APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Feb. 2009, p. 991–997 Vol. 75, No. 4
0099-2240/09/$08.00+0 doi:10.1128/AEM.02024-08

Identification of Freshwater *Phycodnaviridae* and Their Potential Phytoplankton Hosts, Using DNA *pol* Sequence Fragments and a Genetic-Distance Analysis▽†

Jessica L. Clasen[1]‡ and Curtis A. Suttle[2]*

Department of Earth and Ocean Sciences, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z4[1], and Departments of Earth and Ocean Sciences, Botany, and Microbiology and Immunology, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z4[2]

Received 1 September 2008/Accepted 11 December 2008

Viruses that infect phytoplankton are an important component of aquatic ecosystems, yet in lakes they remain largely unstudied. In order to investigate viruses (*Phycodnaviridae*) infecting eukaryotic phytoplankton in lakes and to estimate the number of potential host species, samples were collected from four lakes at the Experimental Lakes Area in Ontario, Canada, during the ice-free period (mid-May to mid-October) of 2004. From each lake, *Phycodnaviridae* DNA polymerase (*pol*) gene fragments were amplified using algal-virus-specific primers and separated by denaturing gradient gel electrophoresis; 20 bands were extracted from the gels and sequenced. Phylogenetic analysis indicated that freshwater environmental phycodnavirus sequences belong to distinct phylogenetic groups. An analysis of the genetic distances "within" and "between" monophyletic groups of phycodnavirus isolates indicated that DNA *pol* sequences that differed by more than 7% at the inferred amino acid level were from viruses that infect different host species. Application of this threshold to phylogenies of environmental sequences indicated that the DNA *pol* sequences from these lakes came from viruses that infect at least nine different phytoplankton species. A multivariate statistical analysis suggested that potential freshwater hosts included *Mallomonas* sp., *Monoraphidium* sp., and *Cyclotella* sp. This approach should help to unravel the relationships between viruses in the environment and the phytoplankton hosts they infect.

Since the "discovery" of the high abundances of viruses in oceans (1), a growing body of research has demonstrated that viruses are dynamic members of aquatic environments (3, 22, 24, 31). Viruses infecting phytoplankton are of particular ecological importance and have been implicated in bloom termination (2, 3, 13), changing community composition (5), and nutrient cycling (12, 25, 28).

Many viruses infecting eukaryotic phytoplankton are members of the *Phycodnaviridae* (3), a family of large (100 to 220 nm in diameter), polyhedral, double-stranded DNA viruses (29). Exploration of the genetic richness of *Phycodnaviridae* in environmental samples has employed PCR-based methods and degenerate primers (algal-virus-specific 1 and 2 [AVS-1 and AVS-2]) that amplify α-like DNA polymerase (*pol*) gene fragments (6, 7, 18). Combining this method with denaturing gradient gel electrophoresis (PCR-DGGE) and phylogenetic analysis has shown that some closely related phycodnaviruses are cosmopolitan in distribution (19–21), while others appear to be restricted to specific environments (18). Moreover, phylogenetic analysis has demonstrated that *Phycodnaviridae* infecting the same species are typically much more closely related than those that infect different species (22).

In our study, we used PCR-DGGE and phylogenetic analysis to examine the genetic richness and relationships among *Phycodnaviridae* within several lakes in the Experimental Lakes Area (ELA) of western Ontario, Canada. These data were analyzed using genetic distance and multivariate statistics to infer the number and identities of potential host species.

MATERIALS AND METHODS

Sampling and virioplankton concentration. Lakes 224, 227, 239, and 240 in the ELA, Ontario, Canada (10), were sampled during the ice-free period of 2004 (Table 1) by submerging (0.5 m) and filling a prerinsed (10% HCl, followed by lake water) 20-liter polyethylene carboy. Each time, an integrated water sampler (16) was used to collect water from the euphotic zone (0 to 16 m) to determine the phytoplankton community composition.

Ultrafiltration through a 10-kDa-cutoff tangential-flow cartridge (S1Y10; Millipore, Billerica, MA) was used to concentrate viruses ~80-fold from 18 liters of 0.45-μm-filtered (142-mm-diameter polyvinylidine difluoride Durapore filters; Millipore) lake water (26). The concentrates were stored at 4°C in the dark until they were processed within 2 to 7 months.

PCR-DGGE. Viral DNA was extracted from each concentrate using the MoBio (Carlsbad, CA) Ultra Clean Soil DNA extraction kit and stored frozen until it was used for PCR. The volume of concentrate extracted was standardized to 50 ml of lake water.

DNA polymerase gene (*pol*) fragments from *Phycodnaviridae* were amplified using two rounds of PCR with AVS-1 and -2 primers (6, 7). In the first round, 5 μl of DNA template was added to a 45-μl PCR mixture containing 5.0 μl of 10× PCR buffer, 1.5 μl of 50 mM $MgCl_2$, 1.0 μl of each of the 2.0 mM deoxynucleotide triphosphates, 1.0 μl of 10 nM AVS-1, 3.0 μl of 10 nM AVS-2, 0.625 U of Platinum *Taq* DNA polymerase (Invitrogen, Burlington, ON, Canada), and water. The negative control was prepared as described above but without a DNA template, while *Micromonas pusilla* virus SP1 (MpV-SP1) was used as a positive

* Corresponding author. Mailing address: Rm. 1461, 6270 University Blvd., Vancouver, BC V6T 1Z4, Canada. Phone: (604) 822-9652. Fax: (604) 822-5558. E-mail: csuttle@eos.ubc.ca.

‡ Present address: Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92697.

† Supplemental material for this article may be found at http://aem.asm.org/.

▽ Published ahead of print on 16 December 2008.

TABLE 1. Environmental parameters for samples from the ELA

Lake no.	Date (2004)	Depth (m)	Temp (°C)	In situ light (μmol of photons m^{-2} s^{-1})
224	20 July	0.5	22.5	728
227	2 September	0.5	15.9	300
239	4 June	0.5	14.1	720
239	16 June	0.5	15.6	716
239	1 July	0.5	17.8	742
239	15 July	0.5	22.0	210
239	26 July	0.5	21.2	127
239[a]	May to October	0.5	7.8–22	22–742
240	15 June	0.5	15.5	412
240	29 June	0.5	17.0	600
240	13 July	0.5	21.3	26
240	10 August	0.5	18.4	164
240	24 August	0.5	15.1	42

[a] From composite Lake 239 viral concentrate; therefore, ranges of sample dates, temperatures, and in situ light levels are given.

control. The PCR conditions were as follows: initial denaturation at 95°C for 90 s, followed by 40 cycles of denaturation at 95°C for 45 s, annealing at 45°C for 45 s, and extension at 72°C for 45 s, and a final extension at 72°C for 7 min. To confirm amplification, 10 μl of PCR product and 2 μl of 6× loading buffer were loaded onto a 1.5% agarose gel flanked by a 100-bp ladder (Invitrogen). The gel was run at 90 V for 60 min in Tris-borate-EDTA buffer (0.5×), stained with ethidium bromide, and viewed with a gel documentation system (AlphaImager 3400; Alpha Innotech, San Leandro, CA). The bands (~800 bp) were plugged with a clean Pasteur pipette and placed in sterile 0.5-ml microcentrifuge tubes. DNA was eluted by adding 100 μl of 1× Tris-acetate-EDTA (TAE) and heating it to 65°C for 60 min. Two microliters of the eluted DNA was used in a second-round PCR prepared as described above, except the number of cycles was reduced to 26. Positive and negative controls were prepared as before; however, the negative control from the first-round reaction was also plugged and included as an internal control in the second-round reaction. The second-round PCR products were stored at −20°C.

TABLE 2. *Phycodnaviridae* isolates[a]

Virus abbreviation	Description	Strain	GenBank accession no.
MpV	*M. pusilla* virus	**PL1**	**U32982**
		SP1	U32975
		SP2	U32976
PBCV	*P. bursaria Chlorella* virus infecting *Chlorella*-like strain NC64a	**PBCV1**	**AJ890364**
		SC1B	AF344238
		AR93.2	AF344203
		NYb1	AF344234
		CH57	AF344210
		AL1A	AF344198
		CA4B	AF344209
		SH6A	AF344239
		NYs1	AF344235
		AR158	AF344202
		NY2A	**AF344230**
	P. bursaria Chlorella virus infecting *Chlorella*-like strain Pbi	CWM1	AF344214
		CVB1	AF344212
		CVR1	AF344215
		CVA1	**AF344211**
PgV	*P. globosa* virus	**3T**	**AY345136**
		4T	AY345137
		6T	AY345139
		10T	AY345142
		5T	AY345138
		7T	AY345140
CbV	*C. breviflum* virus	**PW3**	**U32984**
		PW1	U32983
EhV	*E. huxleyi* virus	**EHV**	**U42580**
		208	AF453867
FsV	*Feldmannia* sp. virus	**FsV**	**AF013260**
ESV	*E. siliculousus* virus	**ESV**	**AF204951**

[a] Sequences of DNA polymerase (*pol*) from known *Phycodnaviridae* isolates were accessed through GenBank. Boldface sequences were used in the environmental phylogeny (Fig. 2).

TABLE 3. Environmental samples[a]

Sample prefix	Location	Reference	GenBank accession no.
L224	Lake 224, ELA, ON, Canada	This study	EU408225 to EU408244
L227	Lake 227, ELA, ON, Canada	This study	EU408225 to EU408244
L239	Lake 239, ELA, ON, Canada	This study	EU408225 to EU408244
L240	Lake 240 , ELA, ON, Canada	This study	EU408225 to EU408244
CL1	Crawford Lake, ON, Canada	18	EU336433 to EU336476
LO1	Lake Ontario, ON, Canada	18	EU336573 to EU336707
CR	Chatfield Reservoir, CO	18	EU336477 to EU336572
SPR	South Platte River, CO	18	EU336708 to EU336803
BS	Barkley Sound, BC, Canada	19	AF405572 to AF405604
ESO2	Marine aerosols, East Sea, Korea	Unpublished[b]	AY436587 to AY436589
JPavs	Jericho Pier, BC, Canada	20	AY145089 to AY145098
MI	Malaspina Inlet, BC, Canada	19	AF405572 to AF405604
OTU	Gulf of Mexico, TX	7	U36931 to U36935
PS	Pendrell Sound, BC, Canada	19	AF405572 to AF405604
SI	Salmon Inlet, BC, Canada	19	AF405572 to AF405604
SO	Southern ocean	19	AF405572 to AF405604
GOS	Halifax to Galapagos Islands	15	Multiple[c]

[a] Freshwater DNA polymerase gene (*pol*) sequences were obtained from samples collected from four lakes at the ELA (L224, L227, L239, and L240). Other environmental DNA *pol* sequences from lakes in Colorado and Ontario, as well as the Gulf of Mexico and the Pacific and southern oceans, were accessed through GenBank. Metagenomic data from the GOS was accessed through the NCBI environmental database; GOS locations are not identified, since sequences often came from several different locations (http://camera.calit2.net/index.php).

[b] B. C. Cho, G. Park, D. H. Choi, and C. Y. Hwang, 2003.

[c] AACY020457048, AACY020716371, AACY020076678, AACY020168926, AACY020008685, AACY023220264, AACY022626042, AACY020325924, AACY021388666, AACY020002798, AACY020040066, AACY020013999, AACY020462121, AACY021701173, AACY021524699, AACY022625378, AACY023197486, AACY023235187, AACY020017444, AACY020013609, AACY020006926, AACY020028759, AACY020011590, AACY020009009, AACY020010354, AACY021532305, AACY020038007, AACY022638056, AACY023984140, AACY020034483, AACY023336016, AACY023234611, AACY020069902, AACY020070781, and AACY023362373.

DGGE was used to separate the AVS-amplified products (21). Second-round PCR products (40 μl) and 6× loading buffer (10 μl) were loaded onto a 6 to 7% polyacrylamide gel, which had a 20 to 40% gradient of denaturant (100% denaturant is defined as 7 M urea and 40% deionized formamide). Samples were run at 60 V for 15 h in a 60°C 1× TAE buffer, using a D-code electrophoresis system (Bio-Rad, Hercules, CA). Upon completion, the gel was stained in a 1× SYBR green I solution (Invitrogen) for >3 h and then visualized and photographed with a gel documentation system.

Cloning and sequencing. Several randomly selected bands from each lake sample were excised from the DGGE gel with a Pasteur pipette and placed in sterile microcentrifuge tubes, and the DNA was eluted by adding 100 μl of 1× TAE and heating it to 95°C for 15 min. Two microliters of the eluted DNA was used as the template in a PCR with AVS primers under the conditions described above for the second-round AVS PCR. The amplified DNA was purified (MinElute PCR purification kit; Qiagen, Germantown, MD), ligated, and transformed into *Escherichia coli* strain TOP10 using a TOPO TA Cloning kit (Invitrogen), and following an overnight incubation (37°C), the inserted *pol* fragments were amplified according to the manufacturer's recommendations. The amplified PCR products (~800 bp) were cleaned with a MinElute PCR purification kit, diluted, and sequenced using Applied Biosystems BioDye v3.1 Terminator Chemistry (Applied Biosystems, Foster City, CA) at the University of British Columbia's Nucleic Acid and Proteins Services Facility.

Sequence analysis. Phylogenetic analysis was used to compare phycodnavirus DNA *pol* sequences from the ELA lakes ($n = 20$) with those from other freshwater phycodnaviruses ($n = 24$) (18) and those from several isolates ($n = 28$) (Tables 2 and 3), while a second analysis included environmental phycodnavirus sequences from the ELA lakes ($n = 20$), other lakes ($n = 6$), and oceans ($n =$

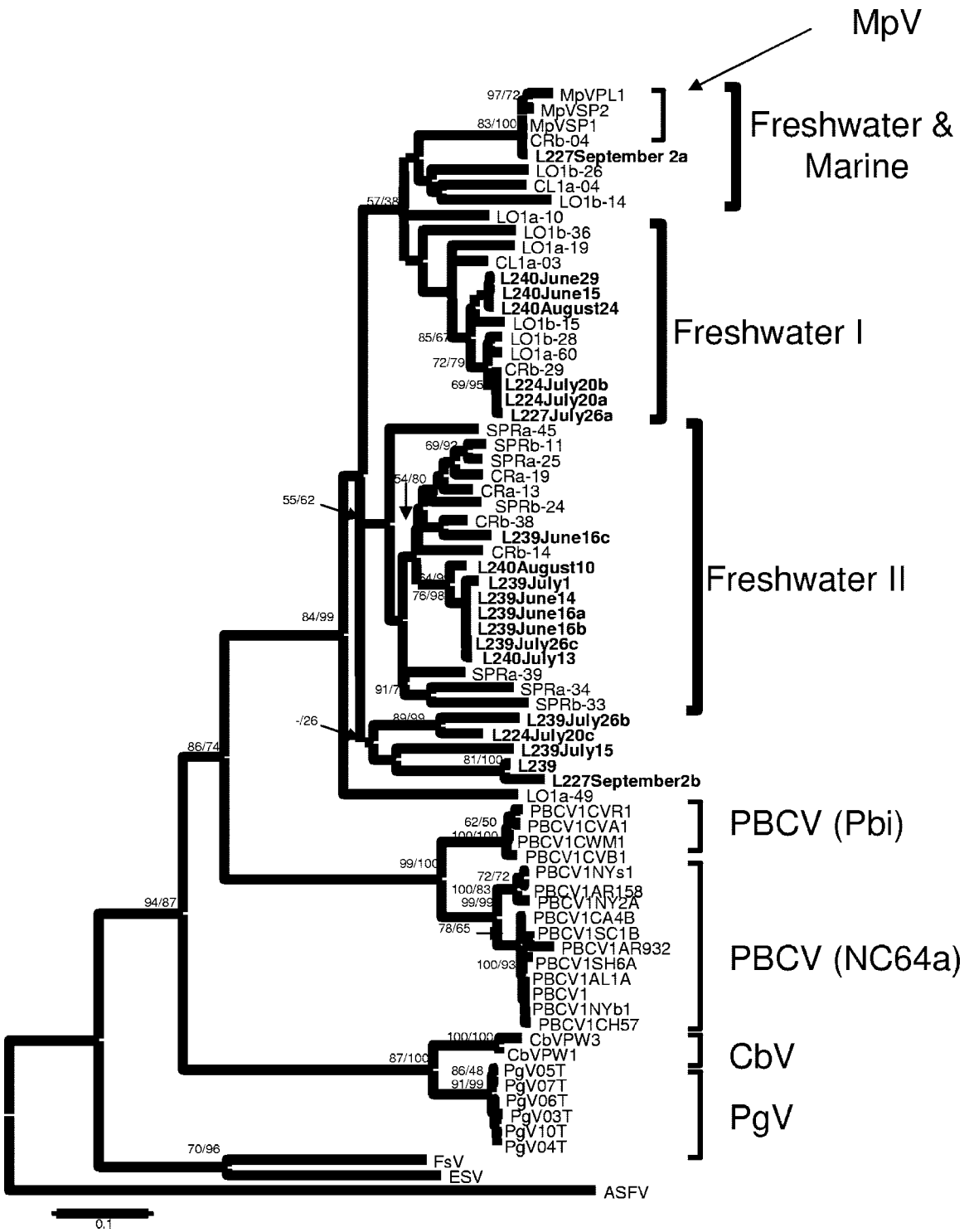
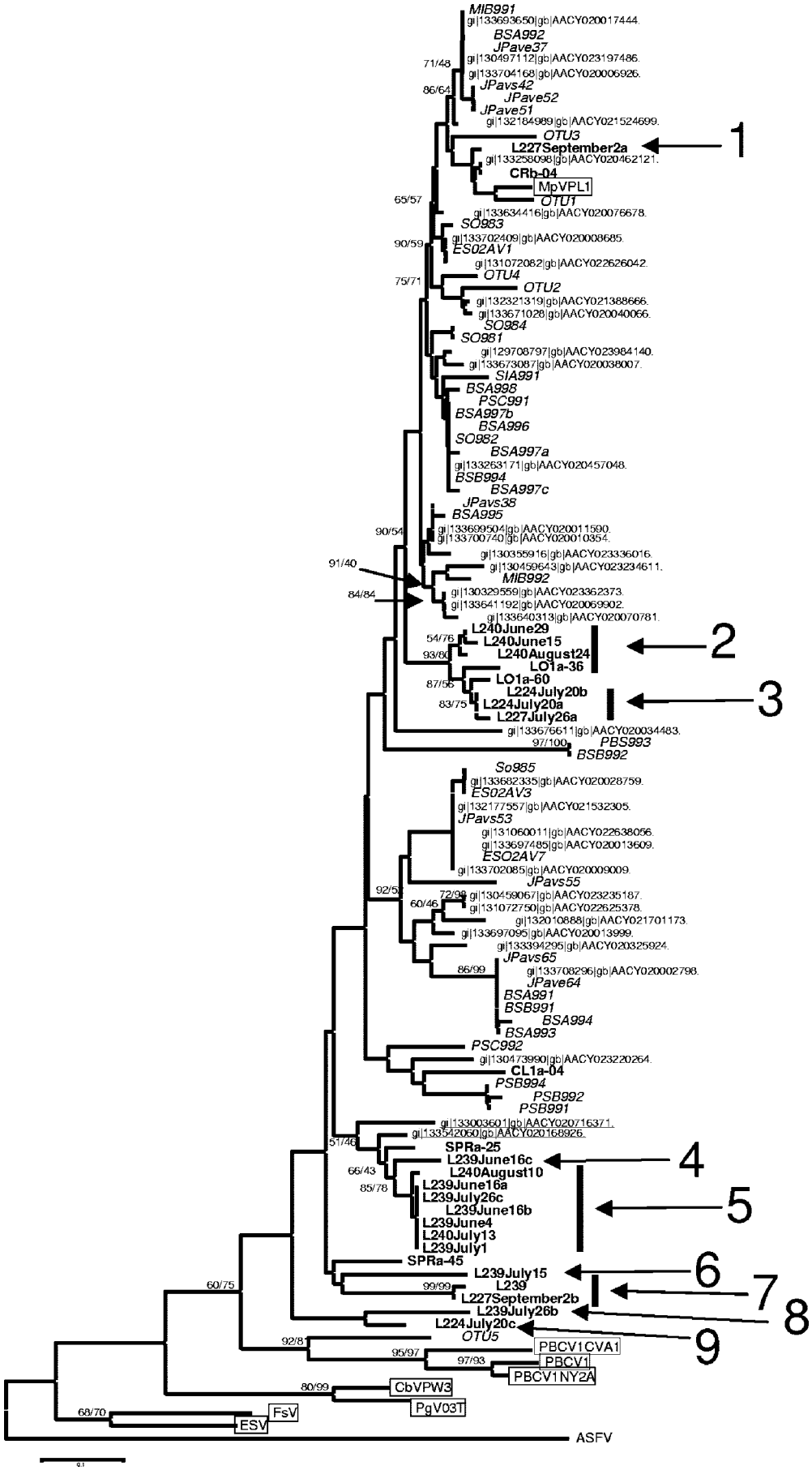


FIG. 1. NJ tree of freshwater phycodnavirus inferred amino acid *pol* sequences. NJ bootstrap values (10,000 replicates) and ML support values of >50 are indicated at the nodes (ML/NJ). The three freshwater clades identified by Short and Short (18) are indicated. Boldface annotations indicate the lake and collection date. Groups of phycodnavirus isolates are indicated on the right of the tree (Table 2). FsV, *Feldmannia* sp. virus; ESV, *Ectocarpus siliculousus* virus; ASFV, African swine fever virus (*Asfarviridae*). The scale bar represents the number of amino acid substitutions per residue.

44), including the top 20 matches from the Global Ocean Survey (GOS) data (15) to each of the 20 ELA lake sequences ($n = 35$) (Tables 2 and 3).

For each phylogeny, translated sequences were aligned in CLUSTAL X and refined by eye in BioEdit. The edited alignments were used to construct bootstrapped neighbor-joining (NJ) trees (10,000 replicates) in MEGA version 4 and quartet-puzzling maximum likelihood (ML) trees in Tree Puzzle (v. 5.2) with African swine fever virus as the outgroup. Phylogenetic trees were drawn using TreeView (v. 1.6.6) and MEGA version 4.



Specificity of AVS primers. To determine if AVS-amplified gene fragments represent the overall richness of the *Phycodnaviridae*, DNA *pol* sequences from the phycodnaviruses *Paramecium bursaria Chlorella* virus 1 (PBCV-1) and *Emiliana huxleyi* virus 86 (EhV) (accession numbers U42580 and AJ890364, respectively) were queried against the GOS data (15). As the virus and GOS sequences were obtained without AVS amplification, they were free of primer bias. The top 20 matches to PBCV and EhV that contained the conserved motif YGDTDS (Asp-Asp), along with sequences from the isolates, were phylogenetically analyzed (as described above) to determine if the tree topology was the same for amplified *pol* sequences as for sequences obtained from whole genomes and environmental shotgun libraries.

Inferring the number of potential hosts. An ML tree of phycodnavirus isolates was generated with sequences from isolates of PBCV, EhV, MpV, *Phaeocystis globosa* virus (PgV), and *Chrysochromulina brevifilum* virus (CbV) (Table 2). The maximum genetic distances between sequences from viruses infecting the same host species (within), and the minimum genetic distances between sequences from viruses infecting different host species (between) were determined. A discriminant analysis of the genetic-distance data was used to predict the likelihood that an "unknown" sample belonged in a particular group and therefore measured the robustness of the grouping criteria (SYSTAT v. 11). Finally, the within- and between-group means, standard errors, standard deviations, and 95% upper and lower confidence intervals were calculated and compared (by analysis of variance [ANOVA]) to determine the threshold genetic distance separating viruses infecting the same species (SYSTAT v. 11). Sequences separated by distances larger than this threshold value were assumed to have originated from viruses that infect different phytoplankton species.

Inferring the identities of potential hosts using multivariate statistics. The eukaryotic phytoplankton community composition was determined on each sampling date by D. L. Findlay at the Freshwater Institute, Winnipeg, Manitoba, Canada. A portion (125 ml) of the integrated (0- to 16-m) water samples were fixed in Lugol's fixative, and 10-ml subsamples were allowed to settle for 24 h in Utermöhl settling chambers (27) as outlined by Nauwerck (14). The cells were counted and identified to species level using an inverted microscope with phase contrast. The composition of the phytoplankton community on each sampling date was converted into a presence/absence binary matrix.

Sequences from viruses thought to infect different hosts based upon the genetic-distance analysis were traced back to specific plugged DGGE bands. On each sampling date, the presence or absence of each of these bands was converted into a binary matrix using GelCompar II (Applied Maths). For the purpose of this analysis, each band was treated as a single genotype, although different sequences can have the same melting temperature and migrate to the same point on the gel.

A monotonic multiple-dimensional scaling (MDS) analysis with the presence/absence data for phytoplankton species and viral sequences from each lake was conducted to infer potential hosts (SYSTAT v. 11). Phytoplankton species and viral *pol* sequences that occupied the same space in an MDS analysis cooccurred 100% of the time (see Fig. S1 in the supplemental material). Although the cooccurrence of viral DNA *pol* fragments and phytoplankton species suggests that the two are associated, it does not necessarily mean that the cooccurring taxa are the viral hosts. Moreover, the analysis does not take into account time lags that may obscure relationships between viruses and phytoplankton.

Nucleotide sequence accession numbers. The sequences obtained in this study have been deposited in GenBank under accession numbers EU408225 through EU408244.

RESULTS AND DISCUSSION

Several significant results stem from the phylogenetic analysis of freshwater environmental *Phycodnaviridae* sequences. First, freshwater phycodnaviruses form groups that are largely distinct from both cultured isolates and their marine counterparts. Second, ~99% of the marine and freshwater environmental sequences were more closely related to viruses infecting *M. pusilla* than other *Phycodnaviridae* isolates. Third, genetic-distance analysis indicated that the ELA freshwater environmental sequences likely originated from viruses that infect at least nine different species of phytoplankton. The significance and ecological importance of these findings are discussed below.

Phylogenies of freshwater *pol* sequences. DNA *pol* sequences from the ELA DGGEs (see Fig. S2 in the supplemental material) clustered in several clades within the *Phycodnaviridae* that are distinct from isolate sequences (Fig. 1). The exception was L227September2a, which clustered with MpV-SP1. Seventy-five percent of the ELA sequences clustered within the previously identified groups (18) Freshwater and Marine and Freshwater I and II, while 25% fell outside these clades. There were no obvious temporal or spatial patterns among the ELA sequences, aside from Freshwater II, which was mostly comprised of sequences from Lake 239 (Fig. 1). Similar to previous results (18), the sequences were more closely related to those from viruses (MpV) infecting the marine phytoplankton *M. pusilla* than to sequences from the PBCV group of viruses, which infect freshwater *Chlorella*-like algae.

Phylogenetic analysis (Fig. 2) of AVS-amplified freshwater and marine environmental *pol* fragments from this study and others (7, 18–20) and GOS (15) data (Table 3) produced NJ and ML trees of similar topologies, with most freshwater and marine sequences clustering separately (Fig. 2). The only GOS data that clustered close to freshwater sequences were from Lake Gatun, Panama (15; http://camera.calit2.net/index.php). BLAST searches of the ELA sequences against Lake Gatun metagenomic data recovered only two *Phycodnaviridae* sequences, both of which clustered with freshwater sequences. This result, along with persuasive tree architecture (Fig. 2) (18), is consistent with most freshwater phycodnaviruses being genetically distinct from their marine counterparts.

Other studies have found that some freshwater phage structural gene sequences cluster into monophyletic groups (17). Similarly, cyanophage *psbA* genes, which encode a core photosynthetic protein, cluster into marine and freshwater clades (8). Together, these studies suggest limited genetic exchange between viruses in marine and fresh waters.

Specificity of the AVS primers. The AVS primers do not amplify some marine phycodnavirus isolates, including EhVs and *Herterosigma akashiwo* viruses (HaVs). This suggests that only a subset of *Phycodnaviridae* sequences are represented in

FIG. 2. NJ tree of environmental phycodnavirus inferred amino acid *pol* sequences. The tree includes sequences from some *Phycodnaviridae* isolates, freshwater and marine environmental samples, and GOS BLAST hits (Tables 2 and 3). NJ bootstrap values (10,000 replicates) and corresponding ML support values of >50 are indicated at the nodes (ML/NJ). The ELA and other freshwater sequences (boldface), phycodnavirus isolates (boxed), and the African swine fever virus (ASFV) outgroup are as described in Table 2 and the legend to Fig. 1. Marine environmental sequences (italics) are from several locations, including the Gulf of Mexico and the Pacific and southern oceans (Table 3). The top 20 matches to each of the freshwater DNA *pol* sequences in the GOS are indicated by their accession numbers. The underlined sequences are the two GOS sequences that cluster nearest to freshwater sequences (see the text). The numbers indicate the potential host species determined from a genetic-distance analysis between viral *pol* sequences. Distances between AVS-amplified sequences of >0.081 amino acid substitution per residue were used to infer freshwater and marine host species. The scale bar represents the number of amino acid substitutions per residue. FsV, *Feldmannia* sp. virus; ESV, *Ectocarpus siliculousus* virus.

TABLE 4. Genetic distances within and between phycodnavirus groups[a]

Virus	Distance within or between:					
	CbV	PgV	EhV	PBCV (Pbi)	PBCV (NC64a)	MpV
CbV	**0.01846**					
PgV	0.19729	**0.01199**				
EhV	1.08917	1.0129	**0.01212**			
PBCV (Pbi)	1.21224	1.13597	1.13175	**0.03074**		
PBCV (NC64a)	1.612868	1.65970	1.65548	0.73471	**0.11830**	
MpV	1.03029	0.95402	0.94980	0.75331	1.27704	**0.03753**

[a] Distances either within (in boldface) a group of phycodnavirus isolates infecting the same host or between the different host groups were determined from branch lengths on an ML tree. Table 2 shows a description of the viral groups.

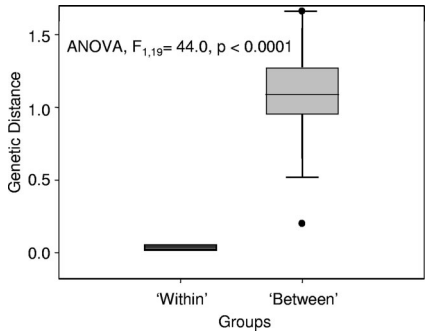


FIG. 3. Genetic distances within and between phycodnavirus groups. In the box plots of within- and between-host group distances, means, quartiles, and outliers are indicated, as are the ANOVA statistics from a comparison between the groups (Table 5).

data sets generated using these primers. The fact that 99% of the environmental *pol* sequences were more closely related to MpV than to other phycodnavirus isolates (Fig. 2) suggested that the primers might preferentially amplify MpV-like sequences. This was tested by BLASTing the same region of DNA *pol* from the complete genomes of EhV-86 and PBCV-1 (11, 29) against the GOS database. Not only did the EhV and PBCV *pol* sequences fall outside of the clade containing MpV (Fig. 1), they were also obtained without the use of the AVS primers. Moreover, as previously mentioned, EhVs do not amplify with the AVS primers. If the AVS amplicons are representative of the natural richness of the *Phycodnaviridae*, then the BLAST searches with the EhV and PBCV sequences should also be more closely related to MpV than to other isolates. All of the retrieved GOS sequences ($n = 20$) were most closely related to MpV (data not shown), indicating that primer bias was not responsible for AVS-amplified environmental sequences being most closely related to MpV.

Number of potential host species. Since phycodnavirus isolates infecting the same host species typically cluster in monophyletic groups (Fig. 1) (4, 6), the number of host species should be reflected in the number of discrete clades (23), which have a genetic distance greater than that which separates viruses infecting the same species. However, there is evidence that exceptions occur (30). Distances within and between monophyletic groups of phycodnavirus isolates were determined from branch lengths (Table 4), and as there was a significant difference between the within- and between-group distances (Fig. 3 and Table 5) (ANOVA, $F_{1,19} = 43.966$; $P < 0.0001$), both groups were used as predictors in a discriminant analysis. Overall, unknown distances were placed in the correct group 95% of the time; however, within-group distances were always placed correctly. As a result, the upper 95% confidence interval around the mean within-group distance (0.081 amino acid substitution per residue) was used as the threshold to distinguish among discrete clades of viruses. Therefore, if the distance between two viral sequences was <0.081 substitution per amino acid, the sequences were assumed to have originated from viruses that infect the same host species. Applying this threshold suggests that the environmental DNA *pol* sequences came from viruses infecting 13 different freshwater and 20 different marine hosts. The 20 ELA sequences are calculated to have come from viruses infecting nine different hosts (Fig. 2). The within-group distances are similar for all six isolate groups (Fig. 3); however, this distance may have to be adjusted as more sequences from isolates become available. An exception to this relationship is PgV-102P, which reportedly clusters with CbVs rather than other PgVs (30), suggesting that lateral gene transfer of a polymerase gene has occurred or that this is an example of an expansion in host range for CbV.

Identities of potential hosts. The identities of potential phytoplankton hosts were inferred using multivariate statistics to assess cooccurrence between viral sequences and phytoplankton species. MDS analysis with phytoplankton and virus presence/absence data revealed that *Mallomonas* sp., *Chrysosphaerella* sp., *Monoraphidium* sp., *Synedra* sp., *Cyclotella* sp., *Trachelomonas* sp., and *Peridinium* sp. were potential phytoplankton hosts, representing species from several ecologically important groups, including dinoflagellates, chlorophytes, chrysophytes, and diatoms (phytoplankton data are available upon request from the Freshwater Institute, Winnipeg, Canada). In this study, the migration distances of DGGE bands and light microscopy were used to infer viral and phytoplankton taxa, respectively. With current technologies, the use of sequences will eliminate potential ambiguities caused by comigration of DGGE bands, while host isolation coupled with sequences will allow different phytoplankton species that are indistinguishable by light microscopy to be identified. Despite these caveats, combining genetic-distance analysis with viral and host presence/absence data and MDS provides a technique for inferring the number and identities of phytoplankton infected by specific groups of viruses. This approach should help unravel the interactions between hosts and viruses in nature.

TABLE 5. DA results and basic statistics from genetic-distance analysis[a]

Group	DA (%)	Mean	SD	SE	95% CI	
					Upper	Lower
Within	100	0.038	0.041	0.017	**0.081**	−0.004
Between	93	1.094	0.383	0.099	1.306	0.882

[a] SD, standard deviation; SE, standard error; upper and lower 95% CI, the upper and lower 95% confidence intervals around the mean from each group. The threshold value used in further analysis is in boldface.

Implications. The results of this study have evolutionary and ecological implications. Phylogenetic analyses place most freshwater and marine phycodnaviruses in discrete evolutionary groups. Even though most of these clusters are not statistically well supported, the fact that they are clustered by environment suggests that this interpretation is valid. Furthermore, distance and MDS analyses allowed the number and identities of host phytoplankton species to be inferred. Such information can focus research on particular phytoplankton species and help unravel the interactions between viruses and hosts and the impact viruses have on structuring phytoplankton communities.

ACKNOWLEDGMENTS

We gratefully acknowledge the staff of the ELA and the Freshwater Institute (Winnipeg, Manitoba, Canada), particularly Dave L. Findlay for his phytoplankton data. We thank past and present Suttle laboratory members for numerous helpful discussions, especially Matthias Fischer. We thank Caroline Chénard, Johan Vande Voorde, and Chris Payne for reviewing and editing drafts of the manuscript.

This work was supported by the Natural Sciences and Engineering Research Council of Canada through a graduate student fellowship to J.L.C. and a Discovery Grant to C.A.S., as well as by a grant from the Tula Foundation for sequencing and analysis.

REFERENCES

1. **Bergh, O., K. Y. Borsheim, G. Bratbak, and M. Heldal.** 1989. High abundance of viruses found in aquatic environments. Nature **340:**467–468.
2. **Bratbak, G., J. K. Egge, and M. Heldal.** 1993. Viral mortality of the marine alga *Emiliania huxleyi* (Haptophyceae) and termination of algal blooms. Mar. Ecol. Prog. Ser. **93:**39–48.
3. **Brussaard, C. P. D.** 2004. Viral control of phytoplankton populations—a review. J. Eukaryot. Microbiol. **51:**125–138.
4. **Brussaard, C. P. D., S. M. Short, C. M. Frederickson, and C. A. Suttle.** 2004. Isolation and phylogenetic analysis of novel viruses infecting the phytoplankton *Phaeocystis globosa* (Prymnesiophyceae). Appl. Environ. Microbiol. **70:** 3700–3705.
5. **Castberg, T., A. Lasen, R. A. Sandaa, C. P. D. Brussaard, J. K. Egge, M. Heldal, R. Thyrhaug, E. J. van Hannen, and G. Bratbak.** 2001. Microbial population dynamics and diversity during a bloom of the marine coccolithophorid *Emiliania huxleyi* (Haptophyta). Mar. Ecol. Prog. Ser. **221**:39–46.
6. **Chen, F., and C. A. Suttle.** 1995. Amplification of DNA polymerase gene fragments from viruses infecting microalgae. Appl. Environ. Microbiol. **61:** 1274–1278.
7. **Chen, F., C. A. Suttle, and S. M. Short.** 1996. Genetic diversity in marine algal virus communities as revealed by sequence analysis of DNA polymerase genes. Appl. Environ. Microbiol. **62:**2869–2874.
8. **Chenard, C., and C. A. Suttle.** 2008. Phylogenetic analysis of genes encoding photosynthesis proteins in cyanophage isolates and natural virus communities. Appl. Environ. Microbiol. **74:**5317–5324.
9. Reference deleted.
10. **Cleugh, T. R., and B. W. Hauser.** 1971. Results of initial survey of experimental lakes area, Northwestern Ontario. J. Fish. Res. Bd. Can. **28:**129–137.
11. **DeAngelis, P. L., W. Jing, M. V. Graves, D. E. Burbank, and J. L. VanEtten.** 1997. Hyaluronan synthase of *Chlorella* virus PBCV-1. Science **278:**1800–1803.
12. **Gobler, C. J., D. A. Hutchins, N. S. Fisher, E. M. Cosper, and S. A. Sanudo-Wilhelmy.** 1997. Release and bioavailability of C, N, P, Se and Fe following viral lysis of a marine chrysophyte. Limnol. Oceanogr. **42:**1492–1504.
13. **Nagasaki, K., M. Ando, S. Itakura, I. Imai, and Y. Ishida.** 1994. Viral mortality in the final stage of *Heterosigma akashiwo* (Raphidophyceae) red tide. J. Plankton Res. **16:**1595–1599.
14. **Nauwerck, A.** 1963. Die beziehungen zwischen zooplankton und phytoplankton in See Erken. Symb. Bot. Ups. **17:**1–163.
15. **Rusch, D. B., A. L. Halpern, G. Sutton, K. B. Heidelberg, S. Williamson, S. Yooseph, D. Wu, J. A. Eisen, J. M. Hoffman, K. Remington, K. Beeson, B. Tran, H. Smith, H. Baden-Tillson, C. Stewart, J. Thorpe, J. Freeman, C. Andrews-Pfannkoch, J. E. Venter, K. Li, S. Kravitz, J. F. Heidelberg, T. Utterback, Y.-H. Rogers, L. I. Falcón, V. Souza, G. Bonilla-Rosso, L. E. Eguiarte, D. M. Karl, S. Sathyendranath, T. Platt, E. Bermingham, V. Gallardo, G. Tamayo-Castillo, M. R. Ferrari, R. L. Strausberg, K. Nealson, R. Friedman, M. Frazier, and J. C. Venter.** 2007. The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. PLoS Biol. **5:**398–431.
16. **Shearer, J. A.** 1978. Two devices for obtaining water samples integrated over depth. Can. Fish. Mar. Serv. Tech. Rep. **772:**1–10.
17. **Short, C. M., and C. A. Suttle.** 2005. Nearly identical bacteriophage structural gene sequences are widely distributed in both marine and freshwater environments. Appl. Environ. Microbiol. **71:**480–486.
18. **Short, S. M., and C. M. Short.** 2008. Diversity of algal viruses in various North American freshwater environments. Aquat. Microb. Ecol. **51:**13–21.
19. **Short, S. M., and C. A. Suttle.** 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.
20. **Short, S. M., and C. A. Suttle.** 2003. Temporal dynamics of natural communities of marine algal viruses and eukaryotes Aquat. Microb. Ecol. **32:**107–119.
21. **Short, S. M., and C. A. Suttle.** 1999. Use of the polymerase chain reaction and denaturing gradient gel electrophoresis to study diversity in natural virus communities. Hydrobiologia **401:**19–32.
22. **Suttle, C. A.** 2000. Ecological, evolutionary, and geochemical consequences of viral infection of cyanobacteria and eukaryotic algae, p. 247–296. *In* C. J. Hurst (ed.), Viral ecology. Academic Press, London, United Kingdom.
23. **Suttle, C. A.** 2003. Viral diversity and its implications for infection in the sea, p. 73–76. *In* F. Briand (ed.), Ecology of marine viruses. CIESM Workshop Monographs 21. CIESM, Monaco.
24. **Suttle, C. A.** 2005. Viruses in the sea. Nature **437:**356–361.
25. **Suttle, C. A.** 2007. Marine viruses—major players in the global ecosystem. Nat. Rev. Microbiol. **5:**801–812.
26. **Suttle, C. A., A. M. Chan, and M. T. Cottrell.** 1991. Use of ultrafiltration to isolate viruses from seawater which are pathogens of marine phytoplankton. Appl. Environ. Microbiol. **57:**721–726.
27. **Utermöhl, H.** 1958. Zur Vervollkommnung der quantitativen Phytoplankton-Methodik. Mitt. Int. Verein. Theor. Angew. Limnol. **9:**1–38.
28. **Wilhelm, S. W., and C. A. Suttle.** 1999. Viruses and nutrient cycles in the sea. Bioscience **49:**781–788.
29. **Wilson, W. H., D. C. Schroeder, M. J. Allen, M. T. G. Holden, J. Parkhill, B. G. Barrell, C. Churcher, N. Harnlin, K. Mungall, H. Norbertczak, M. A. Quail, C. Price, E. Rabbinowitsch, D. Walker, M. Craigon, D. Roy, and P. Ghazal.** 2005. Complete genome sequence and lytic phase transcription profile of a Coccolithovirus. Science **309:**1090–1092.
30. **Wilson, W. H., D. C. Schroeder, J. Ho, and M. Canty.** 2006. Phylogenetic analysis of PgV-102P, a new virus from the English Channel that infects *Phaeocystis globosa*. J. Mar. Biol. Assoc. U.K. **86:**485–490.
31. **Wommack, K. E., and R. R. Colwell.** 2000. Virioplankton: viruses in aquatic ecosystems. Microbiol. Mol. Biol. Rev. **64:**69–114.