



# Metagenomic and whole-genome analysis reveals new lineages of gokushoviruses and biogeographic separation in the sea

Jessica M. Labonté<sup>1†</sup> and Curtis A. Suttle<sup>1,2,3,4\*</sup>

<sup>1</sup> Department of Microbiology and Immunology, University of British Columbia, Vancouver, BC, Canada

<sup>2</sup> Department of Earth, Ocean and Atmospheric Sciences, University of British Columbia, Vancouver, BC, Canada

<sup>3</sup> Department of Botany, University of British Columbia, Vancouver, BC, Canada

<sup>4</sup> Canadian Institute for Advanced Research, University of British Columbia, Vancouver, BC, Canada

## Edited by:

Heather K. Allen, National Animal Disease Center, USA

## Reviewed by:

Andrew D. Millard, University of Warwick, UK

Henry M. Krisch, Centre National de la Recherche Scientifique, France

## \*Correspondence:

Curtis A. Suttle, Department of Earth, Ocean and Atmospheric Sciences, University of British Columbia, #2178-2207 Main Mall, Vancouver, BC V6T 1Z4, Canada  
e-mail: [suttle@science.ubc.ca](mailto:suttle@science.ubc.ca)

## † Present address:

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

Much remains to be learned about single-stranded (ss) DNA viruses in natural systems, and the evolutionary relationships among them. One of the eight recognized families of ssDNA viruses is the *Microviridae*, a group of viruses infecting bacteria. In this study we used metagenomic analysis, genome assembly, and amplicon sequencing of purified ssDNA to show that bacteriophages belonging to the subfamily *Gokushovirinae* within the *Microviridae* are genetically diverse and widespread members of marine microbial communities. Metagenomic analysis of coastal samples from the Gulf of Mexico (GOM) and British Columbia, Canada, revealed numerous sequences belonging to gokushoviruses and allowed the assembly of five putative genomes with an organization similar to chlamydiamicroviruses. Fragment recruitment to these genomes from different metagenomic data sets is consistent with gokushovirus genotypes being restricted to specific oceanic regions. Conservation among the assembled genomes allowed the design of degenerate primers that target an 800 bp fragment from the gene encoding the major capsid protein. Sequences could be amplified from coastal temperate and subtropical waters, but not from samples collected from the Arctic Ocean, or freshwater lakes. Phylogenetic analysis revealed that most sequences were distantly related to those from cultured representatives. Moreover, the sequences fell into at least seven distinct evolutionary groups, most of which were represented by one of the assembled metagenomes. Our results greatly expand the known sequence space for gokushoviruses, and reveal biogeographic separation and new evolutionary lineages of gokushoviruses in the oceans.

**Keywords:** biogeography, ssDNA viruses, *Microviridae*, *Gokushovirinae*, virus diversity, ocean viruses

## INTRODUCTION

Viruses are the most abundant (Suttle, 2005) and diverse (Breitbart et al., 2002; Angly et al., 2006) biological entities in the oceans. By causing lysis of specific subsets of microbial communities, they influence community composition by controlling species evenness and maintaining species richness (Hennes et al., 1995; Thingstad, 2000; Wommack and Colwell, 2000; Middelboe et al., 2001; Weinbauer, 2004; Winter et al., 2010); thereby, influencing nutrient and energy cycling (Fuhrman, 1999; Wilhelm and Suttle, 1999; Suttle, 2007). Moreover, viruses harbor an enormous pool of genetic diversity that can be exchanged among other viruses (Pedulla et al., 2003; Short and Suttle, 2005) and bacteria (Fuhrman and Schwalbach, 2003; Kenzaka et al., 2010). Despite the abundance of bacteriophages in marine systems (often  $>10^7$  ml<sup>-1</sup>) and their important role in marine systems, relatively little is known about the distribution and composition of most groups of marine viruses.

Metagenomic approaches have provided an in-depth look at the molecular diversity of ssDNA viruses in a range of

environments including marine systems (Breitbart et al., 2002; Angly et al., 2006; Bench et al., 2007), the human gut (Zhang et al., 2006; Breitbart et al., 2008; Minot et al., 2011), modern stromatolites (Desnues et al., 2008), and freshwaters (Kim et al., 2008; López-Bueno et al., 2009; Roux et al., 2012a). Recently, 608 ssDNA viral genomes were assembled from marine metagenomic data revealing far greater evolutionary diversity in ssDNA viruses than previously known (Labonté and Suttle, 2013).

Gokushoviruses are ssDNA bacteriophages belonging to the family *Microviridae* and are represented among sequences found in metagenomic data. For example, gokushovirus genomes were assembled from a wide range of environments by mining of metagenomic data, with 42 assembled from a variety of ecosystems (Roux et al., 2012b), and two others from data collected from the North Atlantic Ocean (Tucker et al., 2011), indicating the widespread occurrence of gokushoviruses. These viruses have a ~30 nm icosahedral capsid encompassing a positive ssDNA molecule of 4.4 to 4.8 kb that encodes five major proteins. Based on the phylogeny of the major capsid protein (VP1) of isolates,

the *Microviridae* are divided into two subfamilies (Brentlinger et al., 2002). Members of the *Microvirinae* (e.g., phiX174 and G4) infect enterobacteria including *Escherichia coli* (Godson et al., 1978), while members of the *Gokushovirinae* infect parasitic bacteria. The latter includes Chp1 (Storey et al., 1989), Chp2 (Liu et al., 2000; Everson et al., 2002) and Chp3 (Garner et al., 2004) that infect *Chlamydia* spp., while phiMH2K (Brentlinger et al., 2002) and SpV4 (Chipman et al., 1998) infect *Bdellovibrio* sp. and *Spiroplasma* sp., respectively.

There are no reported gokushovirus isolates, and their hosts remain unknown. Based on bacterial genomic sequences bacteria in the *Bacteroidetes* appear to be hosts for a proposed subfamily, *Alpavirinae* (Krupovic and Forterre, 2011), of previously unknown microviruses. As well, eight ssDNA phages have been isolated that are morphologically similar but genetically different to microviruses (Holmfeldt et al., 2013, 2012).

Our study examined the genetic diversity and relatedness of *Gokushovirinae*-like viruses from temperate and subtropical coastal environments. From three ssDNA-enriched metagenomic datasets we assembled and phylogenetically compared five new gokushovirus genomes. Recruitment of metagenomic reads to these genomes showed spatial differences in the most abundant gokushovirus genotypes. The genetic richness of gokushoviruses was also assessed through amplification of a ~800 bp fragment of the conserved gene encoding the major capsid protein, VP1. These results reveal biogeographic separation and new evolutionary lineages of marine gokushoviruses, and likely reflect the underlying distributions of their hosts.

## MATERIALS AND METHODS

### COLLECTION AND PREPARATION OF SAMPLES

Samples (~20 to ~200 L) were collected using GO-FLO or Niskin bottles either mounted on a CTD rosette or directly on a hydrographic wire [Saanich Inlet (SI)], or by bucket from the surface (lake samples). For each sample, the viruses were concentrated ~10–100-fold (~200 mL final volume) using ultrafiltration (Suttle et al., 1991). Briefly, particulate matter was removed by pressure filtering (<17 kPa) the samples through 142-mm diameter glass-fiber filters (MFS GC50, nominal pore size 1.2 μm) and polyvinylidene difluoride filters (Millipore GVWP, pore size 0.22 μm) connected in series. The viral size fraction in the filtrate was then concentrated by ultrafiltration through a 30 kDa molecular weight cut-off cartridge (Amicon S1Y30, Millipore), and stored at 4°C in the dark until processed.

In order to integrate variation within a region, numerous virus concentrates (VCs) collected from different locations and at different times within a geographic region were combined into a single mix. Except for the SI and freshwater samples, these mixes corresponded to those used by Angly et al. (2006) in which the first ssDNA viral sequences were reported from marine metagenomic data. VCs from the Strait of Georgia (SOG) and surrounding inlets and bays were pooled into the following four mixes by combining 2 mL of each VC: BC1-1999 (23 samples), BC3-2000 (26 samples), BC4-2004 (16 samples), and BC2-Low salinity (19 samples). Similarly, samples from the Gulf of Mexico (GOM) were combined into four mixes from the eastern GOM (8 samples), northern GOM (6 samples), western GOM (6 samples),

and the Texas Coast (13 samples), while samples from the Arctic Ocean were made into mixes from the Beaufort Sea (20 samples), Chukchi Sea (14 samples) and High Arctic (22 samples). To look at the diversity of freshwater gokushoviruses, two mixes were made from Chilliwack (6 samples) and Cultis (8 samples) Lakes. An extensive description of all the samples that were combined in each mix is presented in the Supplementary Material of Labonté and Suttle (2013). SI is unusual as it undergoes seasonal anoxia (Zaikova et al., 2010). For the metagenomic study, we combined surface samples from April 2007, and January, March, May, July, August, and November 2008. PCR amplifications were performed on the following nine samples from SI: 10, 120, and 150 m samples from April 2007, and surface samples from January, March, May, July, August, and November 2008.

### ssDNA PREPARATION

As described in Labonté and Suttle (2013), ssDNA was prepared from 10 mL of 0.22-μm filtered (PDVF; Millipore) pooled mixes from British Columbia (SOG), the GOM, the Lakes (LA), and the Arctic (ARC), or from 10 mL of each individual VCs from SI. Briefly, ssDNA was extracted using a silica column and amplified using multiple-displacement whole-genome amplification (WGA) to convert ssDNA into dsDNA. Pure amplified dsDNA was resuspended in ultrapure H<sub>2</sub>O for pyrosequencing or Tris-HCl for PCR amplification.

### GENOME ANALYSIS, BINNING, AND ASSEMBLY

Metagenomic libraries were constructed from WGA ssDNA from SI, SOG, and GOM ssDNA. The purified WGA DNA was resuspended in 100 μL of RNase- and DNase-free water (Invitrogen) and concentrated using a Millipore YM-30 Microcon centrifugal filter to a final volume of ~50 μL; 3–5 μg of DNA from each sample was sent for pyrosequencing (Roche 454 FLX instrumentation with Titanium chemistry) at Genome Québec, McGill University (SOG) and the Broad Institute at the Massachusetts Institute of Technology (GOM and SI).

The sequences were quality and linker trimmed, and assembled into contiguous sequences (contigs) using the Newbler Assembler (Roche). The individual reads and assembled sequences were compared to a database of all available genomes in GenBank (as of February 2010) from viruses belonging to the *Microviridae* using the tBLASTx algorithm with an *e*-value cut-off of 10<sup>-5</sup>. Reads with significant similarity to gokushoviruses were aligned onto the assembled contigs using the add454Reads.perl script and were reassembled into new contigs using the phredPhrap.perl script of the Consed package (Gordon, 2003). Additional contig analyses (BLAST, circularization of the genomes, annotations, alignments, and phylogeny) were performed within the Geneious Pro package v5.6 (Biomatters).

### PRIMER DESIGN AND PCR AMPLIFICATION

Two forward (MicroVP1-F1, 5'-CGN GCN TAY AAY TTR ATH-3'; MicroVP1-F2, 5'-AGN GCN TAY AAY TTR CTN-3') and two reverse (MicroVP1-R1, 5'-TTY GGN TAY CAR GAR AGN-3'; MicroVP1-R2, 5'-NCT YTC YTG RTA NCC RAA-3') primers with respective degeneracies of 256, 215, 256, and 256 were designed from alignments of the inferred

amino acid sequences of the major capsid protein (VP1) of the chlamydia phages Chp1 (accession number NC\_001741.1), Chp2 (NC\_002194.1), Chp3 (NC\_008355.1), Chp4 (NC\_007461.1), phiCPAR39 (NC\_002180.1), and phiCPG1 (NC\_001998.1), Spiroplasmaphage Sp4 (NC\_003438.1), bdellovibriophage phiMH2K (NC\_002643.1), and the Sargasso Sea Chp1-like assembled genome (Angly et al., 2006; Tucker et al., 2011). The primers amplify a ~800 bp VP1 gene fragment from the subfamily *Gokushovirinae* in the *Microviridae*.

Prior to use in PCR reactions, the purified WGA DNA was resuspended in 100  $\mu$ L of TE, and 10  $\mu$ L was used as a 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 MgCl<sub>2</sub>, 125  $\mu$ M of each deoxyribonucleoside triphosphate, 1  $\mu$ M of each MicroVP1-F1, MicroVP1-F2 and MicroVP1-R1 and MicroVP1-R2 primer and 2.5 U of PLATINUM Taq DNA polymerase (Invitrogen). 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.

#### CLONE LIBRARY CONSTRUCTION AND RFLP ANALYSIS

PCR amplicons were purified with a MinElute PCR purification kit (Qiagen), ligated into pCR2.1-Topo (Invitrogen), and used to transform chemically competent *E. coli* Top10 cells. For each sample, 30 clones were checked by colony PCR to verify that they contained an insert of the correct size. Restriction fragment length polymorphism (RFLP) analysis was then performed on 20 positive clones. For each RFLP reaction, 15  $\mu$ L of colony PCR product was digested with *AluI* (New England BioLabs) in a reaction containing 1 U/ $\mu$ g of DNA and 1 $\times$  NEBuffer 4 (20 nM Tris-acetate, 50 mM potassium acetate, 10 mM magnesium acetate, 1 mM DTT, pH 7.9) by incubating at 37°C for 16 h, followed by heat inactivation at 65°C for 20 min. RFLP products were separated on a 2% agarose gel in 0.5 $\times$  TBE (9 mM Tris base, 9 mM boric acid, 2 mM EDTA, pH 8.0) running at 110 V for ~2 h. Sequencing of representative clones confirmed that each unique restriction pattern could be considered as an operational taxonomic unit (OTU). Forward and reverse sequences (~800 bp) were obtained for each RFLP pattern using Big-Dye Terminator Cycle Sequencing (Applied Biosystems) and ABI 373 Stretch or ABI Prism 377 sequencers (Nucleic Acid Protein Service Unit, UBC).

#### PHYLOGENETIC ANALYSIS

For the whole genome phylogeny, non-coding sequences were removed and the five major open reading frames were ordered. The sequences were aligned using MAFFT (Katoh et al., 2002) and maximum likelihood analysis with 100 bootstrap replicates were performed using PhyML (Guindon et al., 2010).

VP1 from the previously sequenced isolates, environmental sequences, and the degenerate PCR products from this study were trimmed to the PCR-product length (~800 bp) and aligned using MAFFT (Katoh et al., 2002). The alignment was cured with GBlocks to remove unconserved regions that aligned with multiple gaps using the less stringent setting (allowing for smaller final blocks, gap positions within the final blocks and less strict

flanking positions) (Talavera and Castresana, 2007). Bayesian phylogenetic analyses were performed on the cured alignment with MrBayes (Huelsenbeck and Ronquist, 2001). MrBayes uses a Markov chain Monte Carlo (mcmc) approach to approximate prior and posterior probabilities. Under the HKY85 substitution model with an invgamma distribution, two independent analyses of 4 (1 cold and 3 heated) mcmc chains with 20,000,000 cycles were run, sampled every 1000th cycle. The consensus tree was generated in Geneious with a burnin of 25%. Trees were viewed in Fig Tree (<http://tree.bio.ed.ac.uk/software/figtree/>).

#### FRAGMENT RECRUITMENT

Recruitment of the reads from metagenomic data sets onto the assembled genomes was performed using tBLASTx with an *e*-value of 10<sup>-10</sup> and allowing only one hit per read. The metagenomic reads from the marine viromes (Angly et al., 2006) and microbialites (Desnues et al., 2008) were obtained from the CAMERA database, while the metagenomic reads from Lake Pavin and Lake Bourget (Roux et al., 2012a) were obtained from the SEED database. The environmental genomes used were Lake\_Bourget\_052, Lake\_Bourget\_523, Lake\_Pavin\_279 and 68\_Microbialite\_063 from Roux et al. (2012b), and SARssp12 from Tucker et al. (2011).

#### NUCLEOTIDE SEQUENCE ACCESSION NUMBERS

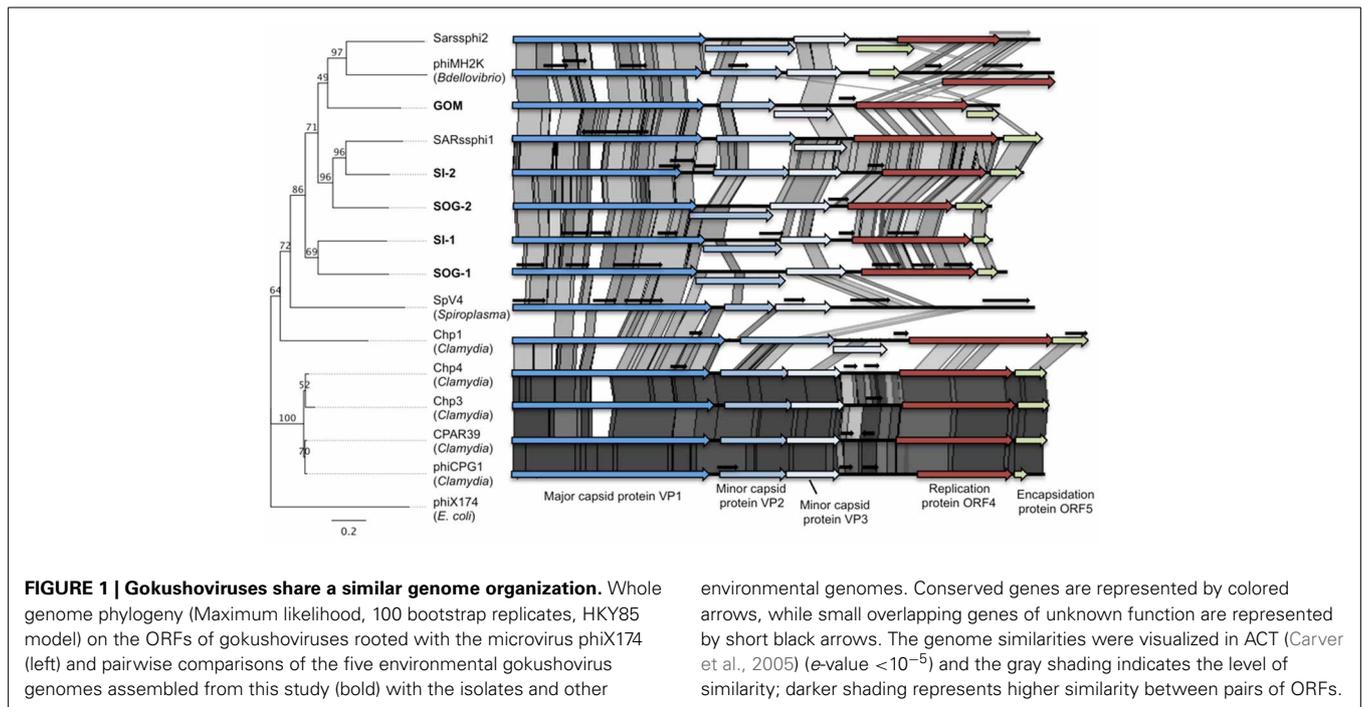
The five complete gokushovirus genomes as well as the 43 environmental PCR product sequences were submitted to Genbank and are available under the accession numbers KC131021-KC131025 and KC130978-KC131020, respectively.

## RESULTS AND DISCUSSION

#### ASSEMBLY OF COMPLETE GOKUSHOVIRUS GENOMES

Sequence analysis of ssDNA metagenomic libraries from the SOG, SI, and the GOM recovered 1733, 374, and 194 sequences, respectively, that were significantly similar to sequences from viruses belonging to the *Microviridae*, with >90% of them being most similar to sequences belonging to the chlamydia microviruses and other gokushoviruses. From these data, five complete circular genomes were assembled with at least 3-fold coverage (two from SOG, two from SI and one from GOM). The genome sizes varied from 4062 to 5386 bp, and were uniformly shorter than those from previously sequenced isolates (Figure 1). Assembly of these genomes represented the accumulation of 95 reads for SOG-1, 58 for SOG-2, 53 for SI-1, 48 for SI-2, and 38 for GOM.

Even though there was only ~30–50% similarity at the nucleotide level among the assembled genomes (Table 1), the chlamydia phages and bdellovibriophage phiMH2K, the gene organization was remarkably similar among them, and included the five proteins required for replication of gokushoviruses (Figure 1), implying a common evolutionary origin. These comprise VP1, the major capsid protein, VP2 that is hypothesized to be involved in host recognition (Chipman et al., 1998) and virus attachment (Everson et al., 2003), VP3 that is a scaffolding protein found in the procapsid only and not in mature virions (Clarke et al., 2004), ORF4 that is a replication initiator involved in ssDNA synthesis (Liu et al., 2000; Garner et al., 2004; Salim et al., 2008), and ORF5 that is involved in DNA packaging (Liu



**FIGURE 1 | Gokushoviruses share a similar genome organization.** Whole genome phylogeny (Maximum likelihood, 100 bootstrap replicates, HKY85 model) on the ORFs of gokushoviruses rooted with the microvirus phiX174 (left) and pairwise comparisons of the five environmental gokushovirus genomes assembled from this study (bold) with the isolates and other

environmental genomes. Conserved genes are represented by colored arrows, while small overlapping genes of unknown function are represented by short black arrows. The genome similarities were visualized in ACT (Carver et al., 2005) (e-value < 10<sup>-5</sup>) and the gray shading indicates the level of similarity; darker shading represents higher similarity between pairs of ORFs.

**Table 1 | Similarity matrix of the coding regions of the five environmental gokushovirus genomes assembled from this study (bold) with the isolates and other environmental genomes.**

	<b>GOM</b>	<b>SI1</b>	<b>SI2</b>	<b>SOG1</b>	<b>SOG2</b>	<b>SAR ssphi2</b>	<b>SAR ssphi1</b>	<b>MH2K</b>	<b>SpV4</b>	<b>Chp1</b>	<b>Chp3</b>	<b>Chp4</b>	<b>CPAR 39</b>	<b>CPG1</b>	<b>X174</b>
GOM	<b>36.5</b>	40.6	38.2	42.6	38.0	39.4	35.7	32.1	33.5	38.7	39.1	39.2	39.2	25.3	
SI1	36.5	<b>38.7</b>	40.5	40.1	35.8	39.4	33.5	31.8	35.4	38.4	38.9	39.0	39.0	25.8	
SI2	40.6	38.7	<b>40.3</b>	49.6	37.4	49.6	35.8	32.2	35.2	41.4	41.9	42.1	42.1	27.1	
SOG1	38.2	40.5	40.3	<b>40.0</b>	35.7	39.3	35.9	31.3	35.1	40.3	40.7	40.8	40.7	26.3	
SOG2	42.6	40.1	49.6	40.0	<b>38.4</b>	44.6	37.3	33.3	34.4	40.5	41.2	41.4	41.4	26.1	
SARssphi2	38.0	35.8	37.4	35.7	38.4	<b>37.3</b>	34.6	28.0	30.1	36.0	36.5	36.7	36.6	24.0	
SARssphi1	39.4	39.4	49.6	39.3	44.6	37.3	<b>35.8</b>	30.7	33.4	38.8	38.8	39.2	39.0	25.5	
phiMH2K (Bdellovibrio)	35.7	33.5	35.8	35.9	37.3	34.6	35.8	<b>29.0</b>	30.7	36.9	37.2	37.4	37.3	24.7	
SpV4 (Spiroplasma)	32.1	31.8	32.2	31.3	33.3	28.0	30.7	29.0	<b>34.6</b>	37.9	39.0	39.1	38.9	24.9	
Chp1 (Chlamydia)	33.5	35.4	35.2	35.1	34.4	30.1	33.4	30.7	34.6	<b>44.6</b>	44.7	44.7	44.5	28.6	
Chp3 (Chlamydia)	38.7	38.4	41.4	40.3	40.5	36.0	38.8	36.9	37.9	44.6	<b>91.0</b>	91.5	91.6	31.3	
Chp4 (Chlamydia)	39.1	38.9	41.9	40.7	41.2	36.5	38.8	37.2	39.0	44.7	91.0	<b>94.3</b>	94.4	32.1	
CPAR39 (Chlamydia)	39.2	39.0	42.1	40.8	41.4	36.7	39.2	37.4	39.1	44.7	91.5	94.3	<b>97.0</b>	31.9	
phiCPG1 (Chlamydia)	39.2	39.0	42.1	40.7	41.4	36.6	39.0	37.3	38.9	44.5	91.6	94.4	97.0	31.8	
phiX174	25.3	25.8	27.1	26.3	26.1	24.0	25.5	24.7	24.9	28.6	31.3	32.1	31.9	31.8	

(i.e. ■ >90%, ■ 45–89%, ■ 30–45%, and □ <30%).

et al., 2000; Garner et al., 2004; Salim et al., 2008). The presence of all five essential genes in the assembled genomes strongly suggests they represent complete sequences from extant viruses in the environment.

Whole genome phylogeny revealed that the environmental genomes cluster more closely with the bdellovibriophage phiMH2K, rather than the chlamydiae (Figure 1), suggesting that the host for these gokushoviruses is more closely related to *Bdellovibrio* spp., which are found in marine waters, than

*Chlamydia* spp. Whole genome pairwise comparisons showed that VP2 and ORF4 are the least conserved genes, with very few regions of conservation. Moreover, there is 91–97% similarity among chlamydiae, while only 28–49% similarity among the environmental phages. A recombination event in which ORF4 and ORF5 are inverted in phiMH2K, which infects the bacterial parasite *Bdellovibrio bacteriovorus*, and in the environmental genome SAR phi2. These genomes also cluster together suggesting a common evolutionary history (Figure 1).

All of the environmental genomes were shorter than those from isolates. Some, such as SI-1 and SOG-1, had multiple overlapping genes of unknown function. It is postulated (Rokyta and Burch, 2006) that ssDNA microviruses, such as the coliphages phiX174 and G4, evolve differently than dsDNA viruses because of strictly lytic life cycles, small genomes, and low rates of horizontal gene transfer (Breitbart and Rohwer, 2005; Comeau and Buenaventura, 2005; Hambly and Suttle, 2005). Novel genes were predicted to originate by overprinting rather than by horizontal gene transfer (Pavesi, 2006).

### GENETIC RELATEDNESS AMONG GENES ENCODING THE MAJOR CAPSID PROTEIN

To look more deeply at the genetic richness of gokushoviruses, degenerate primers were designed to amplify a ~800 bp fragment of the gene encoding the major capsid protein (VP1) that has interspaced conserved and variable regions. For the assembled genomes, the phylogeny of VP1 is congruent with the whole genome; thus, the phylogeny of VP1 can be used to infer viral phylogeny. PCR amplification was performed on samples from the SOG (4 mixes), GOM (4 mixes), SI (9 samples), and Arctic (ARC; 3 mixes) (Table S1). No products were amplified from the ARC, LA, SOG-Low Salinity, or Eastern GOM mixes. This means that gokushoviruses were absent or at low concentrations in these samples, or that they are too divergent to be amplified by the primers.

Twenty VP1 clones from each of the 20 samples were digested using *AluI* to reveal 77 different RFLP patterns. Representative clones from each restriction pattern were sequenced (data not shown). Of these, 43 sequences were at least 98% different at the nucleotide level, and thus identified as unique gokushovirus VP1 sequences. Some sequences occurred in more than one sample from the same geographic region; for example, the sequence SI-07 was sequenced from multiple dates in SI (i.e., January, April, May and Aug), but no sequences occurred in more than one location (Figure 2). Some sequences were found in both SOG and SI, but no sequences were found in GOM and SOG, or GOM and SI. The 85 % of the VP1 sequences that contained the primer sequences and translated into a putative protein were kept for further analysis. The nucleotide alignment revealed multiple regions of conservation, as well as regions that were confined to specific groups, agreeing with observations made during primer design.

The amplified VP1 PCR products were compared to VP1 sequences containing the primer sequences from the assembled genomes, the chlamydiamicrovirus isolates, as well as Genbank environmental sequences from modern stromatolites (Desnues et al., 2008), freshwater (Lake Needwood, MD; Kuzmickas et al., unpublished), rice paddy soil (Kim et al., 2008) and marine genomes (Venter et al., 2004). Few sequences were similar to those from isolates. Phylogenetic analysis using Bayesian (Figure 2) and maximum-likelihood algorithms produced similar trees with gokushoviruses sub-divided into at least seven well-supported new groups containing more than two sequences (Figure 2). Five of the new clades are represented by an assembled genome or sequenced isolate (Figure 2). Several sequences, such as SOG4-04 and SI-18, were too divergent to be assigned to a cluster;

however, as only 20 clones were analyzed for each sample, rarer phylogenetic clusters were poorly sampled.

Sequences from a given location were usually more closely related to ones from the same location; most GOM sequences clustered within ENV6 and ENV7, while ENV2 is represented exclusively by SOG sequences. Sequences found in more than one sample also usually clustered together. For example, the sequences SI-10 and SI-07, which were found in SI on multiple dates, clustered within ENV5, along with sequences from GOM and SI-2. Collectively, these data imply that viruses in the ENV5 group are widespread in nature. Other data from modern stromatolites and marine genomes clustered together as specific phylogenetic groups.

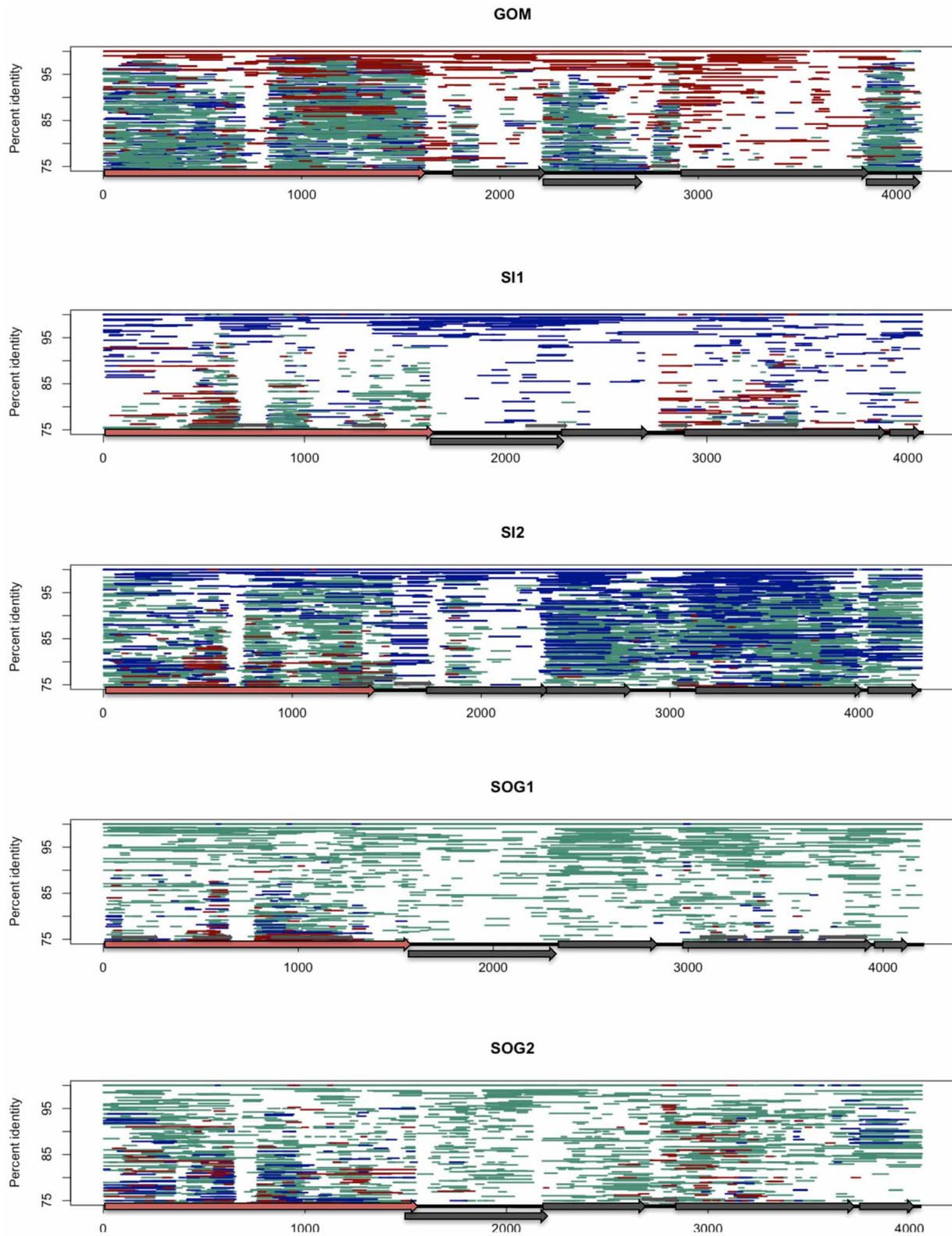
### HOST SPECIFICITY AND GEOGRAPHIC DISTRIBUTION OF ENVIRONMENTAL GOKUSHOVIRUS GENOMES

Isolates in the *Gokushovirinae* infect parasitic bacteria, such as *Chlamydia* spp., *Bdellovibrio* spp., and *Spiroplasma* spp., with host specificity likely being dictated by variable genomic regions. To investigate conserved and variable motifs, metagenomic reads from our ssDNA data, as well as other viral metagenomic data sets from marine (Angly et al., 2006), freshwater (Roux et al., 2012a), and microbialite (Desnues et al., 2008) environments were recruited against environmental gokushovirus genomes (Figure 3). Recruitment was more even when the reads were recruited against genomes assembled from metagenomic data collected from the same region (Figures 3, 4; Figures S1, S2). For the SOG-1, SOG-2, and SI-2 genomes, few reads were recruited from data sets other than those from which the genomes were assembled, suggesting that these genomes are not widespread (Figure 3). In contrast, reads from all of the metagenomic data sets aligned on the GOM genome (Figure 3), indicating a wider geographic distribution of these viruses. The high level of recruitment from other data sets on the GOM genome is also congruent with the phylogenetic clustering of the VP1 gene with other VP1 sequences that were present in multiple samples (Figure 2).

The distribution of reads from other environmental samples that recruited to the assembled genomes was very uneven, showing regions of higher conservation within VP1 and VP3, while few reads were recruited to the VP2 region, indicating high variability in this gene (Figure 4). Since the VP2 sequences of the assembled genomes differ from those of isolates, and VP2 encodes for the minor capsid protein involved in host recognition (Chipman et al., 1998), the environmental sequences are likely not from viruses infecting the genera *Chlamydia*, *Bdellovibrio*, or *Spiroplasma*. Recruitment to ORF4 was limited to the source environment for the assembled genomes, and metagenomic data from British Columbia were not recruited to ORF4 of the GOM genome. Thus, both the pairwise comparison (Figure 1) and the recruitment of metagenomic reads (Figure 3) showed that VP2 and ORF4 are less conserved. Similar patterns were observed with genomes assembled from other data sets (Figures S1, S2), showing that specific gokushovirus genotypes are restricted in distribution.

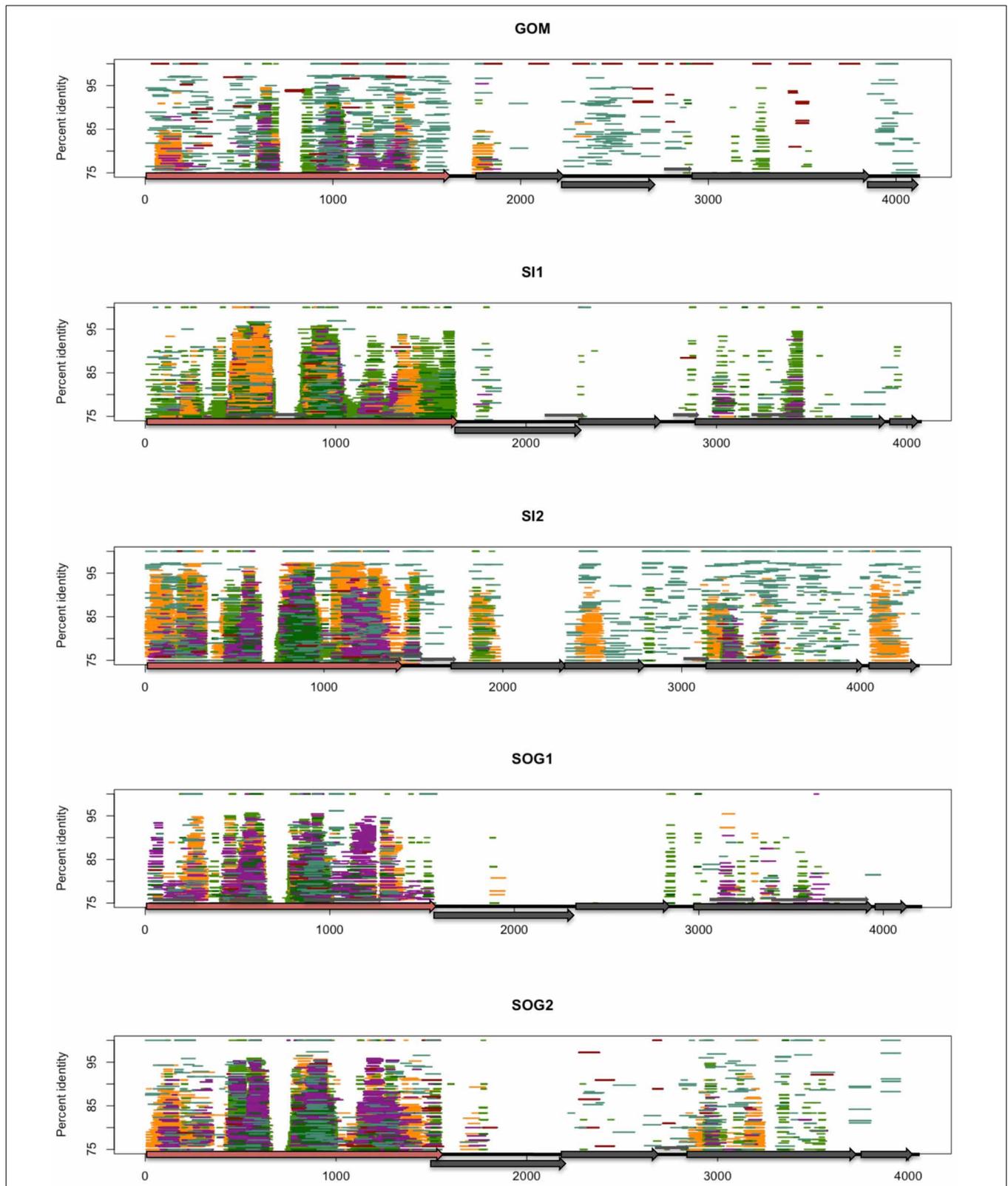
No sequences similar to gokushoviruses were amplified from the Arctic Ocean or two lakes in British Columbia. However,





**FIGURE 3 | Fragment recruitment of the viral ssDNA reads on the genomes from this study to show regions of conservation among ssDNA gokushoviruses.** Each assembled genome (GOM, SI1, SI2, SOG1, SOG2) is represented by a different panel. Each horizontal line represents a metagenomic read from ssDNA data

sets from the Gulf of Mexico (dark red), Saanich Inlet (Dark blue), and Strait of Georgia (aqua) on each of the assembled genomes. Reads were recruited using tBLASTx with an  $e$ -value of  $10^{-10}$ . The position of each line represents the percent similarity of the read to the genome.



**FIGURE 4 | Fragment recruitment of reads from environmental viral metagenomes to show the regions of conservation within different environments.** Each assembled genome (GOM, SI1, SI2, SOG1, SOG2) is represented by a different panel. Each horizontal line represents a read recruited from one of the following publicly available metagenomic data sets:

Gulf of Mexico (dark red), Strait of Georgia (aqua), Sargasso Sea (orange), Lake Bourget (light green), Lake Pavin (dark green), and microbialites (purple). Reads were recruited against each of the assembled genomes using tBLASTx with an  $e$ -value of  $10^{-10}$ . The position of each line represents the percent similarity of the read to the genome. VP1 is represented by a red arrow.

Based on previous work (Labonté and Suttle, 2013), gokushovirus sequences were not the most abundant viruses in our samples, and comprised only 1.6, 0.4, and 0.2% of the metagenomic reads from the SOG, SI, and GOM, respectively. In contrast, in metagenomic data from the Sargasso Sea, gokushovirus sequences comprised nearly 6% of the reads (Angly et al., 2006), while in Lake Bourget they were more than 90% of the sequences (Roux et al., 2012a). These differences may be because the small genomes of gokushoviruses permit rapid replication and high burst sizes, and allow them to dominate following a lytic event, this is consistent with the hypothesis that the most abundant marine viruses are virulent opportunists that replicate rapidly, have high burst sizes and small genomes in order to exploit rapidly growing populations of rare marine bacteria such as *Roseobacter* spp. or *Vibrio* spp. (Suttle, 2007). For example, ~500 genomes are produced each time the chlamydia phage Chp2 infects its parasitic host (Salim et al., 2008). High burst sizes coupled with genomes usually <5 kb support the idea that gokushoviruses are highly virulent and are selected for rapid population growth, which are characteristics of r-strategists. In contrast, many large DNA viruses have a low burst size, large genome and decay slowly, which are characteristics of K-strategists.

Discovering the hosts of marine gokushoviruses is a high priority in order to understand the roles that these viruses play in ecosystems. Given the challenges in culturing marine microbes, culture-independent techniques will likely be needed to determine the hosts for most of these viruses. One approach that we have tried with some success is to use fluorescence *in situ* hybridization (FISH) using labeled VP1 sequences to probe natural microbial communities. Another approach that has been used to visualize phage-infected gammaproteobacterial cells is phage-FISH (Allers et al., 2013), which could be adapted to search for cells infected by gokushoviruses. Finally, single-cell genomics (SCG) allows everything in a cell, including plasmids and viruses to be sequenced (Stepanuskas, 2012). If applied to samples with abundant gokushoviruses, it should be possible to sequence infected cells.

This manuscript presents a new set of degenerate primers that have been used to reveal at least five new evolutionary groups of gokushoviruses, and clearly show they share a common evolutionary history with viruses that infect the obligate intracellular parasitic bacteria *Chlamydia* and *Bdellovibrio*. Phylogenetic analysis of the major capsid protein, combined with fragment recruitment of environmental metagenomic sequences shows that the distribution of some evolutionary groups of gokushoviruses is very environment dependent, whereas others are more cosmopolitan. The high-burst size, rapid replication rates and likely lytic nature of these viruses suggests that they may play an important role as mortality agents in marine systems.

## 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. This research was supported by the Natural Science and Engineering Research Council of Canada (NSERC) through a postgraduate scholarship (Jessica M. Labonté) and Discovery grants (Curtis

A. Suttle). Sample collection was facilitated through ship-time grants from NSERC that supported sample collections from the Strait of Georgia (Curtis A. Suttle) and Saanich Inlet (PD Tortell and SJ Hallam), the US National Science Foundation (Gulf of Mexico), and through the Canadian Arctic Shelf Exchange Study (NSERC) and the Japan/Canada Western Arctic Climate Study. 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 the McGill University and Genome Quebec Innovation Centre.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fmicb.2013.00404/abstract>

**Figure S1 | Fragment recruitment of viral ssDNA reads onto assembled environmental gokushovirus genomes.** Each assembled genome (Lake\_Bourget\_052, Lake\_Bourget\_523, Lake\_Pavin\_279, 68\_Microbialite\_063, and SARssphi2) is represented by a different panel. Each horizontal line represents a read recruited from one of the metagenomic data sets from this study: Gulf of Mexico (dark red), Saanich Inlet (Dark blue), and Strait of Georgia (aqua). Reads were recruited against each of the assembled genomes using tBLASTx with an *e*-value of  $10^{-10}$ . The position of each line represents the percent similarity of the read to the genome. VP1 is represented by a red arrow.

**Figure S2 | Fragment recruitment of reads from environmental viral metagenomes to show the regions of conservation within different environments.** Each assembled genome (Lake\_Bourget\_052, Lake\_Bourget\_523, Lake\_Pavin\_279, 68\_Microbialite\_063, and SARssphi2) is represented by a different panel. Each horizontal line represents a read recruited from one of the following publicly available metagenomic data sets: Gulf of Mexico (dark red), Strait of Georgia (aqua), Sargasso Sea (orange), Lake Bourget (light green), Lake Pavin (dark green), and microbialites (purple) metagenomic data sets on each of the assembled genomes that recruited at using tBLASTx with an *e*-value of  $10^{-10}$ . The height of line represent the percent similarity of the read to the genome. VP1 is represented by a red arrow.

## REFERENCES

- Allers, E., Moraru, C., Duhaime, M. B., Benezé, E., Solonenko, N., Barrero-Canosa, J., et al. (2013). Single-cell and population level viral infection dynamics revealed by phageFISH, a method to visualize intracellular and free viruses. *Environ. Microbiol.* 15, 2306–2318. doi: 10.1111/1462-2920.12100
- 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
- Bench, S. R., Hanson, T. E., Williamson, K. E., Ghosh, D., Radosovich, M., Wang, K., et al. (2007). Metagenomic characterization of Chesapeake Bay virioplankton. *Appl. Env. Microb.* 73, 7629–7641. doi: 10.1128/AEM.00938-07
- Biers, E. J., Sun, S., and Howard, E. C. (2009). Prokaryotic genomes and diversity in surface ocean waters: interrogating the global ocean sampling metagenome. *Appl. Env. Microb.* 75, 2221–2229. doi: 10.1128/AEM.02118-08
- Breitbart, M., Haynes, M., Kelley, S., Angly, F. E., Edwards, R. A., Felts, B., et al. (2008). Viral diversity and dynamics in an infant gut. *Res. Microbiol.* 159, 367–373. doi: 10.1016/j.resmic.2008.04.006
- 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
- Breitbart, M., and Rohwer, F. (2005). Here a virus, there a virus, everywhere the same virus? *Trends Microbiol.* 13, 278–284. doi: 10.1016/j.tim.2005.04.003

- Breitbart, M., Salamon, P., Andresen, B., Mahaffy, J. M., Segall, A. M., Mead, D., et al. (2002). Genomic analysis of uncultured marine viral communities. *Proc. Natl. Acad. Sci. U.S.A.* 99, 14250–14255. doi: 10.1073/pnas.202488399
- 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
- Carver, T. J., Rutherford, K. M., Berriman, M., Rajandream, M.-A., Barrell, B. G., and Parkhill, J. (2005). ACT: the artemis comparison tool. *Bioinformatics* 21, 3422–3423. doi: 10.1093/bioinformatics/bti553
- Chen, F., and Suttle, C. A. (1996). Evolutionary relationships among large double-stranded DNA viruses that infect microalgae and other organisms as inferred from DNA polymerase genes. *Virology* 219, 170–178. doi: 10.1006/viro.1996.0234
- 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
- Clarke, I. N., Cutcliffe, L. T., Everson, J. S., Garner, S. A., Lambden, P. R., Pead, P. J., et al. (2004). Chlamydiophage Chp2, a skeleton in the phiX174 closet: scaffolding protein and procapsid identification. *J. Bacteriol.* 186, 7571–7574. doi: 10.1128/JB.186.22.7571-7574.2004
- Comeau, A. M., and Buenaventura, E. (2005). A persistent, productive, and seasonally dynamic vibriophage population within Pacific oysters (*Crassostrea gigas*). *Appl. Environ. Microb.* 71, 5324–5331. doi: 10.1128/AEM.71.9.5324-5331.2005
- 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
- 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
- Everson, J. S., Garner, S. A., Lambden, P. R., Fane, B. A., and Clarke, I. N. (2003). Host range of chlamydiaphages phiCPAR39 and Chp3. *J. Bacteriol.* 185, 6490–6492. doi: 10.1128/JB.185.21.6490-6492.2003
- Fuhrman, J. A. (1999). Marine viruses and their biogeochemical and ecological effects. *Nature* 399, 541–548. doi: 10.1038/21119
- Fuhrman, J. A., and Schwalbach, M. (2003). Viral influence on aquatic bacterial communities. *Biol. Bull.* 204, 192–195. doi: 10.2307/1543557
- Fuller, N. J., Wilson, W. H., Joint, I. R., and Mann, N. H. (1998). Occurrence of a sequence in marine cyanophages similar to that of T4 g20 and its application to PCR-based detection and quantification techniques. *Appl. Environ. Microb.* 64, 2051–2060.
- 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
- 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
- Gordon, D. (2003). Viewing and editing assembled sequences using Consed. *Curr. Protoc. Bioinform.* 11, 11.2. doi: 10.1002/0471250953.bi1102s02
- 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
- Hambly, E., and Suttle, C. A. (2005). The virosphere, diversity, and genetic exchange within phage communities. *Curr. Opin. Microbiol.* 8, 444–450. doi: 10.1016/j.mib.2005.06.005
- Hambly, E., Tétart, F., Desplats, C., Wilson, W. H., Krisch, H. M., and Mann, N. H. (2001). A conserved genetic module that encodes the major virion components in both the coliphage T4 and the marine cyanophage S-PM2. *Proc. Natl. Acad. Sci. U.S.A.* 98, 11411–11416. doi: 10.1073/pnas.191174498
- Hennes, K. P., Suttle, C. A., and Chan, A. M. (1995). Fluorescently labeled virus probes show that natural virus populations can control the structure of marine microbial communities. *Appl. Environ. Microb.* 61, 3623–3627.
- Holmfeldt, K., Odić, D., Sullivan, M. B., Middelboe, M., and Riemann, L. (2012). Cultivated single-stranded DNA phages that infect marine Bacteroidetes prove difficult to detect with DNA-binding stains. *Appl. Environ. Microb.* 78, 892–894. doi: 10.1128/AEM.06580-11
- Holmfeldt, K., Solonenko, N., Shah, M., Corrier, K., Riemann, L., Verberkmoes, N. C., et al. (2013). Twelve previously unknown phage genera are ubiquitous in global oceans. *Proc. Natl. Acad. Sci. U.S.A.* 110, 12798–12803. doi: 10.1073/pnas.1305956110
- Huelsenbeck, J. P., and Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17, 754–755. doi: 10.1093/bioinformatics/17.8.754
- 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
- Kenzaka, T., Tani, K., and Nasu, M. (2010). High-frequency phage-mediated gene transfer in freshwater environments determined at single-cell level. *ISME J.* 4, 648–659. doi: 10.1038/ismej.2009.145
- Kim, K.-H., Chang, H.-W., Nam, Y.-D., Roh, S. W., Kim, M.-S., Sung, Y., et al. (2008). Amplification of uncultured single-stranded DNA viruses from rice paddy soil. *Appl. Environ. Microb.* 74, 5975–5985. doi: 10.1128/AEM.01275-08
- 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. Microb.* 75, 3634–3640. doi: 10.1128/AEM.02317-08
- Labonté, J. M., and Suttle, C. A. (2013). Previously unknown and highly divergent ssDNA viruses populate the oceans. *ISME J.* 7, 2169–2177. doi: 10.1038/ismej.2013.110
- 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 Alcamí, A. (2009). High diversity of the viral community from an Antarctic lake. *Science* 326, 858–861. doi: 10.1126/science.1179287
- Middelboe, M., Hagström, A., Blackburn, N., Sinn, B., Fischer, U., Borch, N. H., et al. (2001). Effects of bacteriophages on the population dynamics of four strains of pelagic marine bacteria. *Microb. Ecol.* 42, 395–406. doi: 10.1007/s00248-001-0012-1
- Minot, S., Sinha, R., Chen, J., Li, H., Keilbaugh, S. A., Wu, G. D., et al. (2011). The human gut virome: inter-individual variation and dynamic response to diet. *Genome Res.* 21, 1616–1625. doi: 10.1101/gr.122705.111
- Pavesi, A. (2006). Origin and evolution of overlapping genes in the family Microviridae. *J. Gen. Virol.* 87, 1013–1017. doi: 10.1099/vir.0.81375-0
- Pedulla, M. L., Ford, M. E., Houtz, J. M., Karthikeyan, T., Wadsworth, C., Lewis, J. A., et al. (2003). Origins of highly mosaic mycobacteriophage genomes. *Cell* 113, 171–182. doi: 10.1016/S0092-8674(03)00233-2
- Rokyta, D., and Burch, C. (2006). Horizontal gene transfer and the evolution of microvirid coliphage genomes. *J. Bacteriol.* 188, 1134–1142. doi: 10.1128/JB.188.3.1134-1142.2006
- 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., Debroas, 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
- Rusch, D. B., Halpern, A. L., Sutton, G., Heidelberg, K. B., Williamson, S., Yooshef, S., et al. (2007). The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol.* 5:e77. doi: 10.1371/journal.pbio.0050077
- Salim, O., Skilton, R. J., Lambden, P. R., Fane, B. A., and Clarke, I. N. (2008). Behind the chlamydial cloak: the replication cycle of chlamydiophage Chp2, revealed. *Virology* 377, 440–445. doi: 10.1016/j.virol.2008.05.001
- 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. Microb.* 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. Env. Microb.* 68, 1290–1296. doi: 10.1128/AEM.68.3.1290-1296.2002
- Stepanauskas, R. (2012). Single cell genomics: an individual look at microbes. *Curr. Opin. Microbiol.* 15, 613–620. doi: 10.1016/j.mib.2012.09.001
- 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
- Suttle, C. A. (2005). Viruses in the sea. *Nature* 437, 356–361. doi: 10.1038/nature04160
- 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. Env. Microb.* 57, 721–726.
- Talavera, G., and Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* 56, 564–577. doi: 10.1080/10635150701472164
- Thingstad, T. F. (2000). Elements of a theory for the mechanisms controlling abundance, diversity, and role of lytic bacterial viruses in aquatic systems biogeochemical. *Limnology* 45, 1320–1328. doi: 10.4319/lo.2000.45.6.1320
- 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
- Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., et al. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science*. 304, 66–74. doi: 10.1126/science.1093857
- Weinbauer, M. G. (2004). Ecology of prokaryotic viruses. *FEMS Microbiol. Rev.* 28, 127–181. doi: 10.1016/j.femsre.2003.08.001
- Wilhelm, S. W., and Suttle, C. A. (1999). Viruses and nutrient cycles in the sea. *Bioscience* 49, 781–788. doi: 10.2307/1313569
- Winter, C., Bouvier, T., Weinbauer, M. G., and Thingstad, T. F. (2010). Trade-offs between competition and defense specialists among unicellular planktonic organisms: the “killing the winner” hypothesis revisited. *Microbiol. Mol. Biol. R* 74, 42–57. doi: 10.1128/MMBR.00034-09
- Wommack, K. E., and Colwell, R. R. (2000). Virioplankton: viruses in aquatic ecosystems. *Microbiol. Mol. Biol. R* 64, 69–114. doi: 10.1128/MMBR.64.1.69-114.2000
- 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. *Env. Microbiol.* 12, 172–191. doi: 10.1111/j.1462-2920.2009.02058.x
- Zhang, T., Breitbart, M., Lee, W. H., Run, J.-Q., Wei, C. L., Soh, S. W. L., et al. (2006). RNA viral community in human feces: prevalence of plant pathogenic viruses. *PLoS Biol.* 4:e3. doi: 10.1371/journal.pbio.0040003

**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.

Received: 12 October 2013; accepted: 06 December 2013; published online: 24 December 2013.

Citation: Labonté JM and Suttle CA (2013) 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

This article was submitted to *Evolutionary and Genomic Microbiology*, a section of the journal *Frontiers in Microbiology*.

Copyright © 2013 Labonté 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.