Table 2**) parallels the differences in bacterial richness between depths (Zaikova et al., 2010), suggesting more potential hosts in the oxic than in the anoxic zone. However, the high similarity of the sequences that clustered with the SUP05 phages, and the specific geographic distribution of gokushovirus sequences, suggest that each gokushovirus has a narrow host range.

In general, a power-law distribution of viral taxa indicates an environment in which viruses infect the competitively dominant hosts (Hoffmann et al., 2007). The viral taxa at 200 m were not as well described by a power-law function as those at 10 m (**Figure 2**), likely because there were fewer sequences and the sequences were more similar to each other. Because the five most abundant sequences at 200 m were in group SI-5 and the frequency distribution of the dominant taxa did not fit a power-law function, it is likely that most sequences from 200 m were the result of a recent lytic event of cells within the SUP05 group, as suggested in Roux et al. (2014). Phenomena that could explain the presence of similar sequences at 10 and 200 m are the yearly deep-water renewal and the sinking and resuspension of viral particles. Deep-water renewal does not result in complete mixing of the water column. Rather, the oxygenated water flowing into SI is denser than the basin water; therefore, it sinks and displaces deep-basin waters upward (Anderson, 1973). An



alternative hypothesis is that viruses which sediment on particles during stratification are transported upward during renewal.

No marine gokushovirus has been isolated so far, but they likely infect a wide range of hosts throughout the water column. A 16S ribosomal RNA gene survey from SI (Zaikova et al., 2010) revealed that Bacteroidetes, δ -proteobacteria (Nitrospina), Actinobacteria (Microthrix), and Verrucomicrobia were more abundant at 10 m than at 200 m, and are potential host taxa. Prophages with a similar genome organization to gokushoviruses have been found in the genomes of bacteria from the phylum Bacteroidetes (Krupovic and Forterre, 2011), supporting the idea that these bacteria may be hosts for gokushoviruses in SI.

This study demonstrated that the genetic richness of gokushoviruses was much higher in the oxic (10 m) than anoxic (200 m) layers of SI, and that a power-law function better described the taxonomic distribution of gokushoviruses at 10 m than 200 m, reflecting the bacterial diversity through the redoxcline. Finally, the presence of very similar viruses at 10 m and 200 m is likely due to deep water renewal or potentially biomass sinking from the surface. These results suggest that gokushoviruses infect a wide range of

hosts, but that the host range of an individual genotype is narrow.

Acknowledgments

We thank members of the Suttle laboratory for collecting and processing the samples, and the Hallam laboratory for providing filtered water from Saanich Inlet that made this study possible. We also thank the reviewers for their comments which have resulted in an improved manuscript. This research was supported by the Natural Science and Engineering Research Council of Canada (NSERC) through a postgraduate scholarship (JL) and grants awarded to CS, and SH including NSERC Discovery, Canada Foundation for Innovation (CFI), and the Canadian Institute for Advanced Research (CIFAR). Sample collection was facilitated through ship-time grants from NSERC that supported sample collections in Saanich Inlet (P. D. Tortell and SH). Access to sequencing was funded by the Gordon and Betty Moore Foundation through GBMF1799 to the Broad Institute, and by NSERC and the Tula Foundation using facilities at McGill University and Genome Quebec Innovation Centre.

References

- Anderson, J. (1973). Deep water renewal in Saanich Inlet, an intermittently anoxic basin. *Estuar. Coast. Mar. Sci.* 16, 1–10. doi: 10.1016/0302-3524(73)90052-2
- Angly, F. E., Felts, B., Breitbart, M., Salamon, P., Edwards, R. A., Carlson, C., et al. (2006). The marine viromes of four oceanic regions. *PLoS Biol.* 4:e368. doi: 10.1371/journal.pbio.0040368
- Binga, E. K., Lasken, R. S., and Neufeld, J. D. (2008). Something from (almost) nothing: the impact of multiple displacement amplification on microbial ecology. *ISME J.* 2, 233–241. doi: 10.1038/ismej.2008.10
- Breitbart, M., Miyake, J. H., and Rohwer, F. (2004). Global distribution of nearly identical phage-encoded DNA sequences. *FEMS Microbiol. Lett.* 236, 249–256. doi: 10.1111/j.1574-6968.2004.tb09654.x
- Brentlinger, K. L., Hafenstein, S., Novak, C. R., Fane, B. A., Borgon, R., McKenna, R., et al. (2002). Microviridae, a family divided: isolation, characterization, and genome sequence of phiMH2K, a bacteriophage of the obligate intracellular parasitic bacterium *Bdellovibrio bacteriovorus*. *J. Bacteriol.* 184, 1089–1094. doi: 10.1128/jb.184.4.1089-1094.2002
- Chao, A. (1987). Estimating the population size for capture-recapture data with unequal catchability. *Biometrics* 43, 783–791. doi: 10.2307/2531532
- Chen, F., Wang, K., Huang, S., Cai, H., Zhao, M., Jiao, N., et al. (2009). Diverse and dynamic populations of cyanobacterial podoviruses in the Chesapeake Bay unveiled through DNA polymerase gene sequences. *Environ. Microbiol.* 11, 2884–2892. doi: 10.1111/j.1462-2920.2009.02033.x
- Chipman, P. R., Agbandje-McKenna, M., Renaudin, J., Baker, T. S., and McKenna, R. (1998). Structural analysis of the spiroplasma virus, SpV4: implications for evolutionary variation to obtain host diversity among the *Microviridae*. *Structure* 6, 135–145. doi: 10.1016/S0969-2126(98)00016-1
- Clasen, J. L., and Suttle, C. A. (2009). Identification of freshwater phycodnaviridae and their potential phytoplankton hosts, using DNA pol sequence fragments and a genetic-distance analysis. *Appl. Environ. Microbiol.* 75, 991–997. doi: 10.1128/AEM.02024-08
- Dean, F., Nelson, J., Giesler, T., and Lasken, R. (2001). Rapid amplification of plasmid and phage DNA using phi29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Res.* 11, 1095–1099. doi: 10.1101/gr.180501
- Desnues, C., Rodriguez-Brito, B., Rayhawk, S., Kelley, S., Tran, T., Haynes, M., et al. (2008). Biodiversity and biogeography of phages in modern stromatolites and thrombolites. *Nature* 452, 340–343. doi: 10.1038/nature06735
- Edwards, R. A., and Rohwer, F. (2005). Viral metagenomics. *Nat. Rev. Microbiol.* 3, 504–510. doi: 10.1038/nrmicro1163
- Everson, J. S., Garner, S. A., Fane, B. A., Liu, B. L., Lambden, P. R., and Clarke, I. N. (2002). Biological properties and cell tropism of Chp2, a bacteriophage of the obligate intracellular bacterium *Chlamydomonas abortus*. *J. Bacteriol.* 184, 2748–2754. doi: 10.1128/JB.184.10.2748-2754.2002
- Filée, J., Tétart, F., Suttle, C. A., and Krisch, H. M. (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. doi: 10.1073/pnas.0503404102
- Garner, S. A., Everson, J. S., Lambden, P. R., Fane, B. A., and Clarke, I. N. (2004). Isolation, molecular characterisation and genome sequence of a bacteriophage (Chp3) from *Chlamydomonas pecorum*. *Virus Genes* 28, 207–214. doi: 10.1023/B:VIRU.0000016860.53035.f3
- Gobet, A., Böer, S. L., Huse, S. M., van Beusekom, J. E. E., Quince, C., Sogin, M. L., et al. (2012). Diversity and dynamics of rare and of resident bacterial populations in coastal sands. *ISME J.* 6, 542–553. doi: 10.1038/ismej.2011.132
- Godson, G. N., Barrell, B. G., Staden, R., and Fiddes, J. C. (1978). Nucleotide sequence of bacteriophage G4 DNA. *Nature* 276, 236–247. doi: 10.1038/276236a0
- Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010
- Hill, M. (1973). Diversity and evenness: a unifying notation and its consequences. *Ecology* 54, 427–432. doi: 10.2307/1934352
- Hoffmann, K. H., Rodriguez-Brito, B., Breitbart, M., Bangor, D., Angly, F. E., Felts, B., et al. (2007). Power law rank-abundance models for marine phage communities. *FEMS Microbiol. Lett.* 273, 224–228. doi: 10.1111/j.1574-6968.2007.00790.x
- Hopkins, M., Kailasan, S., Cohen, A., Roux, S., Tucker, K. P., Shevenell, A., et al. (2014). Diversity of environmental single-stranded DNA phages revealed by PCR amplification of the partial major capsid protein. *ISME J.* 8, 2093–2103. doi: 10.1038/ismej.2014.43
- Huang, S., Wilhelm, S. W., Jiao, N., and Chen, F. (2010). Ubiquitous cyanobacterial podoviruses in the global oceans unveiled through viral DNA polymerase gene sequences. *ISME J.* 4, 1243–1251. doi: 10.1038/ismej.2010.56

- Huse, S. M., Welch, D. M., Morrison, H. G., and Sogin, M. L. (2010). Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environ. Microbiol.* 12, 1889–1898. doi: 10.1111/j.1462-2920.2010.02193.x
- Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30, 3059–3066. doi: 10.1093/nar/gkf436
- King, A., Adams, M., Carstens, E., and Lefkowitz, E. (2012). “Virus Taxonomy,” in *Ninth Report of the International Committee on Taxonomy of Viruses*, 2nd Edn, eds A. King, M. Adams, E. Carstens, and E. Lefkowitz (San Diego, CA: Elsevier Academic Press).
- Krupovic, M., and Forterre, P. (2011). Microviridae goes temperate: microvirus-related proviruses reside in the genomes of bacteroidetes. *PLoS ONE* 6:e19893. doi: 10.1371/journal.pone.0019893
- Labonté, J. M., Reid, K. E., and Suttle, C. A. (2009). Phylogenetic analysis indicates evolutionary diversity and environmental segregation of marine podovirus DNA polymerase gene sequences. *Appl. Environ. Microbiol.* 75, 3634–3640. doi: 10.1128/AEM.02317-08
- Labonté, J. M., and Suttle, C. A. (2013a). Metagenomic and whole-genome analysis reveals new lineages of gokushoviruses and biogeographic separation in the sea. *Front. Microbiol.* 4:404. doi: 10.3389/fmicb.2013.00404
- Labonté, J. M., and Suttle, C. A. (2013b). Previously unknown and highly divergent ssDNA viruses populate the oceans. *ISME J.* 7, 2169–2177. doi: 10.1038/ismej.2013.110
- Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658–1659. doi: 10.1093/bioinformatics/btl158
- Lilley, M., Baross, J., and Gordon, L. (1982). Dissolved hydrogen and methane in Saanich Inlet, British Columbia. *Deep Sea Res. Part A.* 29, 1471–1484. doi: 10.1016/0198-0149(82)90037-1
- Liu, B. L., Everson, J. S., Fane, B. A., Giannikopoulou, P., Vretou, E., Lambden, P. R., et al. (2000). Molecular characterization of a bacteriophage (Chp2) from *Chlamydia psittaci*. *J. Virol.* 74, 3464–3469. doi: 10.1128/JVI.74.8.3464-3469.2000
- López-Bueno, A., Tamames, J., Velázquez, D., Moya, A., Quesada, A., and Alcami, A. (2009). High diversity of the viral community from an Antarctic lake. *Science* 326, 858–861. doi: 10.1126/science.1179287
- Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376–380. doi: 10.1038/nature03959
- Martin, M. O. (2002). Predatory prokaryotes: an emerging research opportunity. *J. Mol. Microbiol. Biotechnol.* 4, 467–477.
- McDaniel, L. D., Rosario, K., Breitbart, M., and Paul, J. H. (2014). Comparative metagenomics: natural populations of induced prophages demonstrate highly unique, lower diversity viral sequences. *Environ. Microbiol.* 16, 570–585. doi: 10.1111/1462-2920.12184
- Polson, S. W., Wilhelm, S. W., and Wommack, K. E. (2011). Unraveling the viral tapestry (from inside the capsid out). *ISME J.* 5, 165–168. doi: 10.1038/ismej.2010.81
- Quince, C., Lanzen, A., Davenport, R. J., and Turnbaugh, P. J. (2011). Removing noise from pyrosequenced amplicons. *BMC Bioinform.* 12:38. doi: 10.1186/1471-2105-12-38
- R Development Core Team. (2011). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Rodriguez-Brito, B., Li, L., Wegley, L., Furlan, M., Angly, F. E., Breitbart, M., et al. (2010). Viral and microbial community dynamics in four aquatic environments. *Water* 4, 739–751. doi: 10.1038/ismej.2010.1
- Rosario, K., and Breitbart, M. (2011). Exploring the viral world through metagenomics. *Curr. Opin. Virol.* 1, 289–297. doi: 10.1016/j.coviro.2011.06.004
- Roux, S., Enault, F., Robin, A., Ravet, V., Personnic, S., Theil, S., et al. (2012a). Assessing the diversity and specificity of two freshwater viral communities through metagenomics. *PLoS ONE* 7:e33641. doi: 10.1371/journal.pone.0033641
- Roux, S., Krupovic, M., Poulet, A., Debros, D., and Enault, F. (2012b). Evolution and diversity of the *Microviridae* viral family through a collection of 81 new complete genomes assembled from virome reads. *PLoS ONE* 7:e40418. doi: 10.1371/journal.pone.0040418
- Roux, S., Hawley, A. K., Torres Beltran, M., Scofield, M., Schwientek, P., Stepanauskas, R., et al. (2014). Ecology and evolution of viruses infecting uncultivated SUP05 bacteria as revealed by single-cell- and meta- genomics. *Elife* 3:e03125. doi: 10.7554/eLife.03125
- Salim, O., Skilton, R. J., Lambden, P. R., Fane, B. A., and Clarke, I. N. (2008). Behind the chlamydial cloak: the replication cycle of chlamydiae Chp2, revealed. *Virology* 377, 440–445. doi: 10.1016/j.virol.2008.05.001
- Shade, A., Jones, S. E., Caporaso, J. G., Handelsman, J., Knight, R., Fierer, N., et al. (2014). Conditionally rare taxa disproportionately contribute to temporal changes in microbial diversity. *MBio* 5:e01371–e01314. doi: 10.1128/mBio.01371-14
- Short, C. M., and Suttle, C. A. (2005). Nearly identical bacteriophage structural gene sequences are widely distributed in both marine and freshwater environments. *Appl. Environ. Microbiol.* 71, 480–486. doi: 10.1128/AEM.71.1.480-486.2005
- Short, S. M., and Suttle, C. A. (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. doi: 10.1128/AEM.68.3.1290-1296.2002
- Smith, R. J., Jeffries, T. C., Roudnew, B., Seymour, J. R., Fitch, A. J., Simons, K. L., et al. (2013). Confined aquifers as viral reservoirs. *Environ. Microbiol. Rep.* 5, 725–730. doi: 10.1111/1758-2229.12072
- Sogin, M. L., Morrison, H. G., Huber, J. A., Mark Welch, D., Huse, S. M., Neal, P. R., et al. (2006). Microbial diversity in the deep sea and the underexplored “rare biosphere.” *Proc. Natl. Acad. Sci. U.S.A.* 103, 12115–12120. doi: 10.1073/pnas.0605127103
- Storey, C. C., Lusher, M., and Richmond, S. J. (1989). Analysis of the complete nucleotide sequence of Chp1, a phage which infects avian *Chlamydia psittaci*. *J. Gen. Virol.* 70, 3381–3390. doi: 10.1099/0022-1317-70-12-3381
- Sullivan, M. B., Coleman, M. L., Quinlivan, V., Rosenkrantz, J. E., Defrancesco, A. S., Tan, G., et al. (2008). Portal protein diversity and phage ecology. *Environ. Microbiol.* 10, 2810–2823. doi: 10.1111/j.1462-2920.2008.01702.x
- Suttle, C. A. (2007). Marine viruses—major players in the global ecosystem. *Nat. Rev. Microbiol.* 5, 801–812. doi: 10.1038/nrmicro1750
- Suttle, C. A., Chan, A. M., and Cottrell, M. T. (1991). Use of ultrafiltration to isolate viruses from seawater which are pathogens of marine phytoplankton. *Appl. Environ. Microbiol.* 57, 721–726.
- Tucker, K. P., Parsons, R., Symonds, E. M., and Breitbart, M. (2011). Diversity and distribution of single-stranded DNA phages in the North Atlantic Ocean. *ISME J.* 5, 822–830. doi: 10.1038/ismej.2010.188
- Ward, B., Kilpatrick, K., Wopat, A., Minnich, E., and Lidstrom, M. (1989). Methane oxidation in Saanich Inlet during summer stratification. *Cont. Shelf. Res.* 9, 65–75. doi: 10.1016/0278-4343(89)90083-6
- Yoshida, M., Takaki, Y., Eitoku, M., Nunoura, T., and Takai, K. (2013). Metagenomic analysis of viral communities in (had)pelagic sediments. *PLoS ONE* 8:e57271. doi: 10.1371/journal.pone.0057271
- Zaikova, E., Walsh, D. A., Stilwell, C. P., Mohn, W. W., Tortell, P. D., and Hallam, S. J. (2010). Microbial community dynamics in a seasonally anoxic fjord: Saanich Inlet, British Columbia. *Environ. Microbiol.* 12, 172–191. doi: 10.1111/j.1462-2920.2009.02058.x

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Labonté, Hallam and Suttle. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.