Contrasting diversity of phycodnavirus signature genes in two large and deep western European lakes

Xu Zhong and Stéphan Jacquet*

INRA, UMR CARRTEL, 75 Avenue de Corzent, 74203 Thonon-les-Bains cx, France.

Summary

Little is known about *Phycodnavirus* (or double-stranded DNA algal virus) diversity in aquatic ecosystems, and virtually, no information has been provided for European lakes. We therefore conducted a 1-year survey of the surface waters of France's two largest lakes, Annecy and Bourget, which are characterized by different trophic states and phytoplanktonic communities. We found complementary and contrasting diversity of phycodnavirus in the lakes based on two genetic markers, the B family DNA polymerase-encoding gene (*polB*) and the major capsid protein-encoding gene (*mcp*). These two core genes have already been used, albeit separately, to infer phylogenetic relationships and genetic diversity among members of the phycodnavirus family and to determine the occurrence and diversity of these genes in natural viral communities. While *polB* yielded prasinovirus-like sequences, the *mcp* primers yielded sequences for prasinoviruses, chloroviruses, prymnesioviruses and other groups not known from available databases. There was no significant difference in phycodnavirus populations between the two lakes when the sequences were pooled over the full year of investigation. By comparing Lakes Annecy and Bourget with data for other aquatic environments around the world, we show that these alpine lakes are clearly distinct from both other freshwater ecosystems (lakes and rivers) and marine environments, suggesting the influence of unique biogeographic factors.

Introduction

Viruses are the most abundant biological entities in aquatic ecosystems and exhibit the greatest genetic diversity of any group on Earth (Suttle, 2005; Angly *et al.*, 2006). Most studies have focused on viruses that infect prokaryotes

(Weinbauer, 2004; Danovaro *et al.*, 2008), and so there are little data on viruses that infect eukaryotic phytoplankton. Indeed, there is a paucity of studies dealing with all aspects of algal virus ecology, including 'species' diversity and distribution, community dynamics and their role as mortality agents and conductors of biogeochemical cycling of elements (Brussaard, 2004; Brussaard *et al.*, 2008). Viruses infecting eukaryotic algal viruses are primarily double-stranded DNA (dsDNA) viruses. It has been suggested that they are a very diverse group (Suttle, 2005; Dunigan *et al.*, 2006; Bench *et al.*, 2007), and research has shown that they play a role in the control and regulation of phytoplankton, typically for blooming species in marine coastal waters (Bratbak *et al.*, 1993; Jacquet *et al.*, 2002; Tomaru *et al.*, 2008).

The *Phycodnaviridae* infect a variety of eukaryotic microalgae, including Chlorophyta, Dinophyta, Haptophyta and Heterokonta. They belong to the nucleocytoplasmic large DNA viruses, which are characterized by having a polyhedral capsid, no tail or envelope, and large dsDNA genomes ranging from 160 to 560 kb (Van Etten and Meints, 1999; Van Etten *et al.*, 2002; Iyer *et al.*, 2006; Wilson *et al.*, 2009). Currently, the *Phycodnaviridae* consist of six genera named after the hosts they infect: *Chlorovirus, Coccolithovirus, Prasinovirus, Prymnesiovirus, Phaeovirus* and *Raphidovirus* (Dunigan *et al.*, 2006; Wilson *et al.*, 2009). So far, the characterization of phycodnaviruses has been limited to only a few individuals. However, as mortality agents of eukaryotic algae and vehicles of gene transfer, we know that they play a crucial role in host community structuring, population succession, resistance phenomena and the cycling of carbon in aquatic environments (Brussaard, 2004; Larsen *et al.*, 2004; Suttle, 2005; Monier *et al.*, 2008; Tomaru *et al.*, 2008; Thomas *et al.*, 2012).

Until now, the diversity of phycodnaviruses has been estimated using two molecular markers, the *polB* and *mcp* genes, which encode for DNA polymerase and their major capsid protein respectively. Primers that target these genes have already provided varying results regarding phycodnavirus occurrence and diversity (Schroeder *et al.*, 2003; Short, 2012). Investigations using *polB* have revealed a very diverse and wide distribution of phycodnaviruses in marine environments (Chen *et al.*, 1996; Short and Suttle, 2002; 2003; Bellec *et al.*, 2009;

Park *et al.*, 2011), rivers (Short and Short, 2008; Gimenes *et al.*, 2012) and lakes (Short and Short, 2008; Clasen and Suttle, 2009). In these studies, carried out in various marine and freshwater locations, all or almost all of the *polB* sequences obtained were closely related to prasinoviruses. While Larsen *et al.* (2008) developed universal primers that target *mcp*, this marker has only been used twice: once in Norwegian fjords (Larsen *et al.*, 2008) and once in the coastal waters of Korea (Park *et al.*, 2011). These studies revealed that *mcp* may, in fact, be a useful genetic marker for inferring phylogenetic relationships and genetic diversity among members of the phycodnavirus family in natural viral communities.

To the best of our knowledge, only a limited number of freshwater lakes have, so far, been investigated for phycodnaviruses (Short, 2012). Studies have generally focused on one ecosystem and used only one molecular marker, typically *polB*. Our study is unique in that we use primers that target both *mcp* and *polB* and examine two large and deep lakes located in the same eco-area but characterized by different trophic states (oligotrophic vs. oligo-mesotrophic) and phytoplanktonic communities (Jacquet *et al.*, in press). By collecting samples every 3–4 weeks, our investigation also provides new insights into phycodnavirus community structure and diversity in alpine lakes.

Results

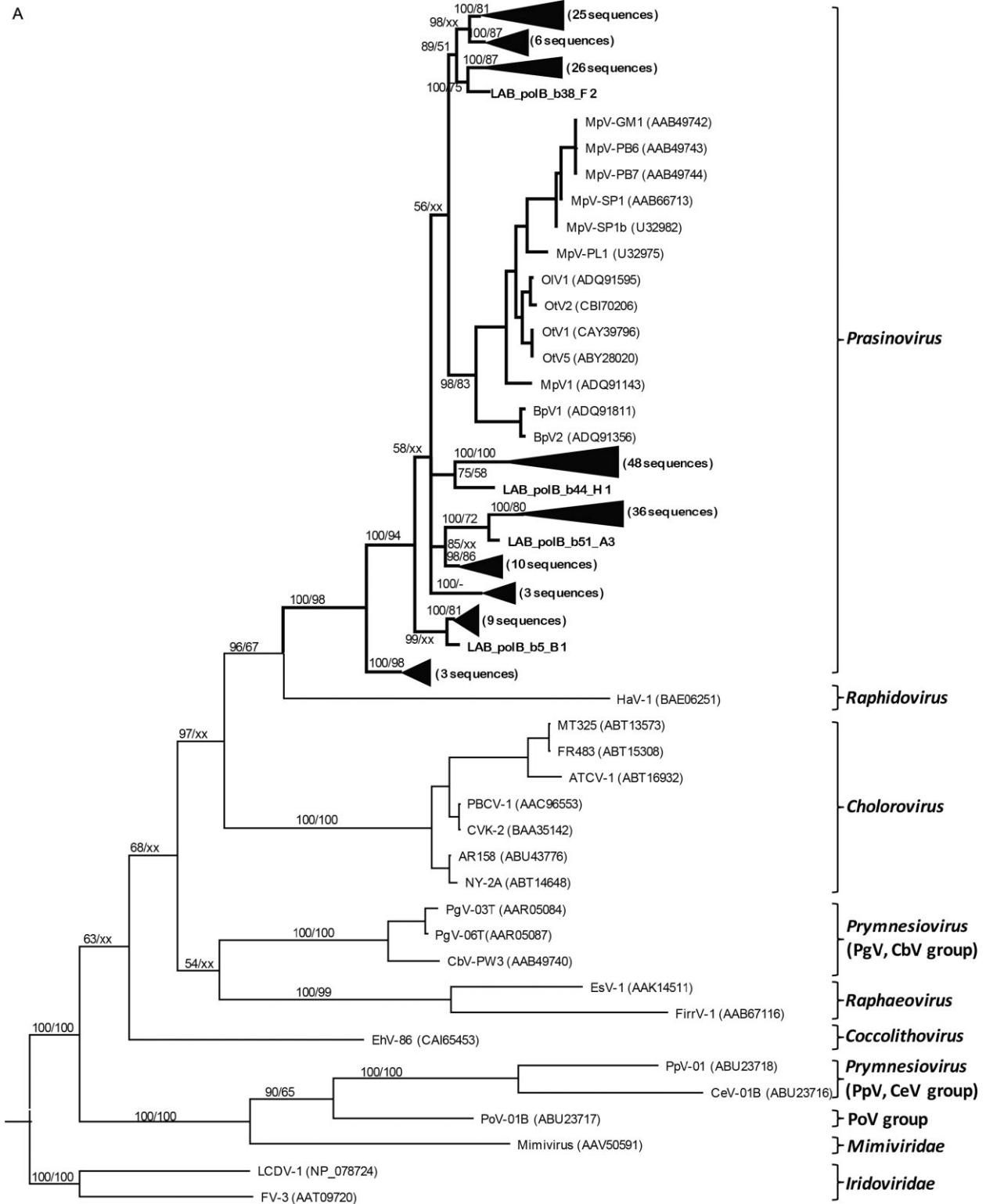### Diversity of phycodnaviruses as estimated from polB

From the excised 48 visibly different bands, we obtained a total of 163 non-redundant sequences from 665 to 683 bp, with nucleotide similarity ranging from 37.7% to 99.4%. The phylogenetic analysis of these sequences with phycodnavirus culture representatives showed that all of the obtained *polB* sequences were grouped within marine prasinoviruses and apart from other phycodnavirus groups (Fig. 1A). When we conducted the metaphylogenetic analysis by introducing all prasinovirus-like *polB* sequences from other published studies (Table 1) and from the GOS (Global Ocean Survey) metagenome database, we discerned five clades (Fig. 2). Three of them were initially identified by Short and Short (2008): Freshwater Cluster I, Freshwater Cluster II and Marine and Freshwater Cluster [with *Micromonas pusilla* virus (MpV), *Ostreococcus tauri* virus (OtV) and *Ostreococcus lucimarinus* virus (OlV)]. The other two clades were Freshwater Cluster III and Marine BpV (*Bathycoccus prasinos* virus) Cluster, which branched off Freshwater Cluster II and Marine and Freshwater Cluster respectively. The three freshwater clusters (I, II and III) contained sequences that originated exclusively from freshwater environments. The Freshwater Cluster II contained 177 sequences, 93.8% of them from rivers or a reservoir (South Platte River,

Chatfield Reservoir, Cuieiras River and Solimões River), with only three sequences from Lake Bourget or Lake Ontario, six from Lake 239 and two from Lake 240 (Fig. 2). In Freshwater Cluster I, all of the sequences originated from lakes. Freshwater Cluster III contained sequences from both rivers and lakes. The Marine and Freshwater Cluster comprised all of the *polB* sequences recruited from GOS, and also sequences from both lakes and rivers. These sequences were closely related to the marine MpV, OtV and OlV prasinovirus groups. The Marine BpV Cluster contained sequences exclusively from marine environments and was closely related to the marine prasinovirus BpV group.

In total, 81.6% of *polB* sequences obtained from Lake Annecy and Bourget belonged to Freshwater Cluster I, 17.2% to the Marine and Freshwater Cluster and only 1.2% to Freshwater Cluster II. No sequences belonged to the Freshwater Cluster III and Marine BpV Cluster (Fig. 2). An examination of the colour ring in the phylogenetic tree shows that the distribution of sequences from either Lake Annecy or Lake Bourget is clearly separated from the others (Fig. 2). A similar clustering pattern was observed for the majority of the sequences obtained. No clade was privileged by sequences from either lake. Based on the pooled 1-year *polB* sequences, the Unifrac analysis revealed no significant differences between Lake Annecy and Lake Bourget (Fig. 3, *P* = 1). The *polB* sequences originating from marine environments seemed different from those obtained in freshwater, as revealed by the principal component analysis (PCA) (Fig. 3, *P* = 0).

Diversity of phycodnaviruses as estimated from mcp

From the excised 60 visibly different representative bands, we obtained 115 non-redundant sequences for both lakes, varying from 362 to 527 bp, and with nucleotide similarity varying between 29.5% and 98.5%. After phylogenetic analysis, the obtained sequences from Lake Annecy and Bourget were clustered into several phycodnavirus groups: *Prymnesiovirus*, *Prasinovirus* and Clusters number 4, 6, 7 and 9 (Fig. 1B). In general, the clustering pattern matched the results of the metaphylogenetic analysis when we included sequences observed in other environments (Table 1) and from the GOS and Antarctica-Aquatic metagenome databases, while keeping both maximum likelihood (ML) bootstrap and Bayesian inference (BI) clade credibility values ≥ 50 (Fig. 4). Very different results were obtained for prymnesioviruses versus prasinoviruses. For the *Prymnesiovirus* group, 60% of the sequences came from Lake Annecy and Bourget, while 44% and 22.6% came from *mcp* sequences in the GOS database and the Antarctica-Aquatic (Ace Lake) metagenome database respectively. For the *Prasinovirus* group, 35% and 71% of the sequences came from *mcp*

A

100/81 (25 sequences)
98/xx
00/87 (6 sequences)
89/51
100/87 (26 sequences)
100/75 **LAB_polB_b38_F 2**

MpV-GM1 (AAB49742)
MpV-PB6 (AAB49743)
MpV-PB7 (AAB49744)
MpV-SP1 (AAB66713)
MpV-SP1b (U32982)
MpV-PL1 (U32975)
OIV1 (ADQ91595)
OtV2 (CBI70206)
OtV1 (CAY39796)
OtV5 (ABY28020)
MpV1 (ADQ91143)
BpV1 (ADQ91811)
BpV2 (ADQ91356)

56/xx

98/83

58/xx
100/100 (48 sequences)
75/58 **LAB_polB_b44_H 1**
100/80 (36 sequences)
100/72 **LAB_polB_b51_A3**
85/xx
98/86 (10 sequences)
100/- (3 sequences)
100/81 (9 sequences)
99/xx **LAB_polB_b5_B1**
100/98 (3 sequences)

100/94

100/98

96/67

Prasinovirus

HaV-1 (BAE06251) *Raphidovirus*

97/xx

MT325 (ABT13573)
FR483 (ABT15308)
ATCV-1 (ABT16932)
PBCV-1 (AAC96553)
CVK-2 (BAA35142)
AR158 (ABU43776)
NY-2A (ABT14648)
100/100

Cholorovirus

68/xx

PgV-03T (AAR05084)
PgV-06T (AAR05087)
CbV-PW3 (AAB49740)
100/100

Prymnesiovirus (PgV, CbV group)

54/xx

EsV-1 (AAK14511)
FirrV-1 (AAB67116)
100/99

Raphaeovirus

EhV-86 (CAI65453) *Coccolithovirus*

63/xx

PpV-01 (ABU23718)
CeV-01B (ABU23716)
100/100

Prymnesiovirus (PpV, CeV group)

90/65

PoV-01B (ABU23717) PoV group

100/100

Mimivirus (AAV50591) *Mimiviridae*

LCDV-1 (NP_078724)
FV-3 (AAT09720)
100/100

Iridoviridae

100/100

Fig. 1. Bayesian phylogenetic tree of *polB* (A) and *mcp* (B) genes based on inferred amino acid sequences from Lake Annecy and Bourget, and nucleocytoplasmic large DNA viruses (NCLDVs) isolates. The phylogeny is based on the alignment of 201 and 112 homologous amino acid positions for *polB* and *mcp* respectively. Values shown at nodes of the main branches are the Bayesian inference (BI) clade credibility and maximum likelihood (ML) bootstrap values, and are reported as BI/ML, where 'xx' indicates a value < 50%. When both BI and ML support values > 80%, the sequences from Lake Annecy and Bourget are grouped in black triangles. The number of sequences is given in parentheses.
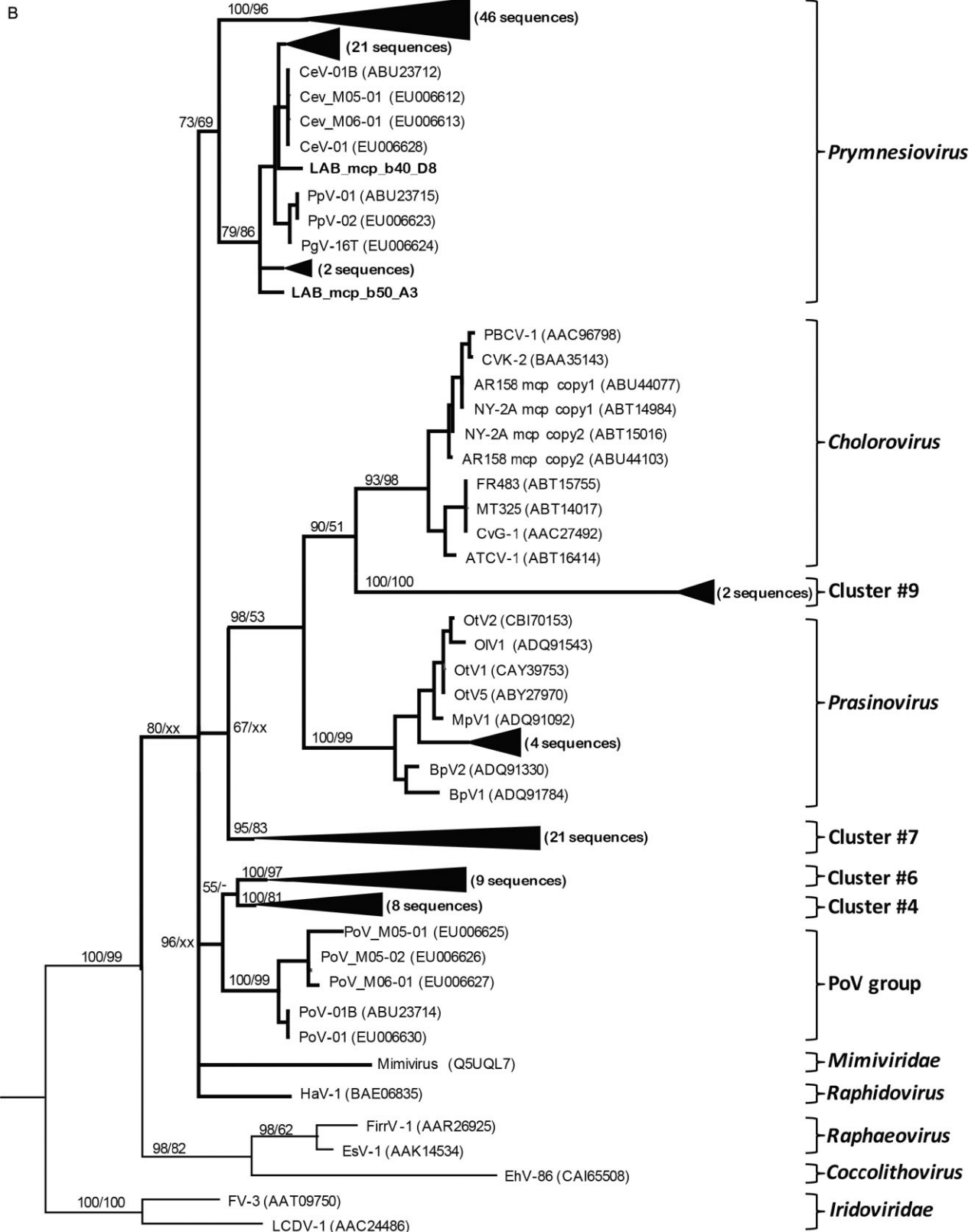
B

Fig. 1. cont.

Table 1. Origins of *polB* and *mcp* environmental sequences used for the phylogenetic analyses reported in Figs 2 and 4.

Gene markers	Environments	Prefix of sequence label	Locations	References	Genbank accession number
polB	Freshwater	L224	Lake 224, ELA, ON, Canada	Clasen and Suttle, 2009	EU408225 to EU408244
		L227	Lake 227, ELA, ON, Canada	Clasen and Suttle, 2009	EU408225 to EU408244
		L239	Lake 239, ELA, ON, Canada	Clasen and Suttle, 2009	EU408225 to EU408244
		L240	Lake 240, ELA, ON, Canada	Clasen and Suttle, 2009	EU408225 to EU408244
		CL1	Crawford Lake, ON, Canada	Short and Short, 2008	EU336433 to EU336476
		CR	Chatfield Reservoir, CO, Canada	Short and Short, 2008	EU336477 to EU336572
		LO1	Lake Ontario, ON, Canada	Short and Short, 2008	EU336573 to EU336707
		SPR	South Platte River, CO, Canada	Short and Short, 2008	EU336708 to EU336803
		CUI	Cuieiras River, AM, Brazil	Gimenes *et al.*, 2012	HQ424349 to HQ424370
		SOL	Solimões River, AM, Brazil	Gimenes *et al.*, 2012	HQ424371 to HQ424430
		LAB_*polB*	Lake Annecy and Bourget, HS, France	this study	KC574505 to KC574667
	Marine	BS	Barkley Sound, BC, Canada	Short and Suttle, 2002	AF405572 to AF405604
		MI	Malaspina Inlet, BC, Canada	Short and Suttle, 2002	AF405572 to AF405604
		PS	Pendrell Sound, BC, Canada	Short and Suttle, 2002	AF405572 to AF405604
		SI	Salmon Inlet, BC, Canada	Short and Suttle, 2002	AF405572 to AF405604
		SO	Southern ocean	Short and Suttle, 2002	AF405572 to AF405604
		OTU	Gulf of Mexico, TX, USA	Chen *et al.*, 1996	U36931 to U36935
		Jpavs	Jericho Pier, BC, Canada	Short and Suttle, 2003	AY145089 to AY145098
		ESO2	Marine aerosols, East Pacific ocean	Unpublished	AY436587 to AY436589
		JCVI	GOS: various sites of world's oceans	Rusch *et al.*, 2007	Multiple
mcp	Freshwater	LAB_*mcp*	Lake Annecy and Bourget, HS, France	this study	KC574390 to KC574504
	Antarctica-Aquatic		Ace Lake, Antarctica	Lauro *et al.*, 2011	Multiple
	Marine	OTU	Norwegian fjords, Norway	Larsen *et al.*, 2008	EU006614 to EU006622
		JCVI	GOS: various sites of world's oceans	Rusch *et al.*, 2007	Multiple

sequences in the GOS database and the Ace Lake database, respectively, but only five sequences from our alpine lakes.

Clusters number 1–8 were unidentified environmental clusters because culture representatives to support their lineage are missing. Clusters number 1, 2, 3, 5 and 8 all contained GOS sequences but no sequences from alpine lakes. Sequences from Norwegian fjords were found in Clusters number 2 and 3, and sequences from Ace Lake in Cluster number 1. The closest homologue to sequences in Cluster number 3 was raphidovirus, but the low bootstrap value for ML did not support the hypothesis that they be grouped together into a single cluster (Fig. 4). Cluster number 7 grouped with prasinovirus and chlorovirus (Fig. 1B), while Clusters number 4 and 6 bunched together with the *Pyraminomas orientalis* virus (PoV) group (Fig. 1B), yet the validity of these clusters was not supported by either the ML bootstrap value (Fig. 1B) or our metaphylogenetic analysis (Fig. 4). The two sequences in Cluster number 9 were closely related to freshwater chloroviruses, exhibiting ML bootstrap and BI clade credibility values of ≥ 50 (Fig. 1B), even though they branched off from the *Chlorovirus* cluster in the metaphylogenetic analysis (Fig. 4). It is noteworthy that one of the denaturing gradient gel electrophoresis (DGGE) bands from which these sequences originated was only detected once in October in Lake Bourget, when chlorophyceae abundance as observed by microscopy was at its highest that year (data not shown). This finding

could suggest that sequences from Cluster number 9 were for viruses infecting chlorophyceae.

Likewise, as shown by the pooled 1-year sequences, *mcp* sequences for neither Lake Annecy or Lake Bourget could form a separate clade in other clusters or viral groups in the phylogenetic tree. The Unifrac analysis also showed no significant differences for these sequences between the two alpine lakes (Fig. 3, $P = 1$). In the PCA, Lake Annecy and Bourget clustered together and were separated from either GOS or Ace Lake, as well as from the Norwegian fjords (Fig. 3).

Discussion

Differences in phycodnavirus diversity depending on the genetic marker used

This study of algal virus diversity demonstrates, for the first time, the prevalence of phycodnavirus signature genes in two European lakes characterized by different trophic states. However, the pattern of phycodnavirus diversity observed in these lakes varied depending on the genetic marker used (note that the markers were used on the same sample). Overall, 60% of the obtained *mcp* sequences were assigned to the *Prymnesiovirus* lineage, while 100% of the *polB* sequences clustered with the marine prasinoviruses. Using *mcp* primers allowed us to target a broader range of phycodnaviruses than with *polB* alone, yielding sequences belonging not only to prasinoviruses, but also to chloroviruses, prymnesioviruses and other
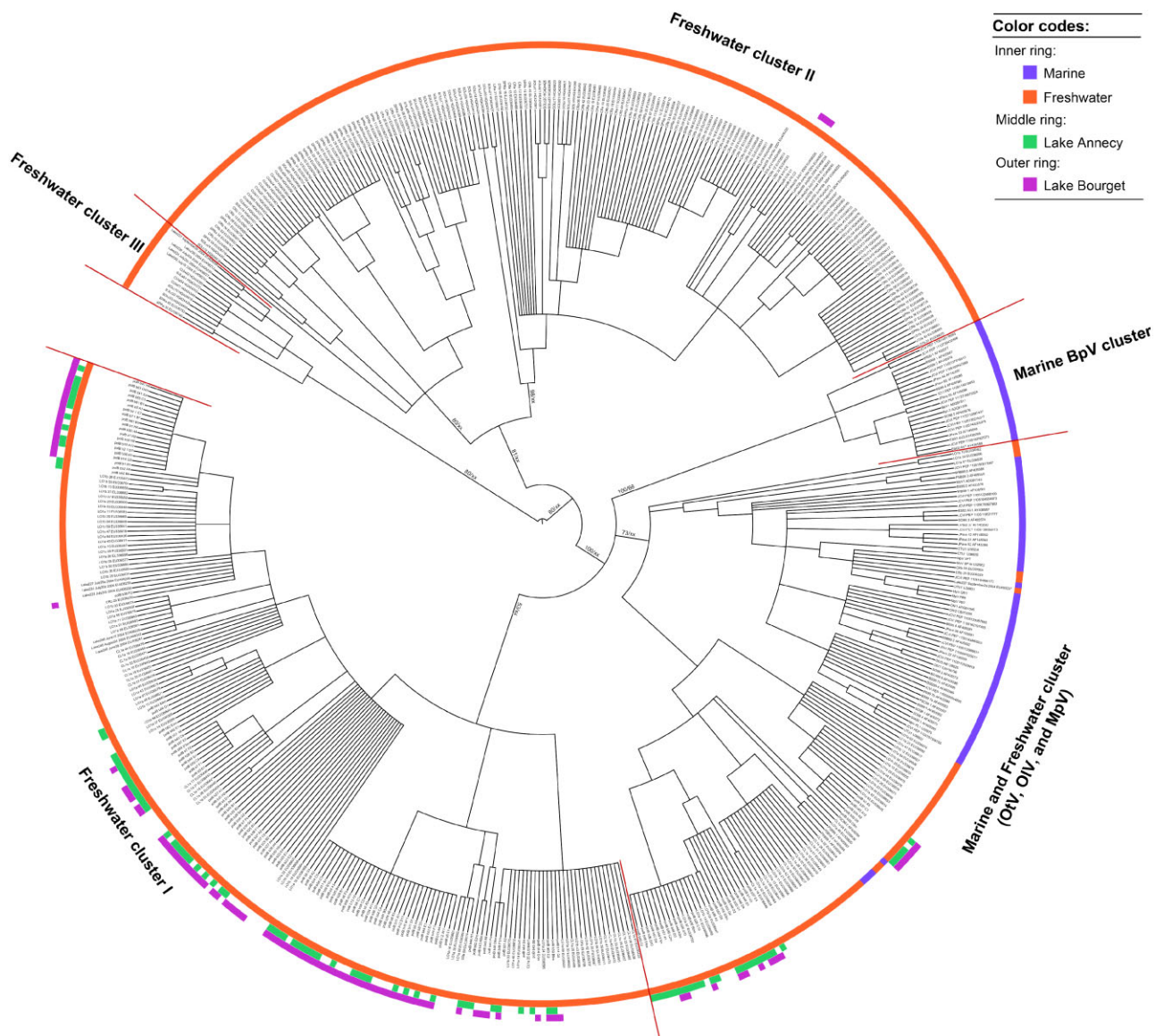
Fig. 2. Bayesian phylogenetic tree based on 94 homologous amino acid positions of family B DNA polymerase-encoding gene (*polB*) from 534 available prasinovirus-like *polB* sequences. The sequences are from Lakes Annecy and Bourget, phycodnavirus isolates, the GOS metagenome database and other different environments, as described in Table 1. The values shown at the nodes of the main branches are the Bayesian inference (BI) clade credibility and maximum likelihood (ML) bootstrap values, and are reported as BI/ML, where 'xx' indicates a value < 50%. The prasinovirus-like *polB* sequences are grouped into five clades, three of which were initially identified by Short and Short (2008): Freshwater Cluster I, Freshwater Cluster II and the Marine and Freshwater Cluster. Freshwater Cluster III and Marine BpV Cluster are new. The three colour rings reflect the habitat type from which the *polB* sequences originated. The inner ring of colour bands shows sequences from the marine (violet) and freshwater (orange) environments. The green colour bands in the middle ring correspond to sequences from Lake Annecy, and the purple colour bands in the outer ring to sequences from Lake Bourget.

unknown phycodnavirus groups. These differences were likely due to the specificity of primers developed to study specific viral populations. The primers AVS1 and AVS2 were initially designed based on chloro-like viruses and are able to amplify *polB* fragments from prasinoviruses and chloroviruses, but not from other phycodnaviruses, such as *Emiliania huxleyi* virus (EhV, *Coccolithovirus*), *Heterosigma akashiwo* virus (HaV, *Raphidovirus*), *Chrysochromulina ericina* virus (CeV, *Prymnesiovirus*),

PoV and *Ectocarpus siliculosus* virus (EsV, *Phaeovirus*) (Short *et al.*, 2011). By contrast, the primer set *mcp*-Fwd/ Rev was designed to amplify two conservative major capsid protein regions of a larger phycodnavirus set that comprises HaV, *Paramecium bursaria Chlorella* virus, PpV (*Phaeocystis poucheti* virus), PoV and CeV (Larsen *et al.*, 2008).

Both gene markers generated sequences attributed to the *Prasinovirus* group. For *polB*, our results are consistent
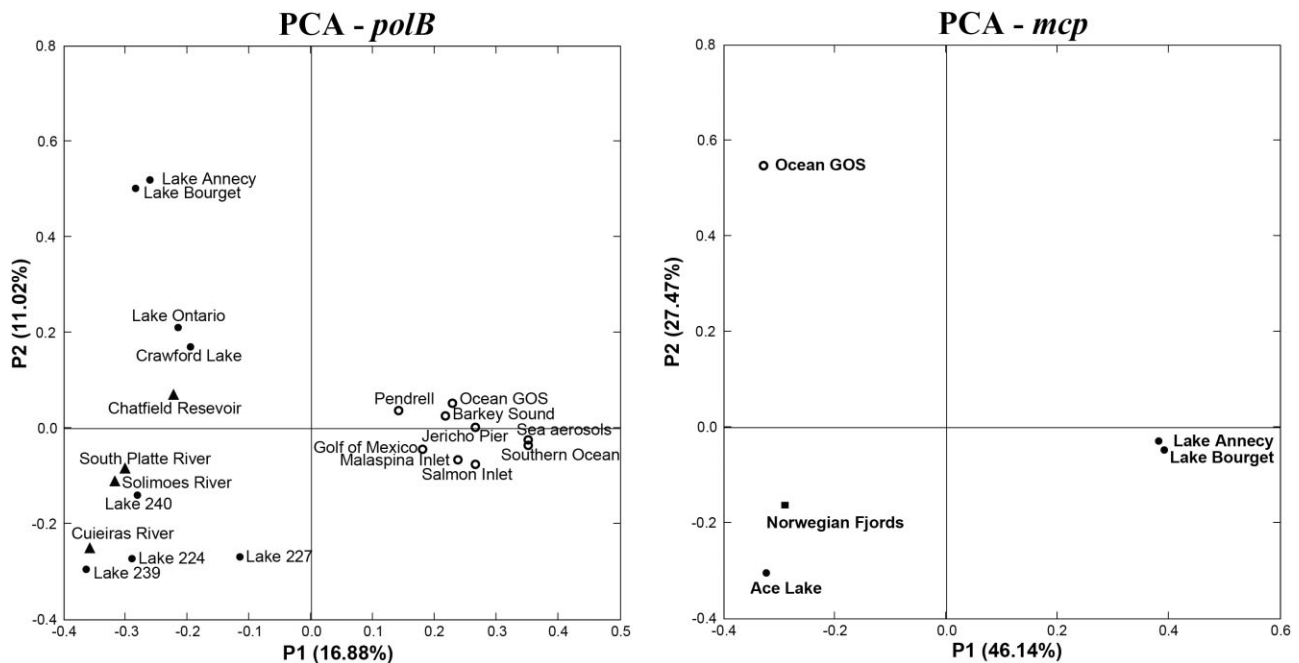
Fig. 3. Principal component analysis (PCA) for *polB* and *mcp* amino acid sequences from different environments using Unifrac distance metric statistical tools available at http://bmf.colorado.edu/unifrac/ (Lozupone and Knight, 2005). The environmental locations from which the sequences originated are shown in Table 1. The Unifrac distance matrix was calculated based on the sequence's branch length of each environment in the phylogenetic trees presented in Fig. 2 and Fig. 4 for *polB* and *mcp* respectively. River and freshwater reservoir ●, lake ▲, fjord ○, marine ■.

with previous studies conducted in both marine and freshwater environments (Chen *et al.*, 1996; Short and Suttle, 2002; 2003; Clasen and Suttle, 2009; Short *et al.*, 2011; Gimenes *et al.*, 2012), which revealed that *polB* sequences are more closely related to marine prainsoviruses than to freshwater chloroviruses (Fig. 1A). By comparison, only 4.4% of the *mcp* sequences, originating from both lakes, grouped with the *Prasinovirus* lineage closely related to MpV, BpV, OlV and OtV. This discrepancy between the two genetic markers for the prasinovirus group is interesting, but little information is available to detect potential hosts. Data on freshwater prasinophyceae are very limited, and we know that Lakes Annecy and Bourget contain few *Pyramimonas* spp., a genus regularly detected in these lakes, based on microscopic counts. These populations are, however, relatively rare, with less than 12 cell ml^{-1} on average detected in 2011 (data not shown). Taib and colleagues (2013), using pyrosequencing of 18S ribosomal RNA (rRNA) amplicons, reported an absence of prasinophyceae-like reads in Lake Bourget, but up to 23 operational taxonomic units (OTUs) belonging to *Prasinophyceae* in three other French lakes. They stated that their study was the first to detect Mamiellales in lakes, while such sequences may constitute the dominant photosynthetic group in the picoplankton 18S rRNA gene clone library in marine surveys, especially in coastal waters and where prasinophytes can account for

45% of the picoeukaryotic community on average when targeted by tyramide signal amplification-fluorescence *in situ* hybridization (TSA-FISH) (Vaulot *et al.*, 2008). It is noteworthy that Mangot *et al.* (2013) also found three prasinophyceae-like OTUs in another alpine lake located in the same eco-area (Lake Geneva). These sequences represented, however, less than 0.01% total reads. Taken together, these results suggest that: (i) freshwater prasinophyceae-like populations are present but rare and (ii) our current knowledge of freshwater prasinophyceae, as reported above, leads us to reject the assumption that all *polB* sequences obtained were related to viruses infecting prasinophyceae. The results for the five prasinovirus-like *mcp* sequences obtained over our year-long study could reasonably reflect such viral diversity in relation to potential prasinophyceae hosts. If this hypothesis is true, it also means that the majority of the prasinovirus-like *polB* sequences that we found could in fact derive from algal groups other than Prasinophyceae. However, they were probably not viruses belonging to *Prymnesiovirus* (CeV and PpV), *Raphidovirus*, *Phaeovirus* or *Coccolithovirus* groups because they did not cluster with these groups. Short and colleagues (2011) speculated that the majority of these prasinovirus-like *polB* sequences may be derived from chlorophyceae or from other closely related phycodnavirus infecting chlorophyta. This conclusion was based on their analysis of diversity in potential
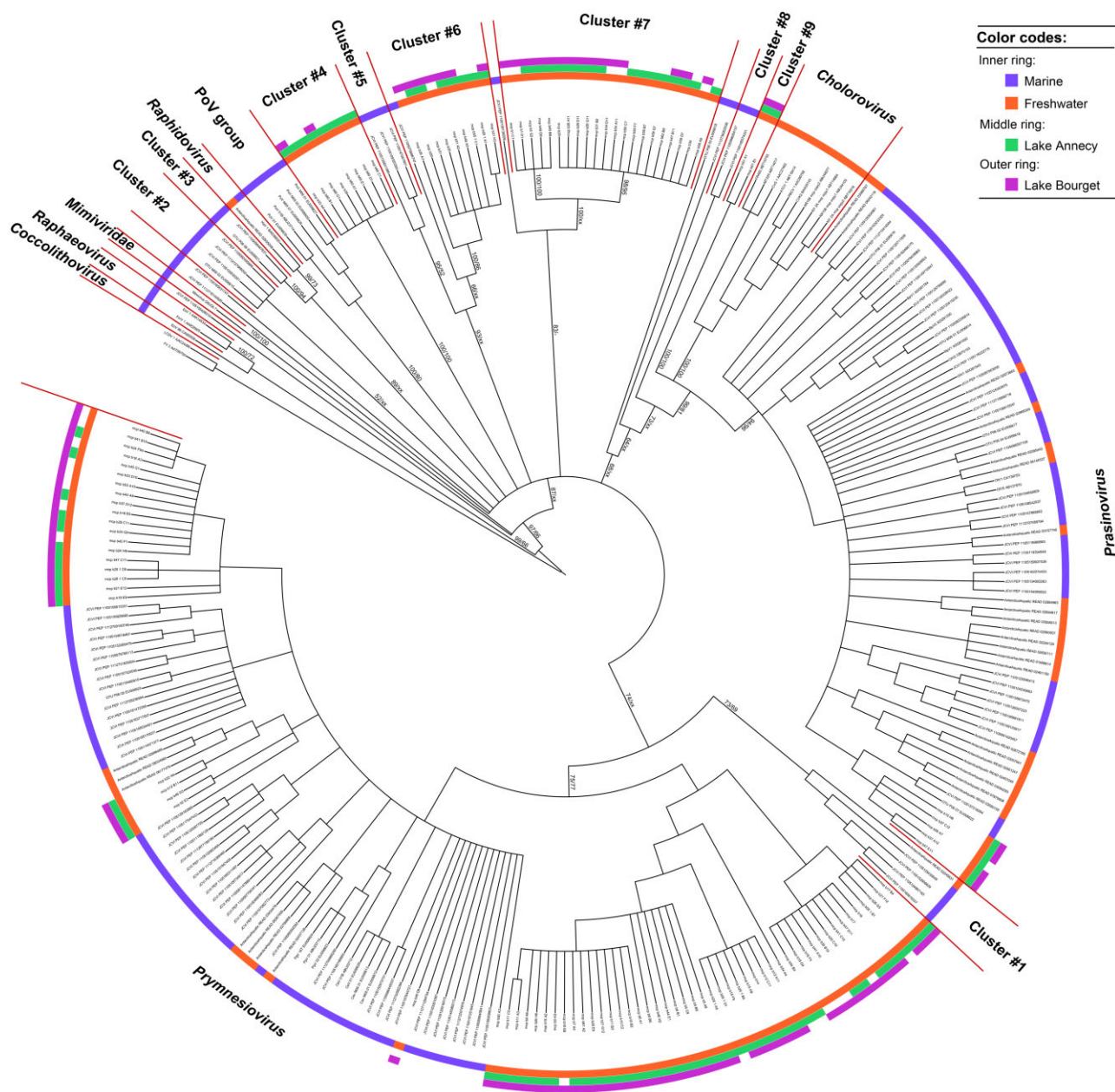
Fig. 4. Bayesian phylogenetic tree based on 97 homologous amino acid positions of major capsid protein-encoding gene (*mcp*) from 290 available *mcp* sequences. The sequences are for Lakes Annecy and Bourget, Norwegian fjords, nucleocytoplasmidic large DNA viruses (NCLDVs) isolates, GOS and Antarctica-Aquatic (Ace Lake) metagenome databases. See legend for Fig. 2.

phytoplanktonic hosts in Lake Ontario, with the relatively diverse *psbA* genotypes associated with chlorophyceae but not prasinophycea. Chlorophyceae are an important component in Lake Annecy and Bourget, accounting for 4.2% and 9.7% of the total annual microalgal abundance in 2011. At least 18 taxa were detected in Annecy, and 35 in Bourget taxa, during this year (Domaizon *et al.*, 2012; Jacquet *et al.*, 2012). Mangot and colleagues (2009), using rRNA probe-based FISH, also reported a high abundance

of chlorophyceae in Lake Bourget, with this class accounting for, on average, 17.9% of all < 5 μm eukaryotes sampled using the same method we did (i.e. over a complete year and from integrated 0–20 m samples). On the other hand, we detected only two chlorovirus-like sequences using *mcp*. This finding may indicate that only a few genotypes of *Chlorophyceae* are potential hosts and/or that in Lakes Annecy and Bourget, these viruses may be infecting *Chlorophyta* members

other than chlorophyceae and prasinophyceae, such as the Charophyceae, among which *Mougeotia* may be important.

We found that 60% of the *mcp* sequences obtained were related to viruses infecting prymnesiophyceae. Lepère and colleagues (2010), using TSA-FISH, reported that prymnesiophyceae could account for as much as 62.8% of total small eukaryotes in the 0–20 m surface layer of Lake Bourget, while only one taxon was unambiguously recognized and counted with microscopy: *Erkenia subaequiciliata*. It should be noted that prymnesiovirus isolates [CeV, PpV, PgV (*Phaeocystis globosa* virus) and/or CbV (*Chrysochromulina brevifilum* virus)] cluster into one single group when using *mcp*. By contrast, they separated into two groups when using *polB*: one consisted of PgV and CbV, the other CeV and PpV, which also clustered with PoV and mimivirus. These results are consistent with the findings of Larsen *et al.* (2008) and suggest that *mcp* may be better at detecting *Prymnesiovirus* than is *polB*, within which gene arrangement or horizontal gene transfer may occur (Larsen *et al.*, 2008; Wilson *et al.*, 2006). Curiously, PoV-01B and other PoV isolates, initially isolated from *P. orientalis* (*Prasinophyceae* and *Pyramimonadales*) did not cluster with viruses infecting Prasinophyceae members of the order Mamiellales (OtV, OlV, BpV and MpV), either for *mcp* or *polB*. This may be because they have different ancestries (Larsen *et al.*, 2008). Yet, PoV-01B was found to be much more closely related to the prymnesioviruses CeV-01B and PpV-01 (Larsen *et al.*, 2008, Short *et al.*, 2011; Fig. S1), and could cross-infect *C. ericina* stain IFM (*Prymnesiophyceae*) (Sandaa *et al.*, 2001). Also, it is characterized by a larger genome size (560 kb) and longer latent period, which make it closer to CeV-01B (510 kb) and PpV-01 (485 kb) than to the prasinoviruses OtV, OlV, BpV and MpV (180 to 200 kb) (Sandaa *et al.*, 2001; Derelle *et al.*, 2008; Nagasaki, 2008; Moreau *et al.*, 2010; Van Etten *et al.*, 2010). These findings likely indicate that PoV has the same ancestry as CeV and PpV. However, the relatively low ML bootstrap value for the *polB* phylogenetic tree (Larsen *et al.*, 2008; Short *et al.*, 2011) and the PoV branching off from prymnesioviruses (CeV, PgV and PpV) in *mcp* Bayesian phylogenetic trees suggest that PoV is a separate group belonging to neither *Prasinovirus* nor *Prymnesiovirus*. To better uncover and understand the evolutionary relationships, we would need a comparative genomic study of these large phycodnaviruses and mimivirus. In short, it appears that *mcp* can detect prynmesioviruses and provide information on their potential prynmesiophyceae hosts.

When analysing *mcp*, we detected three unknown clusters (Clusters number 4, 6 and 7) in our lakes, which contained about 34% of the obtained sequences. The lineage was unidentified and potential hosts were unknown. They could belong to phycodnaviruses other than the six identified groups. Nevertheless, one possible drawback to using *mcp* as a gene marker is the existence of several polyphyletic copies in phycodnavirus genomes (Table S2, Fig. S2), while there is only one *polB* gene per genome (Derelle *et al.*, 2008). Indeed, two to four copies of *mcp* genes can be found in freshwater choloroviruses, six to eight in prasinoviruses and five in the coccolothovirus EhV-86. Because of the lack of (complete) genome sequence, no information for viruses belonging to other groups of phycodnaviruses (e.g. *Prymnesiovirus*, *Raphaeovirus* and *Raphidovirus*) is available. Although a gene marker with polyphyletic copies in the genome could cause confusion on phylogenetic assignment, we circumvented this problem by using the primer set mcp-Fwd/Rev (Larsen *et al.*, 2008), which was designed to target amino acid positions GGQRI and YL^I/$_V$EQ^F/$_L$. This primer set amplifies only one or two monophyletic *mcp* copies from each known chlorovirus (Table S1, Fig. S2). However, because of unknown *mcp* copy numbers as well as their sequences in culture representatives of other phycodnavirus groups (e.g. CeV, PpV, PgV, PoV and HaV), the *mcp* fragments amplified by described primer set were unclear and so may have been either monophyletic or polyphyletic. Therefore, it is possible that the unidentified environmental clusters (Clusters number 4, 6 and 7) also contain polyphyletic *mcp* amplicons of phycodnaviruses belonging to these known groups, in which case some of them could possess several copies targeted by primers mcp-Fwd/Rev. It is hence crucial that researchers obtain the genome sequences of the isolated phycodnaviruses [e.g. CeV, PpV, PgV, PoV, HaV, EsV and *Feldmannia irregularis* virus (FirrV)] and also isolate novel viruses so that we can validate the utility of currently used primers and, eventually, improve them.

In light of the report that mcp-Fwd/Rev fails to amplify MpV-12T (Larsen *et al.*, 2008), we examined all *mcp* copies of available marine-mamiellales-related prasinovirus genomes (OtV1, OtV2, OtV5, OlV1, MpV1, BpV1 and BpV2) (Fig. S2). We found that small modifications on the 5′ end of each primer targeting regions could allow to successfully amplify one copy of the *mcp* gene from marine prasinoviruses (OtV1, OtV2, OtV5, OlV1, MpV1, BpV1 and BpV2), as they shared the amino acid positions (GGQR^I/$_V$ and YL^L/$_V$EQV) with which current primers cannot fully match. We thus designed a new primer set (mcp-Pr-F: GGYGGYCARMGMRTY and mcp-Pr-R: TGIAGYTGYTCRAYIARGTA) targeting these marine prasinoviruses. We did a verification by amplifying some cultures, and a tentative test on our lake samples gave negative results, possibly because of the lack of marine-mamiellales-like prasinoviruses in these lakes. These efforts at least confirmed, once again, that the majority of marine-mamiellales-like *polB* sequences obtained during

the *polB* survey was not related to viruses infecting prasinophyceae. And it is noteworthy that current mcp-Fwd/Rev primers were still able to amplify *mcp* sequences related to prasinoviruses infecting mamiellales from our lakes, as well as from coastal waters (Park *et al.*, 2011). This suggests that there is probably great divergence of prasinovirus-like *mcp* sequences in unknown prasinoviruses found in nature.

Differences in phycodnavirus diversity between environments

When examining a wide range of relatively diverse environments, we noticed that both *polB* and *mcp* sequences derived from marine environments were significantly different to those observed in fresh waters (PCA analysis, $P = 0$). Even for *mcp*, where data are available for far fewer environments than is the case for *polB*, the distribution pattern for Lakes Annecy and Bourget was clearly distinct from that of GOS ecosystems. Phycodnaviruses originating from Ace Lake were not significantly different from those found in the Norwegian fjords ($P = 0.6$), but they clearly differed with respect to the GOS (marine) and alpine lakes (Fig. 3). Our results support making a distinction between the three types of environments: exclusively marine, exclusively freshwater and freshwater/glacier water invaded by seawater (i.e. Norwegian fjords and Ace Lake; Lauro *et al.*, 2011; Larsen *et al.*, 2008). This suggests that water salinity may play a key role in driving the diversification and selection of phycodnaviruses among their hosts. The fact that distinct phycodnavirus populations are maintained in marine versus freshwater habitats may be explained by differing host communities and infrequent marine–freshwater transition between the two types of environment. Gene flow may have been restrained because of a physico-chemical (e.g. salinity gradient) and/or ecological (e.g. antagonism to biological invasion) barrier, which would prevent cross-colonization of both viruses and microbial hosts between ecosystems (Logares *et al.*, 2009). It is noteworthy, however, that the metagenomic analysis carried out by Lauro and colleagues (2011) suggested the possibility of cross-colonization with marine immigrants in the Ace Lake. There, these authors detected indeed some microbial and viral community structures similar to those found in marine surface waters, but there seemed to be strong local selection against these immigrants as species richness was one order of magnitude lower in the lake. Among invaders, abundant marine-like *Mantoniella* (*Prasinophyceae*) were observed, and these could be the host for phycodnaviruses (Lauro *et al.*, 2011). This may explain our finding in the phylogenetic analysis that 71% of *mcp* sequences obtained from Ace Lake were clustered in the prasinovirus group (Fig. 4), which is almost 16-fold higher than in alpine lakes. These

data suggest that marine phycodnavirus immigrants have adapted to freshwater conditions. Such a marine–freshwater transition must have occurred about 7000 years ago, when seawater invaded Ace Lake (Lauro *et al.*, 2011). It explains the significant difference in phycodnavirus sequences between this Lake and either GOS ($P = 0$) or alpine lakes ($P = 0$). The evolutionary adaptation of viruses in this lacustrine ecosystem, as proposed by Logares and colleagues (2009), could result from (i) the co-evolution of viruses and hosts (like *Mantoniella*), or (ii) individual evolution via the cross-infecting of lake inhabitants. Short (2012) reported that some phycodnaviruses can display intraspecies (but not interspecies) host specificity in infectivity. It is therefore possible that marine-derived phycodnaviruses may have cross-infected some lake species; however, such events have not yet been detected.

Genetic exchange in phycodnaviruses between freshwater habitats, typically between rivers and lakes, has also been proposed recently. Gimenes and colleagues (2012) revealed less divergence in *polB* sequences among freshwater environments in the Amazon basin than between freshwater and marine ecosystems. Our results are in agreement with these findings because the freshwater sites all clustered together. However, the Unifrac *P*-test revealed that *polB* sequences were significantly different between lakes and rivers ($P < 0.01$), and this difference was clearly less for marine environments. These results corroborate our finding that 93.8% of *polB* sequences in Freshwater Cluster II originated from rivers, and Freshwater Cluster I contained exclusively sequences derived from lakes (Fig. 2). The genetic differentiation of phycodnaviruses between lake and river may also be due to differences in host communities between these habitats, which are in turn linked to differences in hydrodynamics that affect a number of parameters, including turbulence, water residence, light and nutrient availability. But rivers and lakes are still freshwater, and the physical barrier described above may be easier to cross than the salinity barrier, thereby allowing for river–lake transition (gene flow). When analysing all sequences obtained for either *mcp* or *polB*, phycodnaviruses were shown to be prevalent in Lake Annecy and Lake Bourget, and we found no significant genetic diversity differences between the two ecosystems (Fig. 3, $P = 1$). As is the case for other examined freshwater environments, the phycodnaviruses of Lakes Annecy and Bourget, which share a common glacial origin and are situated in the same eco-region, showed the expected differences *vis-à-vis* marine environments ($P = 0$). Such differences are likely due to host diversity and biology because the phytoplankton community composition in these lakes is indeed different to that characteristic of the marine environment. As shown by Snyder and colleagues (2007), it is possible, however, that immigrants

have been introduced into Lakes Annecy and Bourget through rainfall or air dispersal.

Conclusions

In this study, we investigated eukaryotic algal virus diversity in two large and deep European peri-alpine lakes using polymerase chain reaction (PCR)-based approaches with primers targeting two different gene markers. Our analysis highlighted that AVS1/AVS2 were limited in their capacity to amplify *polB* sequences of phycodnaviruses other than viruses infecting chlorophyta. By contrast, primers targeting the major capsid protein-encoding gene (*mcp*) provided new insights and a complementary view on phycodnavirus diversity by detecting diverse sequences related to non-prasinovirus-like phycodnaviruses. Although the numerous polyphyletic gene copies in some phycodnavirus genomes may be a drawback to using *mcp*, the primers are useful and merit validation in different environments and improvement through obtaining more isolated viruses and genomes. By contrast, because there is only a single copy in phycodnavirus, *polB* is advantageous for culture-free-based diversity assays and the quantitative PCR quantification of viral particles, but here again, we need new primers to explore a broader range of phycodnaviruses. Both host community investigation and novel virus isolation are crucial to achieving better assessment capabilities and phylogenetic assignment/mapping, no matter what gene marker is employed. Based on current available *polB* or *mcp* sequences, our analysis provided evidence of significant differences between marine and freshwater environments, and between different types of freshwater environments (river vs. lake).

Experimental procedures

Sample collection and processing

Water samples were collected once or twice each month (every 3–4 weeks on average) between January and November 2011 at reference stations on Lakes Annecy (GL) and Bourget (point B), corresponding to the deepest part of the lake. The main characteristics of these ecosystems are described elsewhere (Personnic *et al.*, 2009; Berdjeb *et al.*, 2011). We collected 40 l, integrating the water column from surface to 20 m depth, using an electric pump and appropriate tubing. This water was stored in polycarbonate flasks placed at 4°C in the dark. We obtained 14 samples for Lake Annecy and 18 for Lake Bourget. A few hours after sampling, 20 l were prefiltered through a 60 μm nylon mesh, and then through filters with 142 mm diameter and 1 μm pore size (Millipore, Bedford, MA, USA). The sample was concentrated to a final volume of 200–250 ml using a Millipore spiral cartridge with a molecular weight cut-off of 30 000 Da (regenerated cellulose, PLTK Prep/scale TFF, 1 ft^2). The < 1 μm fraction obtained was further filtered through filters with 47 mm diameter and 0.45 μm pore size (Millipore) in order to remove any remaining bacteria. The sample was then checked using flow cytometry, which showed no bacterial contamination (not shown). The < 0.45 μm viral concentrate (VC) was stored at −20°C until further processing.

PCR amplification and DGGE

Prior to running the PCR, VCs were treated using the freeze-thaw method described by Short and Short (2008), which consists of three repetitions of heating for 3 min at 95°C, followed by freezing at −20°C until the liquid becomes solid. To obtain a broad representative sample of sequences from the environment (including rare ones) and to avoid the interference of the GC clamp on natural samples, prior to the DGGE analysis, we conducted the PCR in two stages as recommended by Short and Suttle (2002). The first stage used the treated VC as the template, with the primer set without the GC clamp. A second stage was then performed on the product of the first stage using the GC clamp-containing primer set (i.e. with 40 nt GC clamp attached to the 5′ of forward primer). The PCRs were carried out on the DNA Thermal Cycler T-Professional (Biometra, Göttingen, Germany) to amplify the family B DNA polymerase-encoding gene *polB* using the primer set AVS 1/2 (Chen and Suttle, 1995) and the major capsid protein-encoding gene *mcp* using the primer set mcp-Fwd/Rev (Larsen *et al.*, 2008). Briefly, for all primer sets, the 25 μl reaction mix contained 1 X PCR buffer, 4 mM MgCl$_2$, 200 μM of each deoxyribonucleotide triphosphate (dNTP), 0.4 μM of each primer, 0.5 U of Platinum Taq DNA Polymerase (Invitrogen, Carlsbad, CA, USA) and 1 μl of treated VC (i.e. viral DNA). The typical programme for the first PCR stage was 15 min virion lysing and denaturation at 95°C, followed by 34 cycles of denaturation at 95°C for 30 s, annealing for 30 s, extension at 72°C for 45 s and a final extension at 72°C for 5 min. The programme for the second PCR stage was 5 min denaturation at 95°C, followed by 24 cycles of denaturation at 95°C for 30 s, annealing for 30 s, extension at 72°C for 45 s and a final extension at 72°C for 5 min. Further to optimization tests, we used annealing temperatures of 51°C and 45°C for AVS 1/2 and mcp-Fwd/Rev respectively.

The DGGE was conducted in 6% polyacrylamide gels with an optimized linear denaturing gradient (100% denaturant is defined as 7 M urea and 40% deionized formamide). The linear denaturing gradient was optimal at 40–70% and 45–70% for amplicons of AVS 1/2 and mcp-Fwd/Rev respectively. Twenty microlitres of PCR products (corresponding to 210–270 ng DNA) were loaded into wells with 5 μl of 5 X loading buffer [12.5% ficoll, 25 mM tris, 5 mM Ethylenediaminetetraacetic acid (EDTA) (pH 8.0), 0.5% Sodium dodecyl sulfate (SDS), 0.1% (wt/vol) xylene cyanol and 0.1% (wt/vol) bromophenol blue]. Electrophoresis was carried out for 16 h in 1 X TAE buffer (pH 7.4) (40 mM Tris-base, 20 mM sodium acetate, 1 mM EDTA) at 120 V and a constant temperature of 60°C using the CBS-DGGE 2000 system (C.B.S. Scientific, San Diego, CA, USA). Gels were stained in a 2 X SYBR Green I (Molecular Probles, Invitrogen) solution for 45 min, visualized on a UV transilluminator (Tex-35 M, Bioblock Scientific, Illkirch, France) and photographed with GelDoc (Bio-Rad, Hercules, CA, USA). DGGE banding patterns were analysed using the GelCompare II software package (Applied Maths, Kortrijk, Belgium) as described elsewhere (Berdjeb *et al.*, 2011).

DNA purification, cloning and sequencing

The DNA of each representative DGGE band was eluted from the gel slice, after its excision, by adding 100 μl sterile 1 X TAE buffer and heating at 95°C for 15 min. Three microlitres of eluted DNA served as template in a 22 μl PCR mixture using the corresponding primer set. The PCRs were performed with the same conditions as the first PCR stage described above. The amplicons were first verified by electrophoresis in a 1.5% agarose gel, then purified using the Illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare, Little Chalfont, UK) to finally be cloned into pCR4-TOPO vectors using the TOPO TA Cloning Kit (Invitrogen). Randomly selected clones were sent for sequencing to GATC Biotech (Constance, Germany).

Phylogenetic analysis

After cleaning and correcting sequences using BioEdit 7.0.5.3 (Hall, 1999), we obtained 163 and 115 non-redundant sequences for *polB* and *mcp*, respectively, for Lakes Annecy and Bourget. These sequences have been deposited in GenBank under the reference accession numbers given in Table 1. All sequences were translated to amino acids and aligned with culture phycodnavirus representatives (Table S1) using MAFFT version 6 (Katoh *et al.*, 2002). Multiple alignments were then curated using Gblocks (Castresana, 2000), employing a less stringent option that allowed for gaps inside the final blocks. We constructed the phylogenies using both the BI and ML methods. BI was conducted using MrBayes 3.2.1 (Ronquist *et al.*, 2012), with two runs, four chains, 10^6 generations, sampling every 100 generations, a burn-in value of 25% and mixed models of amino acid substitution. The ML phylogeny was constructed using PhyML 3.0 (Guindon and Gascuel, 2003), with 100 bootstrap replicates, and with the best models of amino acid substitution and rate heterogeneity. The best models for each aligned-sequence dataset were determined using MEGA5 (Tamura *et al.*, 2011). They were the Jones-Taylor-Thornton (JTT) model and gamma-distributed substitution rates for *polB*, and rtREV (Dimmic *et al.*, 2002) model, and gamma-distributed substitution rate for *mcp*. The phylogenetic trees are presented in Fig. 1.

Phylogenetic and statistical analyses of metadata

We used selected inferred amino acid sequences of *polB* and *mcp* genes as queries in our blasts against the GOS (Rusch *et al.*, 2007; Williamson *et al.*, 2008) and Antarctica-Aquatic (Ace Lake; Lauro *et al.*, 2011) microbial metagenome databases. The databases constitute two components of the larger CAMERA database (https://portal.camera.calit2.net/; Seshadri *et al.*, 2007). The BLAST *E-value* was set at ≤ 10^{-20}. We obtained 400 and 733 hits in GOS, and 400 and 82 hits in Antarctica-Aquatic for *polB* and *mcp* respectively. We then removed duplicated sequences and also those that did not fully overlap with our sequences. In the end, we recruited 25 and 100 sequences from GOS, and 0 and 31 sequences from the Antarctica-Aquatic database for *polB* and *mcp* respectively. These non-redundant sequences, together with sequences obtained from different environments (Table 1) and from culture representatives (Table S1), constituted our meta-dataset and were subsequently included in the phylogenetic analysis, as previously described. The 'meta-phylogenetic' tree for *mcp* sequences is shown in Fig. 4. In our study, as reported elsewhere (Chen *et al.*, 1996; Short and Suttle, 2002; 2003; Clasen and Suttle, 2009; Short *et al.*, 2011; Gimenes *et al.*, 2012), all or nearly all *polB* sequences were grouped within the prasinovirus group (Fig. 1); hence, we removed the outer groups in a separate analysis and constructed the 'meta-phylogenetic' tree (Fig. 2). This improved the divergence and resulted in a better association among closely related prasinovirus-like sequences. It is noteworthy that a viral metagenomic study has been recently conducted in Lake Bourget (http://metavir-meb.univ-bpclermont.fr/; Roux *et al.*, 2012). However, no sequence reads were found to be similar to our *mcp* or *polB* gene sequences when we blasted against their virome (e-value = 0.1) using our PCR-generated sequences as queries. It turns out that the sequencing depth used in this new study was not sufficient to generate these relatively rare sequences/biospheres/species.

To evaluate whether *polB* or *mcp* clustering patterns revealed in the phylogenetic reconstructions reflect the environment from which the samples were taken, we carried out statistical analyses using the Unifrac distance metric statistical tools available at http://bmf.colorado.edu/unifrac/ (Lozupone and Knight, 2005). We used the unweighted Unifrac option in order to compare community composition based on presence/absence importance (i.e. on qualitative data). This tool measures the distance between two communities by calculating the fraction of the branch length in a phylogenetic tree (Lozupone *et al.*, 2007). In brief, we used the Bayesian phylogenetic tree and a file mapping sequence labels to their habitats and/or the environmental categories (e.g. marine vs. freshwater) as input for each analysis (*polB* or *mcp*). We then generated the Unifrac distance matrix for communities of defined environmental locations or categories, based on which *P*-test and PCA were conducted.

References

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:** 2121e2131.

Bratbak, G., Egge, J.K., and Heldal, M. (1993) Viral mortality of the marine alga *Emiliana huxleyi* (Haptophyceae) and termination of algal blooms. *Mar Ecol Prog Ser* **93:** 39–48.

Bellec, L., Grimsley, N., Moreau, H., and Desdevises, Y. (2009) Phylogenetic analysis of new prasinoviruses (phycodnaviridae) that infect the green unicellular algae *Ostreococcus, Bathycoccus and Micromonas*. *Environ Microbiol Rep* **1:** 114–119.

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 Environ Microbiol* **73:** 7629–7641.

Berdjeb, L., Ghiglione, J.-F., and Jacquet, S. (2011) Bottom-up *vs.* top-down factors regulating the bacterial community structure in two peri-alpine lakes. *Appl Environ Microbiol* **77:** 3591–3599.

Brussaard, C.P.D. (2004) Viral control of phytoplankton populations – a review. *J Eukaryot Microbiol* **51:** 125–138.

Brussaard, C.P.D., Wilhelm, S.W., Thingstad, F., Weinbauer, M.G., Bratbak, G., Heldal, M., *et al.* (2008) Global-scale processes with a nanoscale drive: the role of marine viruses. *ISME J* **2:** 575–578.

Castresana, J. (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* **17:** 540–552.

Chen, F., and Suttle, C.A. (1995) Amplification of DNA polymerase gene fragments from viruses infecting microalgae. *Appl Environ Microbiol* **61:** 1274–1278.

Chen, F., Suttle, C.A., and Short, S.M. (1996) Genetic diversity in marine algal virus communities as revealed by sequence analysis of DNA polymerase genes. *Appl Environ Microbiol* **62:** 2869–2874.

Clasen, J.L., and Suttle, C.A. (2009) Identification of freshwater *Phycodnaviridae* and their potential phytoplankton hosts, using DNA pol sequence fragments and a genetic-distance analysis. *Appl Environ Microbiol* **75:** 991–997.

Danovaro, R., Dell'Anno, A., Corinaldesi, C., Magagnini, M., Noble, R., Tamburini, C., *et al.* (2008) Major viral impact on the functioning of benthic deep-sea ecosystems. *Nature* **454:** 1084–1087.

Derelle, E., Ferraz, C., Escande, M.L., Eychenié, S., Cooke, R., Piganeau, G., *et al.* (2008) Life-cycle and genome of OtV5, a large DNA virus of the pelagic marine unicellular green alga *Ostreococcus tauri*. *PLoS ONE* **3:** e2250.

Dimmic, M.W., Rest, J.S., Mindell, D.P., and Goldstein, R.A. (2002) rtREV: an amino acid substitution matrix for inference of retrovirus and reverse transcriptase phylogeny. *J Mol Evol* **55:** 65–73.

Domaizon, I., Lainé, L., Lazzarotto, J., Perga, M.E., and Rimet, F. (2012) Suivi de la qualité des eaux du lac d'Annecy 2011. Rapport SILA-INRA, 92 pp.

Dunigan, D.D., Fitzgerald, L.A., and Van Etten, J.L. (2006) Phycodnaviruses: a peek at genetic diversity. *Vir Res* **117:** 119–132.

Gimenes, M.V., Zanotto, P.M., Suttle, C.A., Da Cunha, H.B., and Mehnert, D.U. (2012) Phylodynamics and movement of phycodnaviruses among aquatic environments. *ISME J* **6:** 237–247.

Guindon, S., and Gascuel, O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52:** 696–704.

Hall, T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* **41:** 95–98.

Iyer, L.M., Balaji, S., Koonin, E.V., and Aravind, L. (2006) Evolutionary genomics of nucleo-cytoplasmic large DNA viruses. *Vir Res* **117:** 156–184.

Jacquet, S., Heldal, M., Iglesias-Rodriguez, D., Larsen, A., Wilson, W.H., and Bratbak, G. (2002) Flow cytometric analysis of an *Emiliana huxleyi* bloom terminated by viral infection. *Aquat Microb Ecol* **27:** 111–124.

Jacquet, S., Barbet, D., Cachera, S., Caudron, A., Colon, M., Girel, C., *et al.* (2012) Suivi environnemental des eaux du lac du Bourget pour l'année 2011. Rapport INRA-CISALB, 220 pp.

Jacquet, S., Domaizon, I., and Anneville, O. (in press) Long-term trend of physico-chemical and biological indicators of water-quality and functioning of large peri-alpine lakes (Lakes Geneva, Annecy and Bourget): a comparative study of ecosystem trajectories during re-oligotrophication. *Arch Sci* **65:** 225–242.

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.

Larsen, A., Fonnes Flaten, G.A., Sandaa, R.A., Castberg, T., Thyrhaug, R., Erga, S.R., *et al.* (2004) Spring phytoplankton bloom dynamics in Norwegian coastal waters: microbial community succession and diversity. *Limnol Oceanogr* **49:** 180–190.

Larsen, J.B., Larsen, A., Bratbak, G., and Sandaa, R.A. (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.

Lauro, F.M., DeMaere, M.Z., Yau, S., Brown, M.V., Ng, C., Wilkins, D., *et al.* (2011) An integrative study of a meromictic lake ecosystem in Antarctica. *ISME J* **5:** 879–895.

Lepère, C., Masquelier, S., Mangot, J.F., Debroas, D., and Domaizon, I. (2010) Vertical structure of small eukaryotes in three lakes that differ by their trophic status: a quantitative approach. *ISME J* **4:** 1509–1519.

Logares, R., Bråte, J., Bertilsson, S., Clasen, J.L., Shalchian-Tabrizi, K., and Rengefors, K. (2009) Infrequent marine – freshwater transitions in the microbial world. *Trends Microbiol* **17:** 414–422.

Lozupone, C., and Knight, R. (2005) UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* **71:** 8228–8235.

Lozupone, C.A., Hamady, M., Kelley, S.T., and Knight, R. (2007) Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. *Appl Environ Microbiol* **73:** 1576–1585.

Mangot, J.F., Lepère, C., Bouvier, C., Debroas, D., and Domaizon, I. (2009) Community structure and dynamics of small eukaryotes targeted by new oligonucleotide probes: new insight into the lacustrine microbial food web. *Appl Environ Microbiol* **75:** 6373–6381.

Mangot, J.F., Domaizon, I., Taib, N., Marouni, N., Duffaud, E., Bronner, G., *et al.* (2013) Short-term dynamics of diversity patterns: evidence of continual reassembly within lacustrine small eukaryotes. *Environ Microbiol* **15:** 1745–1758.

Monier, A., Larsen, J.B., Sandaa, R.A., Bratbak, G., Claverie, J.M., and Ogata, H. (2008) Marine mimivirus relatives are probably large algal viruses. *Virol J* **5:** 12.

Moreau, H., Piganeau, G., Desdevises, Y., Cooke, R., Derelle, E., and Grimsley, N. (2010) Marine prasinovirus genomes show low evolutionary divergence and acquisition of protein metabolism genes by horizontal gene transfer. *J Virol* **84:** 12555–12563.

Nagasaki, K. (2008) Dinoflagellates, diatoms, and their viruses. *J Microbiol* **46:** 235–243.

Park, Y., Lee, K., Lee, Y.S., Kim, S.W., and Choi, T.J. (2011) Detection of diverse marine algal viruses in the South Sea regions of Korea by PCR amplification of the DNA polymerase and major capsid protein genes. *Virus Res* **159:** 43–50.

Personnic, S., Domaizon, I., Dorigo, U., Berdjeb, L., and Jacquet, S. (2009) Seasonal and spatial variability of virio, bacterio- and picophytoplanktonic abundances in three peri-alpine lakes. *Hydrobiologia* **627:** 99–111.

Ronquist, F., Teslenko, M., Van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., *et al.* (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* **61:** 539–542.

Roux, S., Enault, F., Robin, A., Ravet, V., Personnic, S., Theil, S., *et al.* (2012) Assessing the diversity and specificity of two freshwater viral communities through metagenomics. *PLoS ONE* **7:** e33641.

Rusch, D.B., Halpern, A.L., Sutton, G., Heidelberg, K.B., Williamson, S., Yooseph, S., *et al.* (2007) The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol* **5:** e77.

Sandaa, R.-A., Heldal, M., Castberg, T., Thyrhaug, R., and 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.

Schroeder, D.C., Oke, J., Hall, M., Malin, G., and Wilson, W.H. (2003) Virus succession observed during an *Emiliana huxleyi* bloom. *Appl Environ Microbiol* **69:** 2484–2490.

Seshadri, R., Kravitz, S.A., Smarr, L., Gilna, P., and Frazier, M. (2007) CAMERA: a community resource for metagenomics. *PLoS Biol* **5:** e75.

Short, C.M., Rusanova, O., and Short, S.M. (2011) Quantification of virus genes provides evidence for seed-bank populations of phycodnaviruses in Lake Ontario Canada. *ISME J* **5:** 810–821.

Short, S.M. (2012) The ecology of viruses that infect eukaryotic algae. *Environ Microbiol* **14:** 2253–2271.

Short, S.M., and Short, C.M. (2008) Diversity of algal viruses in various North American freshwater environments. *Aquat Microb Ecol* **51:** 13–21.

Short, S.M., and Suttle, C.A. (2002) Sequence analysis of marine virus communities reveals that groups of related algal viruses are widely distributed in nature. *Appl Environ Microbiol* **68:** 1290–1296.

Short, S.M., and Suttle, C.A. (2003) Temporal dynamics of natural communities of marine algal viruses and eukaryotes. *Aquat Microb Ecol* **32:** 107–119.

Snyder, J.C., Wiedenheft, B., Lavin, M., Roberto, F.F., Spuhler, J., Ortmann, A.C., *et al.* (2007) Virus movement maintains local virus population diversity. *Proc Natl Acad Sci USA* **104:** 19102–19107.

Suttle, C.A. (2005) Viruses in the sea. *Nature* **437:** 356–361.

Taib, N., Mangot, J.-F., Domaizon, I., Bronner, G., and Debroas, D. (2013) Phylogenetic affiliation of SSU rRNA genes generated by massively parallel sequencing: new insights in the freshwater protist diversity. *PLoS ONE* **8:** e58950.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28:** 2731–2739.

Thomas, R., Jacquet, S., Grimsley, N., and Moreau, H. (2012) Strategies and mechanisms of resistance to viruses in photosynthetic aquatic microorganisms. *Adv Limnol Oceanogr* **3:** 1–15.

Tomaru, Y., Shirai, Y., and Nagasaki, K. (2008) Ecology, physiology and genetics of a phycoDNAvirus infecting the noxious bloom-forming raphidophyte *Heterosigma akashiwo*. *Fish Sci* **74:** 701–711.

Van Etten, J.L., and Meints, R.H. (1999) Giant viruses infecting algae. *Annu Rev Microbiol* **53:** 447–494.

Van Etten, J.L., Graves, M.V., Muller, D.G., Boland, W., and Delaroque, N. (2002) Phycodnaviridae – large DNA algal viruses. *Arch Virol* **147:** 1479–1516.

Van Etten, J.L., Lane, L.C., and Dunigan, D.D. (2010) DNA viruses: the really big ones (giruses). *Annu Rev Microbiol* **64:** 83–99.

Vaulot, D., Eikrem, W., Viprey, M., and Moreau, H. (2008) The diversity of small eukaryotic phytoplankton (≤3 μm) in marine ecosystems. *FEMS Microbiol Rev* **32:** 795–820.

Weinbauer, M.G. (2004) Ecology of prokaryotic viruses. *FEMS Microbiol Rev* **28:** 127–181.

Williamson, S.J., Rusch, D.B., Yooseph, S., Halpern, A.L., Heidelberg, K.B., Glass, J.I., *et al.* (2008) The Sorcerer II Global Ocean Sampling Expedition: metagenomic characterization of viruses within aquatic microbial samples. *PLoS ONE* **3:** e1456.

Wilson, W.H., Schroeder, D.C., Ho, J., and Canty, M. (2006) Phylogenetic analysis of PgV-102P, a new virus from the English Channel that infects *Phaeocystis globosa*. *J Mar Biolog Assoc UK* **86:** 485–490.

Wilson, W.H., Van Etten, J.L., and Allen, M.J. (2009) The Phycodnaviridae: the story of how tiny giants rule the world. *Curr Top Microbiol Immunol* **328:** 1–42.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Fig. S1. Bayesian phylogeny inferred from 279 homologous amino acid positions of complete protein sequences of DNA polymerase-encoding gene (*polB*) from 23 members of the Phycodnaviridae. The tree was rooted using sequences of frog virus 3 (FV-3) and Lymphocystis disease virus 1 (LCDV-1) of the Iridoviridae. Values shown at the nodes of the main branches are the Bayesian inference (BI) clade credibility and maximum likelihood (ML) bootstrap values, and are reported as BI/ML, where 'xx' indicates values < 50%.

Fig. S2. Bayesian phylogenetic inference over each major capsid protein-encoding gene (*mcp*) copies from *Chlorovirus*

(2–4 copies) and *Prasinovirus* (6–8 copies). The phylogeny was based on the aligned complete major capsid protein sequences from the members of the Phycodnaviridae, among which only one *mcp* gene copy was included in the analysis as genome sequences for PpV-01, CeV-01B, PoV-01B, HaV-1, Mimivirus, FirrV-1, EsV-1, and EhV-86 are missing. The tree was rooted using sequences of FV-3 and LCDV-1 of the Iridoviridae. Values shown at the nodes of the main branches are for Bayesian inference (BI) clade credibility. The sequences selected as culture representatives shown in Table S1 are labelled by colours. The *mcp* sequences containing both forward and reverse primer's amino acid positions (GGQR^I/$_V$ and YL^I/$_V$EQ respectively) are marked in violet.

Table S1. *mcp* and *polB* sequences of nucleocytoplasmic large DNA viruses (NCLDVs) isolates used for phylogenetic analysis.

Table S2. Summary of *mcp* and *polB* gene copies in some *Phycodnaviridae.*