

## ORIGINAL ARTICLE

# Quantification of virus genes provides evidence for seed-bank populations of phycodnaviruses in Lake Ontario, Canada

Cindy M Short<sup>1</sup>, Oksana Rusanova<sup>2</sup> and Steven M Short<sup>1,2</sup>

<sup>1</sup>Department of Biology, University of Toronto Mississauga, Mississauga, Ontario, Canada and <sup>2</sup>Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, Ontario, Canada

Using quantitative PCR, the abundances of six phytoplankton viruses DNA polymerase (*polB*) gene fragments were estimated in water samples collected from Lake Ontario, Canada over 26 months. Four of the *polB* fragments were most related to marine prasinoviruses, while the other two were most closely related to cultivated chloroviruses. Two *Prasinovirus*-related genes reached peak abundances of >1000 copies ml<sup>-1</sup> and were considered 'high abundance', whereas the other two *Prasinovirus*-related genes peaked at abundances <1000 copies ml<sup>-1</sup> and were considered 'low abundance'. Of the genes related to chloroviruses, one peaked at ca 1600 copies ml<sup>-1</sup>, whereas the other reached only ca 300 copies ml<sup>-1</sup>. Despite these differences in peak abundance, the abundances of all genes monitored were lowest during the late fall, winter and early spring; during these months the high abundance genes persisted at 100–1000 copies ml<sup>-1</sup> while the low abundance *Prasinovirus*- and *Chlorovirus*-related genes persisted at fewer than ca 100 copies ml<sup>-1</sup>. Clone libraries of *psbA* genes from Lake Ontario revealed numerous *Chlorella*-like algae and two prasinophytes demonstrating the presence of candidate hosts for all types of viruses monitored. Our results corroborate recent metagenomic analyses that suggest that aquatic virus communities are composed of only a few abundant populations and many low abundance populations. Thus, we speculate that an ecologically important characteristic of phycodnavirus communities is seed-bank populations with members that can become numerically dominant when their host abundances reach appropriate levels.

The ISME Journal (2011) 5, 810–821; doi:10.1038/ismej.2010.183; published online 2 December 2010

**Subject Category:** microbial population and community ecology

**Keywords:** algal viruses; phycodnaviruses; phytoplankton; DNA polymerase; quantitative PCR; freshwater

## Introduction

Viruses that infect eukaryotic phytoplankton have been studied for several decades and at least 40 representatives infecting 11 phytoplankton species have been cultivated (Nagasaki and Bratbak, 2010). Although phytoplankton viruses are diverse in capsid size, and genome size and type, most cultivated algal viruses belong to the family *Phycodnaviridae*. Phycodnaviruses are nucleo-cytoplasmic large DNA viruses with capsid sizes ranging from 100–220 nm, and genomes ranging in length from 170 to 560 kb (Van Etten and Meints, 1999; Van Etten *et al.*, 2002; Dunigan *et al.*, 2006). The earliest effort to characterize phycodnavirus communities while circumventing difficulties associated with virus

cultivation involved designing algal-virus specific (AVS) PCR primers that targeted DNA polymerase (*polB*) genes (Chen *et al.*, 1996). Clone libraries of *polB* amplicons, and more recently amplicons of a gene encoding the major capsid protein (Larsen *et al.*, 2008), demonstrated that phycodnavirus communities are diverse, environmental sequences are distinct from cultivated phycodnavirus genes, and closely related sequences can be found in geographically distant locations (Short and Suttle, 2002; Larsen *et al.*, 2008; Short and Short, 2008; Clasen and Suttle, 2009). Although studies of phycodnavirus diversity have provided many insights into their ecology, they do not yield quantitative information to infer their influence on phytoplankton mortality and succession, or the flow of energy and material in aquatic systems.

An important but challenging aspect of algal virus ecology is understanding factors that influence temporal variations in virus and host abundances. Previous reports have documented 10-fold differences in the total abundance of aquatic virus communities between summer and winter months

Correspondence: SM Short, Department of Biology, University of Toronto Mississauga, 3359 Mississauga Road N, Mississauga, Ontario L5L 1C6, Canada.

E-mail: steven.short@utoronto.ca

Received 21 July 2010; revised 20 October 2010; accepted 31 October 2010; published online 2 December 2010

(Wommack and Colwell, 2000; Li and Dickie, 2001), and rapid changes in virus abundances in nutrient amended seawater enclosures (Castberg *et al.*, 2001; Larsen *et al.*, 2001). The abundances of viruses that infect single strains of phytoplankton, such as *Heterocapsa circularisquama* (Tomaru *et al.*, 2007), *Micromonas pusilla* (Cottrell and Suttle, 1995; Zingone *et al.*, 1999) and *Synechococcus* spp. (Suttle and Chan, 1994), vary seasonally and range from undetectable to  $>10^5$  infectious units  $\text{ml}^{-1}$ . Molecular fingerprinting techniques have also been used to study the temporal dynamics of aquatic viruses (for example, Wommack *et al.*, 1999a; Short and Suttle, 2003), and more recently, quantitative PCR (qPCR) was coupled with molecular fingerprinting to study cyanophage seasonality in Norwegian coastal waters (Sandaa and Larsen, 2006). Although these studies have all contributed to a fundamental understanding of the seasonality of aquatic viruses, few, if any, provide a detailed picture of the dynamics of non-cultivated virus taxa.

To study the intra-annual dynamics of non-cultivated phycodnaviruses in Lake Ontario, Canada, qPCR assays were developed. Initially, Lake Ontario algal virus community richness was examined by sequencing clone libraries of *polB* gene fragments (Short and Short, 2008). Using sequences from autumn clone libraries, qPCR was used to monitor the abundances of three operational taxonomic units (OTUs) for an entire year demonstrating that they achieved maximum abundances in the autumn months of 2007 and 2008, and two were persistent throughout the year while the other was ephemeral (Short and Short, 2009). Phylogenetic analysis of these three *polB* genes indicated that they were most closely related to phycodnaviruses that infect marine prasinophytes. As phycodnaviruses infecting the same host taxa typically cluster in monophyletic groups (Chen and Suttle, 1996; Brussaard *et al.*, 2004), it is plausible that the hosts of the viruses encoding these genes are prasinophytes or closely related phytoplankton. Surprisingly, while phytoplankton surveys of Lake Ontario noted the high diversity and abundance of chlorophyte phytoplankton, no prasinophytes were ever reported (Munawar and Munawar, 1982). It is possible that prasinophytes were present but missed because they were too small or too few for microscopy-based identification; many prasinophytes are picoplankton ( $<3\ \mu\text{m}$ ), and entire clades of unknown prasinophytes have been discovered recently via molecular surveys (Viprey *et al.*, 2008). The facts that nearly all phycodnavirus genes amplified from Lake Ontario are close relatives of prasinoviruses, yet no prasinophycean algae have been reported in Lake Ontario phytoplankton surveys suggests that knowledge of Lake Ontario phycodnaviruses and their hosts is extremely limited.

The aims of this study were to extend previous work by targeting a wider variety of phycodnavirus *polB* genes and to look for molecular signatures of

potential phycodnavirus hosts. More specifically, the goals were to: (1) monitor other phycodnavirus *polB* genes obtained from spring and summer water samples to determine if their seasonal timing differed from the previously studied *polB* genes, (2) extend the length of the time series by including an additional year to determine whether previously observed intra-annual fluctuations represent recurring patterns of abundance, (3) to determine whether a common feature of Lake Ontario phycodnaviruses is water-column persistence throughout the year and (4) to determine whether amplification of a photosynthesis marker gene (*psbA*) can be used to identify potential hosts for Lake Ontario phycodnaviruses.

## Materials and methods

### Sample collection

Water samples were collected at least twice a month during the spring, summer and fall months, and at least once a month during the winter from a jetty on Lake Ontario, Ontario, Canada ( $43^{\circ}32.614'N$ ,  $79^{\circ}34.995'W$ ) from 18 September 2007 to 26 October 2009. Within 1 h of collection, virus size fractions were isolated by gravity filtration through  $35\ \mu\text{m}$  Nitex mesh followed by vacuum filtration (100 mm Hg) through 47 mm diameter  $0.45\ \mu\text{m}$  pore-size Durapore PVDF membrane filters (Millipore, Billerica, MA, USA). For samples collected from 18 September 2007 to 22 January 2009, 72 ml of the filtrate was centrifuged at 182 000 g for 3 h in six 12 ml ultra-clear centrifuge tubes in a Beckman SW40Ti rotor (Beckman Coulter, Fullerton, CA, USA), and for samples collected from 13 February 2009 to 26 October 2009 the same volume of filtrate was centrifuged at 118 000 g for 3.5 h in two 36 ml polyallomer tubes in a Beckman SW32Ti rotor (Beckman Coulter). Importantly, using a conservative sedimentation coefficient of 65S for phycodnaviruses, the calculated time to pellet all phycodnaviruses with the SW40Ti and SW32Ti rotors would be *ca* 2.3 and 3.3 h, respectively; the estimated range of sedimentation coefficients for viruses is 42S to  $>1000\text{S}$  (Lawrence and Steward, 2010). Following ultracentrifugation, the pelleted material in each tube was left overnight at  $4^{\circ}\text{C}$  in 100 or  $300\ \mu\text{l}$  of 10 mM Tris-HCl (pH 8.0) for 12 ml or 36 ml tubes, respectively. The pelleted material in each tube was resuspended by pipetting and brief vortexing, and for each sample resuspended material from individual tubes was combined into a single screw-cap microcentrifuge tube and stored at  $-20^{\circ}\text{C}$ . The final volume of the concentrated virus size fraction was determined by weighing the screw cap microcentrifuge tubes before and after sample addition. Phytoplankton cells were collected by vacuum filtration (100 mm Hg) of 500 ml lake water onto a 47 mm GC50 ( $0.5\ \mu\text{m}$  nominal rating) glass fiber filters (Advantec MFS, Dublin, CA, USA) that were stored at  $-80^{\circ}\text{C}$  until DNA extraction.

**Table 1** Quantitative PCR primers and probes (5' to 3')

Target sequence	Forward primer	Reverse primer	TaqMan probe
LO.16jul08.14	ATACGAGTCGTTCCAAGT	CTTGCGTACTGTTTCAG	AGCAGGCTCGGCACATTC
LO.16jul08.20	CAGTTGGCTACAAGATT	CCTTCATGGTGACAGTTG	TGGAACGCAAGGCAACATACC
LO.20may09.33	GATACAGATTCGGTTATGGT	CATCTTGAAGTGTGCCTC	TGTCCACAGTTCCGTCCTCT
LO.08may08.08	GTAAGGGATTCTTGCCATGTGTTT	TCATTGATGTCTTCTTGGTTCTTCC	AGTCACGGCAACTGGTCGCAACAT

#### DNA extraction, PCR amplification, cloning and sequencing

Using previously described methods (Short and Short, 2008), virus DNA was extracted and *polB* gene fragments were PCR amplified from samples collected on 24 April and 16 July 2008 using a C1000 thermal cycler (Bio-Rad Laboratories, Hercules, CA, USA). Phytoplankton DNA was extracted from glass fiber filters from samples collected on 8 May, 19 June, 1 August, 28 August and 2 October 2008. Filters were cut into several pieces and were placed into two 2 ml screw cap vials that contained 0.25 g each of washed and sterilized 212–300 µm, and 425–600 µm glass beads. After the addition of 550 µl of buffer AP1 and 5.5 µl RNase A (Qiagen, Mississauga, ON, Canada), cells were disrupted in a Mini-Beadbeater-96 (BioSpec Products, Bartlesville, OK, USA) for a total of 6 min. The supernatant was removed and DNA was extracted using a DNeasy Plant Mini Kit (Qiagen) following the manufacturer's recommendations. Phytoplankton *psbA* gene fragments were PCR amplified using modified versions of previously published primers (Zeidner *et al.*, 2003) with the forward primer biased toward Chlorophyte algae (Supplementary Table 1); the forward and reverse primer sequences were *psbA-F*: 5'-GGIATGMGICCITGGATHGCGT-3'; and *psbA-R*: 5'-GGRAARTTRTGIGCRTTICKYTCRTGC-3', respectively. Thermal cycling was conducted on a Bio-Rad C1000 thermal cycler using previously published conditions (Zeidner *et al.*, 2003) except the annealing temperature was raised to 64 °C. PCR reactions were electrophoresed and amplified fragments were excised and extracted as previously described (Short and Short, 2008).

Gel-extracted amplicons were cloned using a pGEM-T Vector System II (Promega, Madison, WI, USA) as previously described (Short and Short, 2008), and purified plasmids were quantified using a NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). Plasmid inserts were sequenced at The Centre for Applied Genomics (TCAG) at The Hospital for Sick Children, Toronto, Ontario, Canada. Only full-length sequences without ambiguous base calls were used for further analysis. All sequences were examined using the chimera-checking web application Bellerophon, available at <http://comp-bio.anu.edu.au/bellerophon/bellerophon.pl> (Huber *et al.*, 2004).

#### Virus sequence analysis and primer and probe design

All 60 phycodnavirus *polB* sequences obtained in this study were compared with 135 previously published

sequences from 18 September to 10 October 2007 (Short and Short, 2008). Sequences from the 24 April to 16 July 2008 clone libraries were assigned GenBank accession numbers HM750157 to HM750180, and HM750181 to HM750214, respectively. All sequences were translated to amino acids and aligned using ClustalW (Thompson *et al.*, 1994) using default parameters in Mega 4.0 (Tamura *et al.*, 2007). Aligned amino acid sequences were used to generate nucleotide alignments, which were in turn compared via a neighbor-joining tree with the maximum composite likelihood substitution model in Mega. Two groups of nearly identical sequences (that is, >97% nucleotide identity) from the spring (24 April 2008) and summer (16 July 2008) clone libraries were each selected as targets for qPCR primer and probe design. Additionally, two other sequences (GenBank accession numbers HM629733 and HM629734) that were most closely related to cultivated chloroviruses, were also selected for qPCR primer and probe design. Primers and TaqMan probes for qPCR were designed using Beacon Designer 7.0 (Premier Biosoft International, Palo Alto, CA, USA) (Table 1), and probes were 5' labelled with FAM (6-carboxyfluorescein) and 3' labelled with Iowa Black FQ (Integrated DNA Technologies, Coralville, IA, USA).

Phylogenetic comparison with cultivated nucleocytoplasmic large DNA viruses was used to identify the gene fragments selected for qPCR. After intein and intron sequences were removed from certain *polB* genes as previously described (Short and Short, 2008), amino acid sequences were aligned and curated using MUSCLE (Edgar, 2004) and Gblocks (Castresana, 2000), respectively, with default parameters in Phylogeny.fr (Dereeper *et al.*, 2008). Phylogenies were recreated using both Bayesian inference and maximum likelihood. Bayesian inference was conducted using Mr Bayes 3.1.2 (Ronquist and Huelsenbeck, 2003) with two runs, four chains, 10<sup>6</sup> generations, sampling every 100 generations, a burnin value of 0.25, and mixed models of amino acid substitution. A maximum likelihood phylogeny was reconstructed using the approximate likelihood-ratio test in PhyML (Guindon and Gascuel, 2003) with branch support from the minimum of SH-like and Chi<sup>2</sup>-based values. Mega 4.0 and Adobe Illustrator CS (Adobe Systems, San Jose, CA, USA) were used for tree viewing and drawing.

#### Quantitative PCR

qPCR via the 5' nuclease assay was used to quantify six phycodnavirus *polB* fragments (two previously

described and four designed for this study, Table 1) in separate reactions as previously described (Short and Short, 2009), except reactions using the primers and probes designed for this study contained 5.0 mM MgCl<sub>2</sub>. For each set of reactions, eight 10-fold serially diluted standards (ranging from *ca* 3 × 10<sup>0</sup> to 3 × 10<sup>7</sup> molecules per reaction) were run in duplicate along with three no-template control reactions containing 2 µl nuclease-free water. Standards were cloned fragments of the target *polB* molecule that were linearized by restriction digest, purified by agarose gel electrophoresis, extracted from agarose gel slices with a QIAquick gel extraction kit (Qiagen), and quantified using a NanoDrop ND-1000 spectrophotometer. Probe and primer sets were tested for efficiency and specificity as previously described (Short and Short, 2009). Time series of target gene abundances were statistically analyzed using the autocorrelation function in IBM SPSS Statistics 17.0 (IBM Corp., Armonk, NY, USA); because samples could not be collected on a weekly basis during winter months, gene abundances were binned into 4 week blocks and averaged before autocorrelation analyses were conducted.

#### *Phytoplankton psbA sequence analysis*

Owing to the high level of conservation among *psbA* sequences, only nucleotides were compared. For each phytoplankton *psbA* clone library, percent coverage was determined using the calculation  $C = 1 - (N/n)$  where  $C$  was the homologous coverage,  $N$  was the number of singleton sequences (that is, nucleic acid sequences <97% identical to any other) and  $n$  was the total number of sequences in the sample. All 220 sequences (GenBank accession numbers HM629511 to HM629730) obtained from five clone libraries were pooled and representative sequences from unique OTUs (that is, singletons and groups of nucleic acid sequences >97% identical to each other) were selected for phylogenetic analysis via comparison with sequences from cultivated phytoplankton, phage and higher plants; sequence identity matrices were generated using BioEdit 7.0.5 (Hall, 1999). Nucleotides were aligned using MUSCLE with default parameters in Phylogeny.fr (Dereeper *et al.*, 2008) and a maximum likelihood phylogeny was reconstructed using the approximate likelihood-ratio test in PhyML (Guindon and Gascuel, 2003) with branch support from the minimum of SH-like and Chi<sup>2</sup>-based values. Mega 4.0 (Tamura *et al.*, 2007) and Adobe Illustrator CS were used for tree viewing and drawing, and sequence identity matrices and clone library composition were analyzed using Microsoft Excel 2007 (Microsoft Corp., Redmond, WA, USA).

## Results and discussion

This study demonstrated that freshwater algal viruses can have annually recurring patterns of abundance marked by variable durations of peak

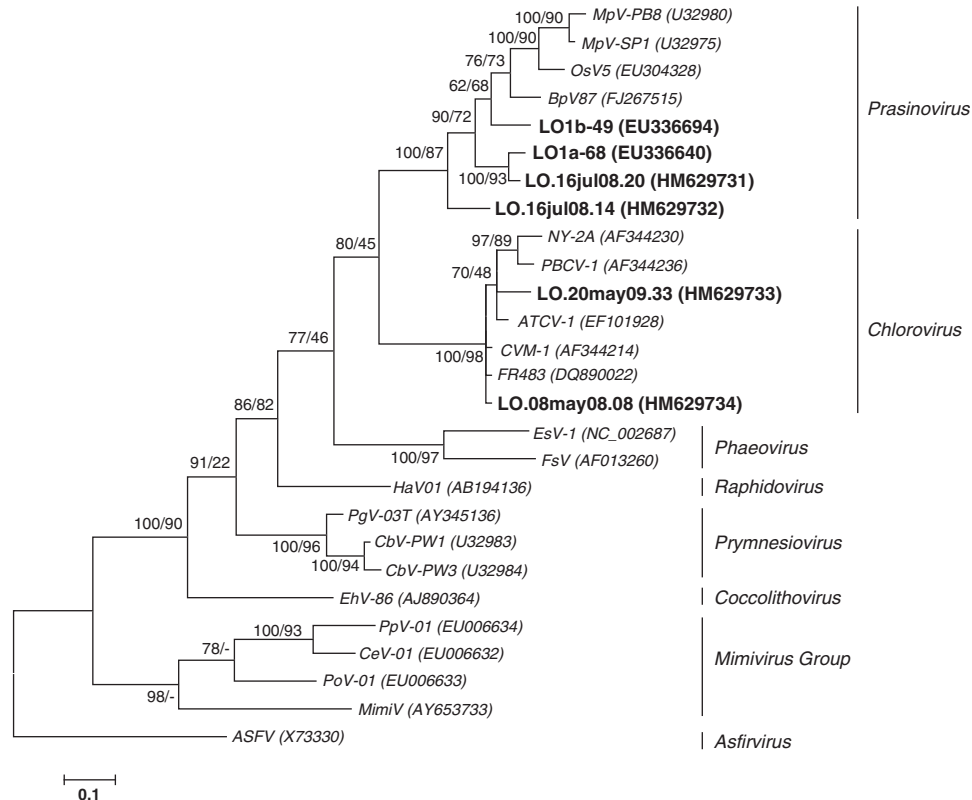
abundance followed by long periods of minimal abundance wherein individual virus types form a seed-bank population. Despite the fact that likely hosts for Lake Ontario phycodnaviruses observed were not identified in previous lake-wide phytoplankton surveys, molecular methods provided insights into the potential hosts of Lake Ontario phycodnaviruses. Before these results and their implications can be considered in more detail, technical aspects of this study must be considered.

Early in 2009, the centrifuge rotor used to concentrate viruses was changed. To determine whether this change lead to differences in sample concentration efficiency, a single water sample was collected and processed as described in the Materials and methods section, and equal volumes were concentrated simultaneously using each rotor. The samples were resuspended and extracted at the same time, and a single *polB* gene fragment (LO.16jul08.20) was quantified using qPCR. Predicted numbers of gene copies from triplicate amplifications was 6.71 × 10<sup>2</sup> (s.d. = 1.80 × 10<sup>2</sup>) for the SW40Ti rotor with ultraclear tubes, and 5.85 × 10<sup>2</sup> (s.d. = 1.03 × 10<sup>2</sup>) for the SW32Ti rotor with polyallomer tubes. A paired *T*-test demonstrated that this difference in mean gene copies was not statistically different ( $P = 0.52$ ). Moreover, although maximal numbers of *Prasinovirus polB* fragments were observed in 2008, the *Chlorovirus polB* fragments reached comparable or even higher abundances in 2009 (Figure 2) providing evidence that inter-annual differences in peak abundance were not merely the result of the different centrifuge rotors used in 2008 and 2009.

All sequences reported in this study were obtained from clone libraries of environmental PCR products and therefore, it was possible that some unique sequence types were chimeric artefacts. Of the 195 virus *polB* sequences used for qPCR primer and probe design, only two sequences (LO1b-25 and LO.16jul08.24) were identified as possible chimeric molecules and these sequences were not considered for qPCR primer and probe design. Of the 220 *psbA* sequences obtained in this study, only the following four were flagged as possible chimeras: LO.01aug08.97 (OTU 'AA'), LO.01aug08.63 (OTU 'W'), and LO.01aug08.80 and LO.01aug08.50 (OTU 'N') (Figures 3 and 4).

#### *polB identification and qPCR*

Phylogenetic analyses of AVS-amplified *polB* sequences from spring, summer and fall clone libraries revealed that, with the exception of a single sequence, they were all most closely related to cultivated prasinoviruses when compared with other phycodnaviruses. These results are consistent with those from previous studies demonstrating that all, or nearly all environmental sequences obtained with the AVS-1 and 2 primers cluster with *polB* genes from marine prasinoviruses (for example, Chen *et al.*, 1996; Short and Suttle, 2002; Short and Short,



**Figure 1** DNA polymerase (*polB*) phylogeny based on inferred amino acid sequences. The phylogeny is based on the alignment of 82 homologous amino acid positions from cultivated nucleo-cytoplasmic large DNA viruses (NCLDV) (in italics) and the Lake Ontario gene fragments used as qPCR targets (in bold type), and was constructed using Bayesian inference (BI). The qPCR targets LO1a-68 and LO1b-49 are from Lake Ontario clone libraries from 18 September to 10 October 2007, respectively (Short and Short, 2009). The other four qPCR-targeted *polB* fragments are labelled with the location (LO = Lake Ontario), the date of sample collection, and an arbitrary clone number. Values at the nodes indicate clade credibility (as percent probability) and where tree topology was conserved, the second values indicate approximate likelihood-ratio test (aLRT) branch support from a maximum likelihood phylogeny. The scale bar indicates the proportion of substitutions per site.

2008; Clasen and Suttle, 2009; Culley *et al.*, 2009). It is known that these primers do not amplify *polB* genes from other phycodnaviruses, such as *Emiliania huxleyi* virus, *Heterosigma akashiwo* virus, *Chrysochromulina ericina* virus and *Pyramimonas orientalis* virus, for example, (Sandaa *et al.*, 2001; Larsen *et al.*, 2008). Thus, it is apparent that the AVS primers amplify only a subset of phycodnaviruses and might even be biased for prasinoviruses. However, there is at least some evidence suggesting that the AVS primers are not particularly biased (acknowledging that they do not amplify certain phycodnavirus *polB* genes) and the preponderance of *Prasinovirus*-like sequences in environmental clones libraries may actually reflect their predominance in many aquatic environments (Clasen and Suttle, 2009).

After comparing all 195 *polB* fragments, which formed >14 clusters of sequences >97% identical to one another with respect to nucleotides, representative groups of sequences were selected for qPCR primer and probe design (Table 1) only if the group contained sequences found only in the summer clone library (for example, LO.16jul08.14), or if it contained sequences from all three seasons

(for example, LO.16jul08.20). qPCR primers and probes were also developed for two Lake Ontario *Chlorovirus polB* sequences. All *polB* fragments monitored via qPCR clustered among cultivated phycodnaviruses when compared with other nucleo-cytoplasmic large DNA viruses (Figure 1). The fragments LO1b-49, LO1a-68, LO.16jul08.20 and LO.16jul08.14 were most closely related to cultivated prasinoviruses, while LO.20may09.33 and LO.08may08.08 clustered among cultivated chloroviruses. For the sake of efficiency, but acknowledging the speculative nature of sequence identification, LO1b-49, LO1a-68, LO.16jul08.20 and LO.16jul08.14 will be referred to as '*Prasinovirus* fragments', and LO.20may09.33 and LO.08may08.08 will be referred to as '*Chlorovirus* fragments' from here on out.

On the basis of amplification of serially diluted standards, all qPCR primer and probe sets produced amplification efficiencies between 95% and 100%. The specificity of each primer and probe set was tested by comparing the cycle threshold from the amplification of  $10^7$  target molecules with amplification of the same number of non-target molecules (Table 2). The 'non-targets' used in these assays were

**Table 2** Properties of sequences used to test primer and probe specificity

Target sequence	Closest relative	Base-pair mismatches for closest relative				Ct from 10 <sup>7</sup> target molecules	Ct from 10 <sup>7</sup> non-target molecules
		Percent nucleotide identity	Forward primer	Reverse primer	Probe		
LO.16jul08.14	LO1a-34	77	7	5	5	14.77	No Ct
LO.16jul08.20	LO1a-68	81	2	4	4	13.92	31.49
LO.20may09.33	UTM.09jun09.47	69	6	7	13	12.34	33.14
LO.08may08.08	UTM.21sept09.07	76	6	7	4	12.30	No Ct

cloned Lake Ontario gene fragments that were most closely related to the target sequence. Only two primer and probe sets produced a detectable signal from non-target molecules, and the differences in Ct between targets and non-targets were 17.57 and 20.80 for the LO.16jul08.20 and LO.20may09.33 primer and probe sets, respectively. These Ct differences represent between five and six orders of magnitude differences in the detection limit for these assays. Therefore, as previously argued (Short and Short, 2009), it is extremely unlikely that non-target sequences contributed to estimates of gene abundance from amplification of *bona fide* target molecules.

Two *Prasinovirus* fragments (LO1b-49 and LO.16jul08.20) represented the most abundant genes monitored and were deemed 'high abundance prasinoviruses'; the peak abundances of LO1b-49 and LO.16jul08.20 were *ca* 5400 and 11 000 copies ml<sup>-1</sup>, respectively (Figure 2). In contrast, the other two *Prasinovirus* fragments LO.16jul08.14 and LO1a-68 reached only 960 and 880 copies ml<sup>-1</sup>, respectively, and were considered 'low abundance prasinoviruses', and the *Chlorovirus* fragments LO.20may09.33 and LO.08may08.08 peaked at *ca* 1600 and 300 copies ml<sup>-1</sup>, respectively. The striking differences in peak abundance observed in this study may be explained by differences in host abundance, virus burst size or even host range if some viruses are able to infect >1 host taxa. Moreover, it has been argued that viruses with high versus low environmental abundances may be due to different life strategies (Suttle, 2007). Following this reasoning, it is possible that the higher abundance prasinoviruses and chloroviruses observed in this study are more virulent, undergo more rapid replication and are more *r*-selected relative to their lower abundance counterparts.

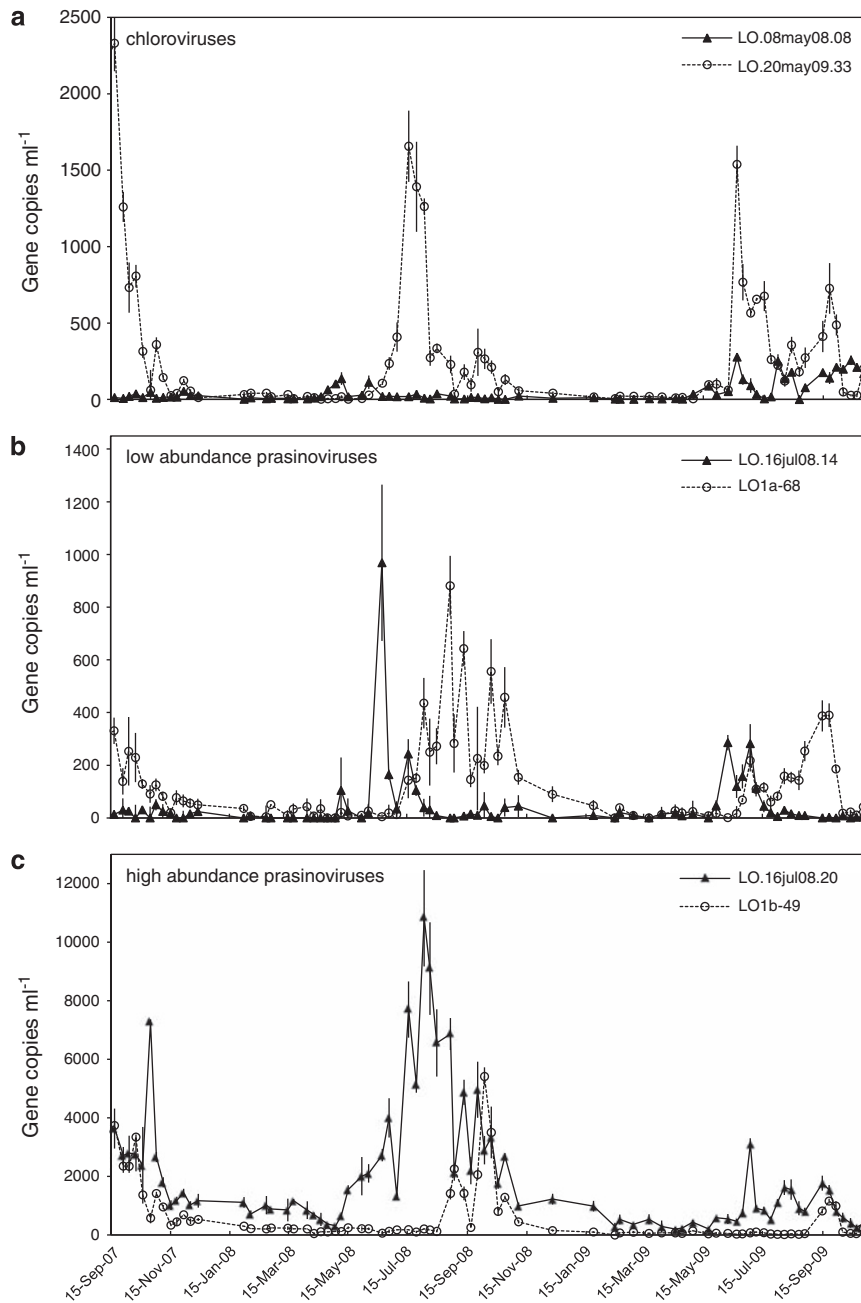
The timing and magnitude of peak abundances, and even the predictability of seasonal patterns differed for each *polB* fragment monitored (Figure 2). Both years of the study, the *Chlorovirus* fragment LO.20may09.33 reached its highest abundance later in the year (16 July 2008 and 18 June 2009) compared with LO.08may08.08, which reached its highest abundance on 8 May 2008 and 18 June 2009. The low abundance *Prasinovirus* fragment

LO.16jul08.14 peaked in the spring (19 June 2008 and 9 June 2009) whereas LO1a-68 peaked in the late summer (28 August 2008 and 21 September 2009). Similarly, the high abundance prasinoviruses LO.16jul08.20 and LO1b-49 peaked in the summer (1 August 2008 and 2 July 2009) or autumn (2 October 2008 and 21 September 2009) of each year, respectively. Between 2008 and 2009, most *polB* fragments reached different maximum abundances. This was especially true for the *Prasinovirus* fragments whose peak abundances in 2009 were, on average, only 31% of the peak abundances of 2008. The opposite trend was observed for the *Chlorovirus* fragment LO.08may08.08 whose peak abundance in 2009 was roughly double that of 2008, and the *Chlorovirus* fragment LO.20may09.33 reached similar peak abundances each year. Although the cause of these differences in maximum abundance are unknown, it can be speculated that the *Chlorovirus* hosts were less sensitive to changing environmental conditions and their inter-annual variability in peak abundance was reduced relative to the hosts of the prasinoviruses. As an additional comment on abundance patterns, autocorrelations revealed significant positive coefficients and indicated roughly annual patterns with lags of 13, 14, and 11 4-week periods for LO1b-49 ( $r=0.0423$ ), LO1a-68 ( $r=0.238$ ) and LO.20may09.33 ( $r=0.353$ ), respectively. Although significant positive coefficients were not observed for the other gene fragments, these autocorrelation analyses must be interpreted cautiously because of the limited size of the data set. Overall, the differences observed in the seasonal patterns of abundance support the speculation that the viruses encoding the monitored *polB* fragments infect different phytoplankton that have different seasonality themselves.

All *polB* fragments monitored displayed similar abundance patterns wherein peaks were followed by times of relatively low, but stable abundance through the late fall and winter months, and most fragments were detectable throughout the entire year except for LO.16jul08.14, which was not detected from mid February to early May 2008. Throughout the winter and early spring months (November to April) of both years of the study, the average abundances of the *Chlorovirus* fragments were  $11.79 \pm 7.58$  and

$26.1 \pm 9.0$  copies  $\text{ml}^{-1}$  for LO.08may08.08 and LO.20-may09.33, respectively. During the same period, the average abundances of the 'low abundance prasinoviruses'

were  $36.28 \pm 17.93$  and  $8.37 \pm 9.49$  copies  $\text{ml}^{-1}$  for LO1a-68 and LO.16jul08.14, respectively, whereas the 'high abundance prasinoviruses' averaged

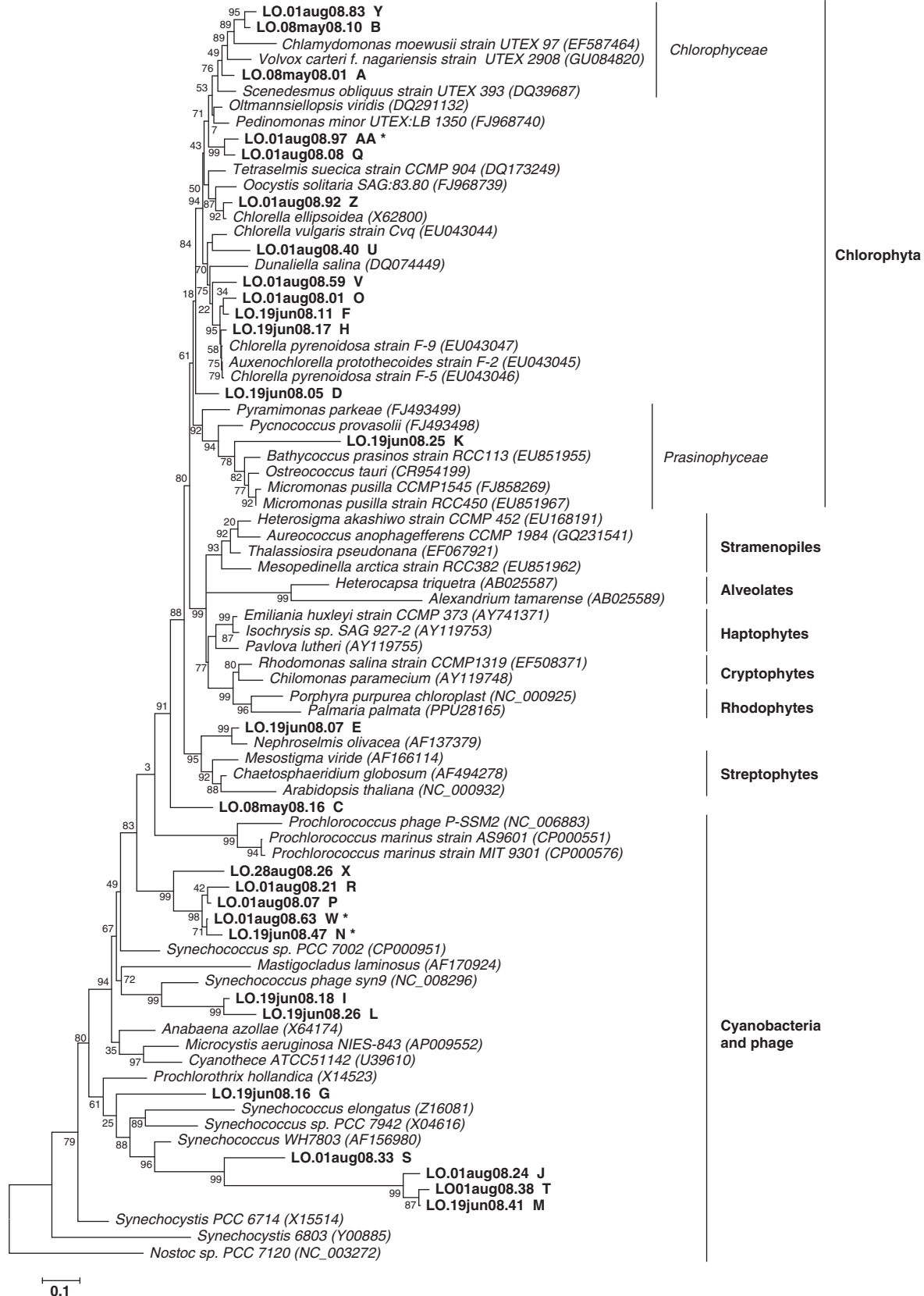


**Figure 2** The abundance of putative *Chlorovirus* (a) and *Prasinovirus* (b and c) *polB* gene fragments from September 2007 to November 2009. The prasinoviruses are split into two panels based on their maximum abundances (b: low abundance = maximum abundances  $< 1000$  copies  $\text{ml}^{-1}$ ; and c: high abundance = maximum abundance  $> 1000$  copies  $\text{ml}^{-1}$ ). Error bars represent s.d. from triplicate qPCR reactions.

**Figure 3** Maximum likelihood phylogeny of 564 aligned nucleotides of *psbA* genes. Maximum likelihood approximate likelihood-ratio test (aLRT) support values are indicated at the nodes. *psbA* sequences amplified in this study are indicated in bold and are labelled with the location (LO = Lake Ontario), the date of sample collection, and an arbitrary clone number. The letter following the clone number is an arbitrary label for the OTU (that is, group of sequences  $> 97\%$  identical) that corresponds the labels in Figure 4. Only a single representative sequence from each OTU was used for the phylogenetic analysis. Sequence groups containing potential chimeras are indicated with an \* after the clone name. The scale bar indicates the proportion of substitutions per site.

199.7 ± 32.5 and 747.2 ± 150 copies ml<sup>-1</sup> for LO1b-49, and LO.16jul08.20, respectively. These observations lead to questions of how some viruses survive at low

abundances for long periods of time, and how some viruses can maintain relatively high abundances throughout the year.





Although phycodnavirus persistence can be explained by chronic or latent infections, or even the infection of multiple host taxa, these seem unlikely because there are so few examples of these lifestyles known for cultivated phycodnaviruses. It is possible that some phycodnaviruses are incredibly efficient allowing them to be maintained through times of low host abundance. For example, the *Emiliania huxleyi* virus isolate  $\phi 28$  is highly infectious, and multiplicities of infection as low as  $10^{-5}$  can clear host cultures of only  $10^2$  cells  $\text{ml}^{-1}$  (Vaughn *et al.*, 2010). Yet, another possibility is that sediments are a reservoir that seeds viruses into the water column. The persistence of *polB* fragments in Lake Ontario is also interesting because it is known that viruses in aquatic environments are rapidly removed by a variety of biological, chemical and physical processes (Wommack and Colwell, 2000). In Lake Ontario and other environments subject to seasonal dynamics, persistence could be explained if low abundances allow viruses to avoid density-dependent sources of loss (for example, grazing), or if certain sources of destruction (for example, ultraviolet radiation) are reduced during the winter months. No matter the mechanism, and despite the fact that our sampling regime cannot capture the dynamics of these viruses on time scales relevant to their life histories (that is, hours or days), the observation of water column virus persistence suggests that Lake Ontario phycodnaviruses are not very abundant or active in winter, but form a reservoir from which any population can initiate infections when its host abundance increases. This is consistent with the Bank model proposed by Breitbart and Rohwer (2005), wherein only a small portion of a virus community is active and abundant at any given time and most populations are rare and inactive, forming a seed-bank that can ‘Kill-the-Winner’ (*sensu* Thingstad, 2000) when their hosts reach critical thresholds of abundance. For phytoplankton viruses in temperate environments, it is plausible that the majority form seed-bank populations during the winter months with most infections occurring in the spring, summer and autumn.

#### *psbA* identification and clone library analysis

One drawback to cultivation-free studies of phycodnaviruses is that host identities are unknown and can only be inferred from phylogenetic analyses of virus sequences. To gain better insights into Lake Ontario’s phytoplankton communities during the study period, *psbA* genes were PCR amplified, cloned and sequenced to >85% coverage from five samples collected on dates staggered throughout the 2008 growing season (Table 3). Phylogenetic analysis of Lake Ontario *psbA* fragments revealed that 12 of 27 sequence types or OTUs (arbitrarily labelled A through AA) were more closely related to cyanobacteria and cyanophage than eukaryotes (Figure 3). The remaining 15 OTUs clustered within the large

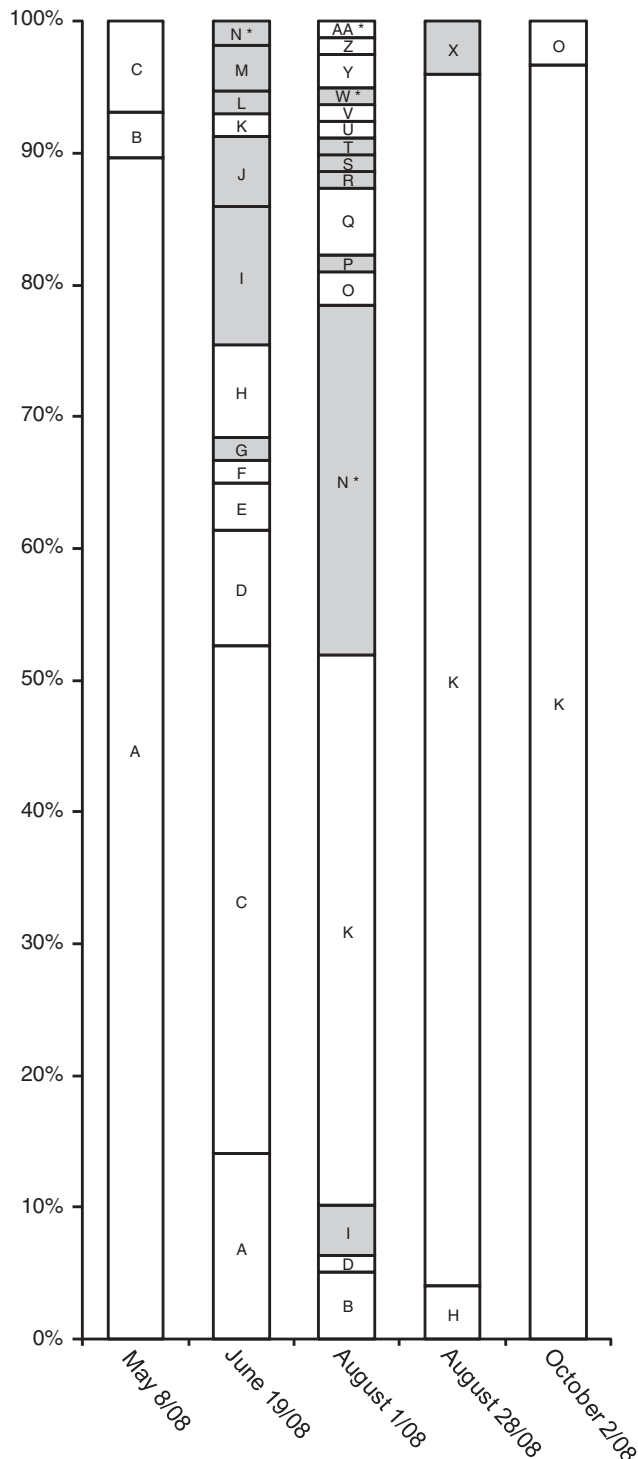
**Table 3** Properties of *psbA* clone libraries

Sample	Total number of sequences	Number of singletons	Sequencing coverage (%)
LO.08may08	29	1	96.6
LO.19jun08	57	5	91
LO.01aug08	79	10	87.3
LO.28aug08	25	2	92
LO.02oct08	30	1	96.7

clade ‘Chlorophyta’ except E, which was closely related to the prasinophyte *Nephroselmis olivacea* and C, which resolved to a branch between the eukaryotes and prokaryotes.

Potential hosts for prasinoviruses and chloroviruses were particularly interesting as these types of viruses were the presumed source of qPCR-monitored *polB* fragments. Currently, the only GenBank records of *psbA* sequences from phytoplankton taxa known to inhabit Lake Ontario (for phylogenetic review of Lake Ontario see Munawar and Munawar, 1996) were *Chlorella* sp., *Chlorella vulgaris* and *Chlamydomonas globosa*. Hence, most *psbA* sequences from Lake Ontario were difficult to identify based solely on their relationships to cultivated phytoplankton sequences. Nonetheless, the most likely source of several Lake Ontario OTUs was chlorophycean algae. For example, OTUs O, F and H were closely related to *Chlorella pyrenoidosa* (95.2, 97.1 and 98.0% nucleotide identity, respectively), OTU Z was 96.9% identical to *Chlorella ellipsoidea* and OTU U was 89.0% identical to *Chlorella vulgaris*. Similarly, OTUs Y and B were >88% identical to either *Volvox carterii* or *Chlamydomonas moewusii* (Figure 3). Despite the lake’s relatively high abundance and diversity of chlorophyte algae, previous phylogenetic studies of Lake Ontario have not noted prasinophycean taxa (Munawar and Munawar, 1996), yet two *psbA* OTUs could represent sequences from prasinophytes. OTU E was 94.8% identical (nucleotides) to the prasinophyte *Nephroselmis olivacea*, and OTU K clustered, albeit on a relatively long branch, among prasinophyte taxa that formed the monophyletic group ‘Prasinophyceae’ (Figure 3). Interestingly, *Nephroselmis olivacea* does not cluster with the other Prasinophyceae in our phylogeny. However, this is consistent with previous observations of prasinophyte systematic based on *psbA* (Zeidner *et al.*, 2003) and small subunit ribosomal DNA (Pröschold and Leliaert, 2007) sequences. In addition, OTU D, which clustered between prasinophytes and other green algae, is also a candidate host for the putative prasinoviruses observed in this study.

In all, 80% of the *psbA* sequences from Lake Ontario were related to *psbA* sequences from eukaryotes, and the remainder were most closely related to cyanobacteria or cyanophage *psbA* genes (Figures 3 and 4). Sequences closely related to



**Figure 4** Composition of Lake Ontario *psbA* clone libraries. The X axis is labelled with the sample collection dates for each clone library of PCR-amplified *psbA* gene fragments, and the Y axis represents the proportion of each unique OTU within an individual clone library. The letters are arbitrarily assigned labels used to identify unique OTUs. Shaded boxes indicate OTUs closely related to known cyanobacteria or phage, and OTUs containing potential chimeric sequences are indicated with an asterisk (\*).

*Chlorella* spp. were detected in all clone libraries except 8 May 2008. However, the sequences most closely related to a *Chlorella* sp. (OTUs H, F and O)

did not represent a high proportion of any clone libraries (Figure 4). Of these, OTU H (98% nucleotide identity to *Chlorella pyrenoidosa* strain F-9) was the most frequently observed with only four sequences in the 19 June clone library, and one in the 28 August clone library. As these *Chlorella*-like OTUs never represented a substantial proportion of any library it is difficult to speculate that they are hosts for any Lake Ontario chloroviruses. Nonetheless, if these OTUs represent *Chlorovirus* hosts and their frequency in the *psbA* clone libraries reflects their low abundance in nature, their low concentration could be the major factor driving the relatively low abundances of the *Chlorovirus* fragments monitored, as shown for certain scenarios of virus–host dynamics in the ‘killing the winner model’ (Winter *et al.*, 2010). It is interesting that although we only found two *psbA* sequences closely related to members of the *Prasinophyceae*, most sequences in the Lake Ontario AVS libraries are closely related to algal viruses known to infect prasinophytes. One explanation for this discrepancy in prasinovirus and prasinophyte hosts diversity may be that the majority of the ‘putative prasinoviruses’ do not actually infect prasinophytes, but rather they infect other closely related phytoplankton belonging to the Chlorophyta. Another possibility may be that host diversity was greatly underestimated in our study because the *psbA* libraries were only sequenced to 85% coverage.

A surprising result from this study was that the *psbA* clone libraries from May, late August and October were dominated by a single OTU. In all, 90% of the sequences from 8 May 2008 were >97% identical to OTU A, a close relative of *Volvox* and *Scenedesmus* (Figure 4). Later in the year, OTU A dropped to *ca* 14% of the 19 June clone library and disappeared from the summer and fall libraries. Although it is highly speculative based only on coincidence, OTU A is a possible host for the virus encoding the LO.16jul08.14 *polB* fragment because OTU A’s decline and disappearance coincided with the sharp spike in the abundance of the putative *Prasinovirus* fragment LO.16jul08.14. Furthermore, although this *polB* fragment was deemed a ‘low abundance *Prasinovirus*’ and is most closely related to cultivated prasinoviruses, it resides on a branch between cultivated prasinoviruses and chloroviruses (Figure 1), and may actually represent a novel type of virus that infects algae from another class of green algae. Overall, the sequence that was most often detected in the *psbA* clone libraries and dominated the 28 August and 2 October libraries was OTU K, a relative of prasinophyte algae (Figure 3). Coincidentally, the peak abundances noted for the *Prasinovirus* fragments LO1a-68 and LO1b-49 also occurred on 28 August and 2 October 2008. Thus is possible that OTU K is a host for Lake Ontario prasinoviruses that stably coexists with its virus parasites like some cyanobacteria and their phage (for example, Waterbury and Valois, 1993).

## Conclusions

Our data suggest that a common ecological feature of Lake Ontario phycodnaviruses is episodic 'blooms' followed by months of persistence at relatively low abundances. It has been hypothesized that the vast majority of viruses in any environment are dormant with only a few members that are active and abundant at any point in time, and that this activity is directly influenced by the host abundances and growth. Recent studies employing the cultivation-based assays (Marston and Sallee, 2003), metagenomics (Breitbart *et al.*, 2002, 2004), molecular fingerprinting (Wommack *et al.*, 1999a; Parada *et al.*, 2008), and now the qPCR data reported here support this 'Bank model' (*sensu* Breitbart and Rohwer, 2005) hypothesis. Further, differences in the average, and peak abundances of the *polB* fragments monitored suggest that there are differences in life history strategies of different phycodnaviruses. Although we can speculate that the hosts for the viruses encoding the *polB* fragments we tracked are prasinophytes or *Chlorella*-like phytoplankton, we do not yet know their exact identity. Nevertheless, identification of phytoplankton taxa coinciding with the peaks in phycodnavirus *polB* fragment abundances has provided plausible host taxa and leads for future cultivation efforts. Only by fulfilling Koch's postulates will the hosts for Lake Ontario phycodnaviruses be truly known, yet the cultivation-free efforts reported have provided insights into phycodnavirus dynamics, and suggest that virus decay rates and mechanisms for persistence need to be re-examined for a more complete understanding of the ecology of aquatic viruses and their hosts.

## Acknowledgements

We are grateful to Natasha Musrap and Michael Staniewski for their help collecting and processing water samples. This study was supported by the Canadian Foundation for Innovation Leaders Opportunity Fund and NSERC Discovery grants awarded to SMS.

## References

- Breitbart M, Felts B, Kelley S, Mahaffy JM, Nulton J, Salamon P *et al.* (2004). Diversity and population structure of a near-shore marine-sediment viral community. *Proc R Soc Lond B Biol Sci* **271**: 565–574.
- Breitbart M, Rohwer F. (2005). Here a virus, there a virus, everywhere the same virus? *Trends Microbiol* **13**: 278–284.
- Breitbart M, Salamon P, Andresen B, Mahaffy JM, Segall AM, Mead D *et al.* (2002). Genomic analysis of uncultured marine viral communities. *Proc Natl Acad Sci USA* **99**: 14250–14255.
- Brussaard CPD, Short SM, Frederickson CM, Suttle CA. (2004). Isolation and phylogenetic analysis of novel viruses infecting the phytoplankton *Phaeocystis globosa* (Prymnesiophyceae). *Appl Environ Microbiol* **70**: 3700–3705.
- Castberg T, Larsen A, Sandaa RA, Brussaard CPD, Egge JK, Heldal M *et al.* (2001). Microbial population dynamics and diversity during a bloom of the marine coccolithophorid *Emiliania huxleyi* (Haptophyta). *Mar Ecol Prog Ser* **221**: 39–46.
- Castresana J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* **17**: 540–552.
- Chen F, Suttle CA. (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.
- Chen F, Suttle CA, Short SM. (1996). Genetic diversity in marine algal virus communities as revealed by sequence analysis of DNA polymerase genes. *Appl Environ Microbiol* **62**: 2869–2874.
- Clasen JL, Suttle CA. (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.
- Cottrell MT, Suttle CA. (1995). Dynamics of a lytic virus infecting the photosynthetic marine picoflagellate *Micromonas pusilla*. *Limnol Oceanogr* **40**: 730–739.
- Culley AI, Asuncion BF, Steward GF. (2009). Detection of inteins among diverse DNA polymerase genes of uncultivated members of the Phycodnaviridae. *ISME J* **3**: 409–418.
- Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F *et al.* (2008). Phylogeny fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res* **36**: W465–W469.
- Dunigan DD, Fitzgerald LA, Van Etten JL. (2006). Phycodnaviruses: a peek at genetic diversity. *Virus Res* **117**: 119–132.
- Edgar RC. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792–1797.
- Guindon S, Gascuel O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**: 696–704.
- Hall TA. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* **41**: 95–98.
- Huber T, Faulkner G, Hugenholtz P. (2004). Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. *Bioinformatics* **20**: 2317–2319.
- Larsen A, Castberg T, Sandaa RA, Brussaard CPD, Egge J, Heldal M *et al.* (2001). Population dynamics and diversity of phytoplankton, bacteria and viruses in a seawater enclosure. *Mar Ecol Prog Ser* **221**: 47–57.
- Larsen JB, Larsen A, Bratbak G, Sandaa RA. (2008). Phylogenetic analysis of members of the Phycodnaviridae virus family, using amplified fragments of the major capsid protein gene. *Appl Environ Microbiol* **74**: 3048–3057.
- Lawrence JE, Steward GF. (2010). Purification of viruses by centrifugation. In: Wilhelm SW, Weinbauer MG, Suttle CA (eds). *Manual of Aquatic Viral Ecology*. American Society of Limnology and Oceanography: Waco, pp 166–181.
- Li WKW, Dickie PM. (2001). Monitoring phytoplankton, bacterioplankton, and virioplankton in a coastal inlet (Bedford Basin) by flow cytometry. *Cytometry* **44**: 236–246.
- Marston MF, Sallee JL. (2003). Genetic diversity and temporal variation in the cyanophage community

- infecting marine *Synechococcus* species in Rhode Island's coastal waters. *Appl Environ Microbiol* **69**: 4639–4647.
- Munawar M, Munawar IF. (1982). Phycological studies in Lakes Ontario, Erie, Huron, and Superior. *Can J Botany* **60**: 1837–1858.
- Munawar M, Munawar IF. (1996). *Phytoplankton Dynamics in the North American Great Lakes: Lake Ontario, Erie and St Clair* vol. 1 SPB Academic Publishing: Amsterdam. p 282.
- Nagasaki K, Bratbak G. (2010). Isolation of viruses infecting photosynthetic and nonphotosynthetic protists. In: Wilhelm SW, Weinbauer MG, Suttle CA (eds). *Manual of Aquatic Viral Ecology*. American Society of Limnology and Oceanography: Waco. pp 82–91.
- Parada V, Baudoux AC, Sintès E, Weinbauer MG, Herndl GJ. (2008). Dynamics and diversity of newly produced viroplankton in the North Sea. *ISME J* **2**: 924–936.
- Pröschold T, Lelièvre F. (2007). Systematics of the green algae: conflict of classic and modern approaches. In: Brodie J, Lewis J (eds). *Unravelling the Algae*. CRC Press: Boca Raton. p. 376.
- Ronquist F, Huelsenbeck JP. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572–1574.
- Sandaa RA, Heldal M, Castberg T, Thyrhaug R, Bratbak G. (2001). Isolation and characterization of two viruses with large genome size infecting *Chrysochromulina ericina* (Prymnesiophyceae) and *Pyramimonas orientalis* (Prasinophyceae). *Virology* **290**: 272–280.
- Sandaa RA, Larsen A. (2006). Seasonal variations in virus-host populations in Norwegian coastal waters: focusing on the cyanophage community infecting marine *Synechococcus* spp. *Appl Environ Microbiol* **72**: 4610–4618.
- Short SM, Short CM. (2008). Diversity of algal viruses in various North American freshwater environments. *Aqua Microbial Ecol* **51**: 13–21.
- Short SM, Short CM. (2009). Quantitative PCR reveals transient and persistent algal viruses in Lake Ontario, Canada. *Environ Microbiol* **11**: 2639–2648.
- Short SM, Suttle CA. (2002). Sequence analysis of marine virus communities reveals that groups of related algal viruses are widely distributed in nature. *Appl Environ Microbiol* **68**: 1290–1296.
- Short SM, Suttle CA. (2003). Temporal dynamics of natural communities of marine algal viruses and eukaryotes. *Aqua Microbial Ecol* **32**: 107–119.
- Suttle CA. (2007). Marine viruses - major players in the global ecosystem. *Nat Rev Microbiol* **5**: 801–812.
- Suttle CA, Chan AM. (1994). Dynamics and distribution of cyanophages and their effect on marine *Synechococcus* spp. *Appl Environ Microbiol* **60**: 3167–3174.
- Tamura K, Dudley J, Nei M, Kumar S. (2007). MEGA4: molecular evolutionary genetics analysis (MEGA4) software version 4.0. *Mol Biol Evol* **24**: 1596–1599.
- Thingstad TF. (2000). Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. *Limnol Oceanogr* **45**: 1320–1328.
- Thompson JD, Higgins DG, Gibson TJ. (1994). Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**: 4673–4680.
- Tomaru Y, Hata N, Masuda T, Tsuji M, Igata K, Masuda Y *et al*. (2007). Ecological dynamics of the bivalve-killing dinoflagellate *Heterocapsa circularisquama* and its infectious viruses in different locations of western Japan. *Environ Microbiol* **9**: 1376–1383.
- Van Etten JL, Graves MV, Muller DG, Boland W, Delaroque N. (2002). *Phycodnaviridae* - large DNA algal viruses. *Arch Virol* **147**: 1479–1516.
- Van Etten JL, Meints RH. (1999). Giant viruses infecting algae. *Annu Rev Microbiol* **53**: 447–494.
- Vaughn JM, Balch WM, Novotny JF, Vining CL, Palmer CD, Drapeau DT *et al*. (2010). Isolation of *Emiliania huxleyi* viruses from the Gulf of Maine. *Aqua Microbial Ecol* **58**: 109–116.
- Viprey M, Guillou L, Ferreol M, Vaulot D. (2008). Wide genetic diversity of picoplanktonic green algae (Chloroplastida) in the Mediterranean Sea uncovered by a phylum-biased PCR approach. *Environ Microbiol* **10**: 1804–1822.
- Waterbury JB, Valois FW. (1993). Resistance to cooccurring phages enables marine *Synechococcus* communities to coexist with cyanophages abundant in seawater. *Appl Environ Microbiol* **59**: 3393–3399.
- Winter C, Bouvier T, Weinbauer MG, Thingstad TF. (2010). Trade-offs between competition and defense specialists among unicellular planktonic organisms: the 'killing the winner' hypothesis revisited. *Microbiol Mol Biol Rev* **74**: 42–57.
- Wommack KE, Colwell RR. (2000). Viroplankton: viruses in aquatic ecosystems. *Microbiol Mol Biol Rev* **64**: 69–114.
- Wommack KE, Ravel J, Hill RT, Chun JS, Colwell RR. (1999a). Population dynamics of Chesapeake Bay viroplankton: total community analysis by pulsed-field gel electrophoresis. *Appl Environ Microbiol* **65**: 231–240.
- Wommack KE, Ravel J, Hill RT, Colwell RR. (1999b). Hybridization analysis of Chesapeake Bay viroplankton. *Appl Environ Microbiol* **65**: 241–250.
- Zeidner G, Preston CM, Delong EF, Massana R, Post AF, Scanlan DJ *et al*. (2003). Molecular diversity among marine picophytoplankton as revealed by *psbA* analyses. *Environ Microbiol* **5**: 212–216.
- Zingone A, Sarno D, Forlani G. (1999). Seasonal dynamics in the abundance of *Micromonas pusilla* (Prasinophyceae) and its viruses in the Gulf of Naples (Mediterranean Sea). *J Plankton Res* **21**: 2143–2159.

Supplementary Information accompanies the paper on The ISME Journal website (<http://www.nature.com/ismej>)