Freshwater Biology

Freshwater Biology (2014) **59,** 1577–1595 doi:10.1111/fwb.12365

Differing assemblage composition and dynamics in T4-like myophages of two neighbouring sub-alpine lakes

XU ZHONG AND STEPHAN JACQUET
INRA, UMR CARRTEL, Thonon-les-Bains Cedex, France

SUMMARY

1. Bacteriophages play an important role in plankton population dynamics and biogeochemical cycling, but their community dynamics and diversity are still poorly known, especially in fresh waters.
2. We conducted a 1-year investigation of the T4-like bacteriophages in the surface waters of two Western European peri-alpine lakes (lakes Annecy and Bourget), using fingerprinting and cloning-sequencing approaches targeting the major capsid protein *g23* gene.
3. Our results suggest that T4-like bacteriophages are diverse (we found several new sequences) and differed substantially between the two lakes. Further, these phages displayed seasonal patterns with marked shifts in community composition.
4. Examining the dynamics of some individual constituents (e.g. DGGE bands), we found that (i) the majority of these myoviruses were rare (mean relative abundance <1%) and only a few were abundant (mean relative abundance >5%); (ii) the dominance of the most abundant DGGE bands changed throughout the year and their dynamics were very different between the two lakes; (iii) only a few bands (e.g. <20%) were detectable throughout the year and their proportion was higher in the oligo-mesotrophic Lake Bourget.
5. Our results suggest a highly dynamic T4-like myoviral assemblage. T4-like-viruses seem to conform to a standard model of community organisation and a common type of species abundance curve, with mainly rare organisms that occasionally may become abundant (termed a 'seed bank' for virus assemblages).

Keywords: community structure and dynamics, diversity, lakes, T4-like myoviruses

Introduction

Viruses are the most abundant biological entities on Earth. Through the direct mortality of specific host species and the release of cellular nutrients, viruses influence the succession of individual populations of microorganisms, biogeochemical cycles and, ultimately, microbial community structure (Weinbauer, 2004; Suttle, 2007; Wilhelm & Matteson, 2008). Metagenomic studies have now revealed that viruses of the family *Myoviridae* (King *et al.*, 2011), a subset of the tailed dsDNA viruses (*Caudiviridae*), dominate the virioplankton in a variety of aquatic environments (Breitbart *et al.*, 2002; DeLong *et al.*, 2006; Yooseph *et al.*, 2007; Fancello *et al.*, 2013). T4-like myoviruses (*Tevenviridae*) belong to the *Myoviridae* and share morphological similarities with the enterobacteriophage T4. Such T4-like myoviruses have been isolated from a variety of hosts, ranging from the phylum *Proteobacteria* (such as *Escherichia coli*, *Shigella*, *Vibrio*, *Aeromonas*, *Klebsiella*, *Acinetobacter*, *Pseudomonas*, *Salmonella*, *SAR11*) to *Cyanobacteria*, with genome sizes ranging between approximately 150 and 250 kb (Petrov *et al.*, 2010; Martis *et al.*, 2013; Zhao *et al.*, 2013). Based on the genetic divergence of their structural proteins (*g18*, *g19* and *g23*), T4-like myoviruses have been classified into four subgroups (T-evens, Pseudo T-evens, Schizo T-evens and Exo T-evens) (Desplats & Krisch, 2003). After examination of major capsid protein (MCP, encoded by the *g23* gene) sequences obtained from the GOS metagenome (about 1400 sequences), Comeau & Krisch (2008) assembled MCP homologues into three

Correspondence: Stéphan Jacquet, INRA, UMR CARRTEL, Station d'Hydrobiologie Lacustre, 74203 Thonon cedex, France.
E-mail: sjacquet@thonon.inra.fr

major clusters: Far T4 (including the *Rhodothermus* phage RM378), Near T4 (including T-evens and PseudoT-evens, SchizoT-evens) and Cyano T4 (including Exo-T4).

Filée *et al.* (2005) designed the first primer set MZIA1-bis/A6, which targeted the conservative MCP-encoding gene *g23* of T4 (T-even), *Aeromonas* phage Aeh1 (SchizoT-even) and *Synechococcus* cyanophage S-PM2 (ExoT-even), to investigate the diversity of T4-like myoviruses. They revealed that *g23* homologues of the T4 are ubiquitous in aquatic environments. Using this primer set, a few studies then investigated the diversity of T4-like myoviruses in both marine and freshwater ecosystems (Filée *et al.*, 2005; López-Bueno *et al.*, 2009; Butina *et al.*, 2010, 2013; Huang *et al.*, 2011; Jamindar *et al.*, 2012; Parvathi, Zhong & Jacquet, 2012; Bellas & Anesio, 2013; Zheng *et al.*, 2013), but also in soil and sediments (Jia *et al.*, 2007; Fujii *et al.*, 2008; Wang *et al.*, 2009a,b, 2011; Fujihara *et al.*, 2010; Liu *et al.*, 2011, 2012; Li *et al.*, 2012; Bellas & Anesio, 2013). These studies showed that T4-like myoviruses are prevalent viral components in nature and exhibit distinct diversity between geographical locations or environments. Although numerous *g23* sequences have been generated so far, the majority of the hosts of these viruses are still largely unknown outside the cyanobacteria (Chow & Fuhrman, 2012). Jamindar *et al.* (2012) and Zhong *et al.* (2014) recently made a connection between genome size and *g23* myovirus diversity through the examination of DNA extracted from gel slices of PFGE bands. They found that the *g23*-gene-estimated T4-like myoviruses may have a wider range in genome size than previously appreciated: from 23 kb (Jamindar *et al.*, 2012) to 317 kb (Zhong *et al.*, 2014).

Recently, a new primer set named T4superF1/R1 (Chow & Fuhrman, 2012) has been designed, based on the alignment of about 1400 *g23* MCP sequences obtained from the GOS metagenome and from isolates used in Comeau & Krisch (2008). This set targets roughly the same *g23* region as MZIA1bis/A6, but with higher degeneracy and a shorter fragment. To the best of our knowledge, only three studies have used this primer set to investigate the diversity and community dynamics of marine T4-like myophages, using T-RFLP (Chow & Fuhrman, 2012; Needham *et al.*, 2013; Pagarete *et al.*, 2013). They showed that T4-like myoviruses could exhibit seasonal dynamics, recurring regularly in monthly samples over a 2- to 3-year period (Chow & Fuhrman, 2012; Pagarete *et al.*, 2013). Some OTUs displayed 'boom–and-bust' dynamics (appearing like 'predator–prey' patterns) and oscillations in abundance that could be related to changes in environmental conditions and/or hosts. Needham *et al.* (2013) worked on daily samples over a time series of 78 days and examined simultaneously the community structure of bacteria using ARISA. They observed correlations between the dynamics of some T4-like myovirus OTU and bacteria (e.g. members of SAR11 and Actinobacteria) over a scale of a few days.

The microbial, including viral, ecology of Western European peri-alpine lakes (and fresh waters in general) has been poorly investigated until now. For viruses, studies have largely focussed on community dynamics, vertical distribution and prokaryotic mortality and have only been investigated on a few occasions in these ecosystems (e.g. Jacquet *et al.*, 2005, 2007; Duhamel *et al.*, 2006; Sime-Ngando *et al.*, 2008; Personnic *et al.*, 2009a,b; Thomas *et al.*, 2011). Very recently, we reported the prevalence of a variety of algal virus signature genes in the two largest natural French peri-alpine lakes, Annecy and Bourget, and revealed high and specific algal virus and cyanophage diversity in these ecosystems (e.g. Zhong & Jacquet, 2013, 2014). However, the diversity and the assemblage dynamics of T4-like myoviruses have not yet been investigated. In the present study, we used a PCR–DGGE approach followed by cloning-sequencing analyses to examine the structure and dynamics of individual constituents of the T4-like myovirus assemblage in these two neighbouring (<50 km distant) peri-alpine lakes (also the largest lakes in France). At the same time, we quantified biotic (e.g. potential hosts) and abiotic (nutrients, light, temperature, etc...) factors. Our aims were to (i) uncover the diversity and assemblage dynamics of T4-like myoviruses; (ii) compare the observed patterns between two ecosystems in the same eco-region, but characterised by differing productivity; and (iii) relate viral assemblage structure and dynamics to the abundance of potential hosts and to abiotic variables. Our ultimate goal was to find whether freshwater T4-like myophages displayed patterns similar to those encountered in marine ecosystems, that is, whether these viruses form a reservoir from which a population can initiate infections when its host abundance increases. This would be consistent with the seed bank model proposed by Breitbart & Rohwer (2005), wherein only a small portion of a viral assemblage is active and abundant at any given time and most populations are rare and/or inactive, forming a seed bank of viruses that can 'kill-the-winner' (*sensu* Thingstad, 2000) when hosts reach critical thresholds of abundance.

Methods

Sample collection and processing

Water samples were collected once or twice per month between January and November 2011 from lakes Annecy

and Bourget (France). For each lake, the samples were taken at a single reference station located at the deepest part of the lake (e.g. 45.8727°N 6.1645°E for Lake Annecy and 45.94167°N 6.0305°E for Lake Bourget). We obtained 14 samples for Lake Annecy and 18 for Lake Bourget. We collected >20 L of water integrated from the surface to a depth of 20 m using an electric pump and tubing. The water was stored in a polycarbonate flask in the dark at 4 °C until filtration (performed only a few hours after sampling). Viruses were prepared from a 20-L sample that was sequentially filtered through a 60-μm mesh, then through 142-mm-diameter, 1-μm pore-size polycarbonate filters (Millipore, Bedford, MA, U.S.A.) and the <1-μm filtered water was finally concentrated to a final volume of 200–250 mL using a 30 000-molecular-weight-cut-off spiral-wound ultrafiltration cartridge (regenerated cellulose, PLTK Prep/scale TFF, 1 ft^2; Millipore). Up to six cartridges were rinsed using hot NaCl and milliQ water between each use, and the same cartridge was used only for one lake (thus avoiding contamination between uses and lakes). The <1-μm fraction thus obtained was further filtered twice through 47-mm-diameter 0.45-μm pore-size filters (Millipore) to remove any remaining bacteria [checked using both flow cytometric (FCM) and PCR]. Ten mL aliquots of viral concentrate (VC) was then stored at −20 °C for PCR-DGGE (see below).

For FCM analysis (see below), extra 50 mL water samples were collected from different specific depths (2 or 3, 10, 15 and 20 m) in both lakes to assess the vertical distribution of viruses and potential bacterial hosts.

Environmental parameters

Nutrient concentrations [total nitrogen (N_{total}), N-NH_4^-, N-NO_3^-, SiO_2, P-PO_4^- and total phosphorus (P_{total})] were measured at each station and date for the water column integrated 0–20 m, according to the standard French AFNOR protocols (http://www.afnor.org). A conductivity–temperature–depth measuring device (CTD Seabird SAB 19 Seacat profiler) and a chlorophyll (Chl) fluorescence fluoroprobe (BBE Moaldenke, Schwentinental, Germany) were used to obtain vertical profiles for water temperature, pH, conductivity, dissolved oxygen and chlorophyll *a* concentration.

Flow cytometric (FCM) analysis

Autotrophic picocyanobacteria, heterotrophic prokaryotes and virus-like particles (of which three groups, referred to as VLP1, VLP2 and VLP3, could be distinguished) were counted using a FACSCalibur flow cytometer (Becton Dickinson) as described previously (Duhamel *et al.*, 2006; Personnic *et al.*, 2009a).

Light microscopy

Water samples were immediately fixed with Lugol's solution on board the ship. From each sample, 25 mL was transferred into an Utermöhl counting chamber and left to settle for at least 12 h in the dark. Cyanobacteria counts were then carried out using an inverted microscope (Zeiss, Oberkochen, Germany) following Druart & Rimet (2008).

Transmission electron microscopy (TEM)

Tailed bacteriophages (myoviruses, podoviruses and siphoviruses), identified on the basis of their morphological structures, were observed and counted in glutaraldehyde-fixed sample (0.25% final concentration, grade I; Merck, Darmstadt, Germany) using transmission electron microscopy (TEM) as described elsewhere (Pradeep-Ram *et al.*, 2010).

PCR amplification and DGGE

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, before DGGE analysis, we conducted the PCR in two stages as recommended by Short & Suttle (2002). The first stage used the VC as the template, with the primer set without the GC clamp. A second stage was then performed on the DNA product of the first stage using the GC-clamp-containing primer set (i.e. with 40nt GC clamp attached to the 5′ of forward primer). The PCRs were carried out in a DNA Thermal Cycler T-Professional (Biometra, Göttingen, Germany) to amplify the MCP-encoding gene *g23* using the primer set MZIA1-bis/A6 (Filée *et al.*, 2005). The mix was made with 25 μL of 1X PCR buffer, 4 mM MgCl2, 200 μM of each dNTP, 0.4 μM of each primer, 0.5 U of Platinum® Taq DNA polymerase (Invitrogen, Carlsbad, CA, U.S.A.) and 1 μL of VC (correspond to *c.* 10^6 VLPs estimated using FCM) or DNAs. The program for the first-stage PCR was a 15-min virion lysing and initial denaturation at 95 °C, followed by 34 cycles of denaturation at 95 °C for 30 s, annealing at 50 °C for 30 s, extension at 72 °C for 45 s and a final extension at 72 °C for 5 min. The program for the second-stage PCR was the same as the first-stage PCR, but only with a 5-min initial denaturation at 95 °C

and 24 cycles of amplification. Note that the number of cycles was optimised to our samples.

The DGGE was conducted in a 6% polyacrylamide gel with an optimised linear denaturing gradient (100% denaturant is defined as 7 M urea and 40% deionised formamide). The linear denaturing gradient was optimal at 45–70% for *g23* amplicons. Two hundred nanograms (estimated using gel densitometry) of amplicons from the second-stage PCR was loaded in each well. The electrophoresis was carried out for 16 h in 1× TAE buffer (40 mM Tris-base, 20 mM sodium acetate, 1 mM EDTA; pH 7.4) at 120 V and a constant temperature of 60 °C using the CBS-DGGE 2000 system (C.B.S. Scientific, San Diego, CA, U.S.A.). The gels were stained in a 30 mL TAE solution (40 mM Tris, 20 mM sodium acetate, 1 mM EDTA; pH 7.4) containing 2× SYBR Green I (10 000×, Molecular Probes; Invitrogen) for 45 min before visualising on a UV transilluminator (Tex-35M; Bioblock Scientific, Illkirch, France) and photographing with a GelDoc (Bio-Rad, Hercules, CA, U.S.A.). Due to the limited number of wells in our DGGE device, we ran samples in two gels (one for Lake Annecy and the other for Lake Bourget). To identify the DGGE bands shared by two lakes, samples of the other lake (those possessing typical distinct banding pattern) were also loaded at the same time in the DGGE migration to serve as a reference. Note that this standardisation was conducted twice to confirm the results.

The DGGE banding patterns were then analysed using the GelCompare II software package (Applied Maths, Kortrijk, Belgium) as reported elsewhere (Berdjeb *et al.*, 2013). Briefly, the fingerprinting patterns were first standardised against a reference pattern placed on the left, middle and right side of the gels. The bands were selected one by one by ourselves and not automatically by the software. The clustering analysis was then realised using Pearson's similarity index based on the absence/presence and intensity of the bands for which positions were discriminated unambiguously with a 1% tolerance. The UPGMA method was used to construct the dendrogram from a binary matrix of similarity values.

DNA purification, cloning and sequencing

Gel slices of 74 visibly different DGGE band representatives were excised from the gels. The DNA of each DGGE band representative was eluted from the excised gel slice by adding 100 μL sterile 1× TAE buffer and heated for 15 min at 95 °C. Three μL of eluted DNA served as template in a 22 μL PCR mixture using the corresponding primer set. PCRs were performed with the same conditions of the first-stage PCR, as described above. The amplicons were first verified by electrophoresis in a 1.5% agarose gel, then purified by the Illustra™ GFX™ PCR DNA and Gel Band Purification Kit (GE Healthcare, Little Chalfont, UK) and finally cloned into pCR®4-TOPO® vectors using the TOPO TA Cloning® Kit (Invitrogen). Eight randomly selected clones for representatives of each band were then sent for sequencing to GATC Biotech (Konstanz, Germany). After cleaning and correcting sequences using BioEdit 7.0.5.3 (Hall, 1999), we obtained 190 different non-redundant *g23* sequences for lakes Annecy and Bourget. The absence of chimera sequences was examined using the online Chimera detection tool (http://biotech.inbre.alaska.edu/fungal_portal/?program=chimera_test). These sequences have been deposited in GenBank with accession numbers KF377602 to KF377791.

Phylogenetic analysis and UniFrac statistical analysis

The nucleotide sequences were manually edited using BioEdit to obtain the correct/proper reading frame, by respecting the criteria: (i) the absence of stop codons in the whole sequence, and (ii) the obtaining of expected share features when aligned with MCP of T4-like virus isolates. At last, the nucleotide sequences were translated into amino acids using the online translate tool (http://www.hiv.lanl.gov/content/sequence/TRANSLATE/translate.html). The amino acids deduced were aligned using MAFFT version 7 (Katoh *et al.*, 2002), with representatives of the cultured T4-like myoviruses and environmental sequences of different defined clusters from other studies (Filée *et al.*, 2005; Butina *et al.*, 2010, 2013; Jamindar *et al.*, 2012; Zheng *et al.*, 2013). 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 Bayesian inference and maximum-likelihood methods. Bayesian inference was conducted using MrBayes 3.2.1 (Ronquist *et al.*, 2012) with two runs, four chains, two million generations, sampling every 100 generations, a burn-in value of 25% and mixed models of amino acid substitution. The maximum-likelihood phylogeny was constructed using PhyML 3.0 (Guindon & Gascuel, 2003) with 100 bootstrap replicates and with the best model of acid-amino substitution and rate heterogeneity. The best model for the aligned sequence data set was determined using MEGA 5 (Tamura *et al.*, 2011) and was the JTT model and gamma-distributed substitution rates.

To evaluate whether the *g23* T4-like myovirus assemblage of Lake Annecy differed from that of Lake Bourget, we carried out statistical analyses using the UniFrac distance metric statistical tools available at http://bmf.colorado.edu/unifrac/ (Lozupone & Knight, 2005; Lozupone *et al.*, 2007). We used the unweighted UniFrac option to compare community composition based on the presence/absence (i.e. on qualitative data) from the Bayesian phylogenetic tree. Finally, we conducted a *P*-test based on the UniFrac distance matrix generated for *g23* T4-like myovirus assemblages of Lake Annecy and Lake Bourget.

Statistical analysis

We ran a Pearson's correlation analysis to identify the pairwise relationships between the different parameters. To investigate the relationships between community structure of the T4-like myophages (i.e. the band patterns) and environmental variables as explanatory factors (i.e. temperature, nutrients, abundances and/or type of bacteria and viruses, etc.; inferred from real or averaged values calculated from all data obtained from the 0-20 m integrated layer), a canonical correlation analysis (CCA) was performed using the software package XLSTAT-ADA (Addinsoft, Paris, France). CCA generates an ordination plot that shows the main pattern of variation in community structure as accounted for by the environmental variables measured. The various variables were submitted to the forward selection procedure, in which the statistical significance of the term was tested by the unrestricted Monte Carlo permutation test (999 permutations). Explanatory variables with *P*-values >0.05 were excluded from further analyses.

Results

Environmental parameters

Table 1 reports the minimum, maximum and mean values for the different variables monitored during the study. For both lakes, temperature of surface waters followed the expected pattern: it began to increase in late February, reached its maximum in September and then decreased until November. The concentration of N_{total} and $N–NO_3$ started to decrease in March. In Lake Bourget, P_{total} and $P–PO_4$ decreased after reaching a peak in early April and February, respectively, while in Lake Annecy, they started to increase only in late summer. Concentrations of $N–NH_4$ showed a significant increase in spring in Lake Bourget, while they remained relatively stable throughout the year in Lake Annecy. The

Table 1 Minimum, maximum and mean values of the different biological and environmental variables presented in this study

	Lake Bourget				Lake Annecy			
Variables	Min. (month)	Max. (month)	Mean	SD	Min. (month)	Max. (month)	Mean	SD
Temperature (°C)	5.5 (Feb)	16.9 (Sep)	12.5	4	5.2 (Feb)	17.4 (Sep)	12.6	4.5
Light (μmol m^{-2} s^{-1})	12.6 (Jan)	404.6 (Jun)	227	122.7	29.7 (May)	484.9 (Jun)	233.3	144.4
Dissolved oxygen (mg L^{-1})	7 (Sep)	12.8 (Apr)	9.5	1.6	9.5 (Jan)	15.4 (Apr)	11.5	1.4
Total phosphorus (μg L^{-1})	7 (Jul)	27 (Apr)	12	5	4 (May)	11 (Aug)	5.9	2
Orthophosphates (μg L^{-1})	2 (Oct)	10 (Feb)	5	2	1 (Oct)	4 (Sep)	2.5	0.9
Total nitrogen (μg L^{-1})	333 (Nov)	758 (Mar)	537	109	167 (Nov)	397 (Mar)	284	65
Nitrates (μg L^{-1})	175 (Nov)	510 (Mar)	318	96	10 (Nov)	243 (Mar)	111	67
Ammonium (μg L^{-1})	2 (Jan)	57 (Apr)	16	16	1 (Jan)	8 (Oct)	4.6	2.2
Silicates (μg L^{-1})	790 (Sep)	3030 (Feb)	1737	762	907 (Nov)	3503 (Mar)	2532	891
Chlorophyll *a* (mg L^{-1})	1.47 (Jan)	13.67 (Apr)	4.17	3.1	0.11 (Mar)	0.32 (Aug)	0.24	0.06
Heterotrophic bacteria (10^5 cells mL^{-1})	13.8 (Feb)	45 (Apr)	26.4	7.6	5.76 (Jul)	33.9 (Oct)	18.1	6.92
Synechococcus spp. (10^4 cells mL^{-1})	0.22 (May)	16 (Aug)	4.14	4.36	4.45 (Jun)	15.6 (Mar)	9.17	3.35
Unicellular cyanobacteria (cells L^{-1})	0 (Jun, Nov)	204 (Apr)	63.7	71.1	0 (*)	36 (Apr)	7.2	13.3
Colonial cyanobacteria (cells L^{-1})	0 (Feb)	233 (Nov)	77.8	81.8	0 (Jun)	1094 (Oct)	308.7	320.8
Filamentous cyanobacteria (cells L^{-1})	0 (May)	890 (Nov)	97.5	229.2	0 (*)	122 (Jun)	13.4	35.6
Virus-like particles 1 (10^7 part. mL^{-1})	3.5 (Jan)	27.5 (May)	8.88	5.04	2.75 (Jun)	10 (Sep)	5.14	2.2
Virus-like particles 2 (10^5 part. mL^{-1})	20.6 (Feb)	66.9 (May)	44.5	13.6	4.3 (Jun)	85 (Sep)	41.4	20.9
Virus-like particles 3 (10^5 part. mL^{-1})	2.1 (Feb)	13.7 (Apr)	7.51	3.93	2.21 (Jul)	9.41 (Nov)	4.82	2.45
Myoviruses (%)	19 (Jun)	56 (Aug)	37	11	10 (May)	54 (Nov)	24	14
Podoviruses (%)	9 (Jul; Aug)	18 (Apr)	13	3	10 (Nov)	46 (Apr)	32	12
Siphoviruses (%)	18 (Nov)	59 (Jun)	38	13	16 (Nov)	49 (Jul)	31	9

*This value is observed at different occasions. For unicelluar cyanobacteria in Lake Annecy: March, May to August and October. For filamentous cyanobacteria in Lake Annecy: February to May, July and September to November.

mean concentration of P and N nutrients was about 1.5–8 times higher in Lake Bourget than in Lake Annecy. The concentration of chlorophyll *a* (Chl*a*) increased between early March and early April in both lakes, reaching a higher, sustained level in Lake Bourget.

Vertical distribution of bacteria and viruses

Abundance of VLP1 displayed greater temporal and vertical fluctuations in Lake Bourget than in Lake Annecy, where concentration increased threefold to sevenfold at the end of spring – early summer and late autumn, respectively (Fig. 1c, h). Both VLP2 and VLP3 displayed more striking temporal and vertical variations than VLP1 in both lakes (Fig. 1d, i, e, j). The patterns were rather similar between these two viral groups from April in Lake Annecy and from June in Lake Bourget until the end of the year. Slight differences were detected in Lake Annecy in February and March, when VLP2 were more abundant than VLP3, and in Lake Bourget in April, when VLP3 increased markedly. Picocyanobacterial and heterotrophic bacterial concentrations varied differently in lakes Annecy and Bourget. Picocyanobacterial abundance was high in August and October in Lake Bourget, extending down to 15 and 10 m, respectively (Fig. 1f). By contrast, the abundance of picocyanobacteria was high down to 20 m in late winter/early spring in Lake Annecy, but was generally low thereafter, excepting occasional peaks in abundance in summer below 15 m. This assemblage then increased again in autumn (Fig. 1a). For the heterotrophic bacteria, abundance was high in autumn in Lake Annecy. In Lake Bourget, however, abundance was high in April, June, August and October. For both lakes, the heterotrophic bacteria were more strongly coupled to patterns of VLP2 and VLP3 than to VLP1. There were significant and positive relationships ($P < 0.01$) between these groups ($r = 0.52$, 0.73 and 0.8 between the heterotrophic bacteria and VLP1, VLP2 and VLP3, respectively, for Lake Bourget; $r = 0.55$, 0.73 and 0.67 between the heterotrophic bacteria and VLP1, VLP2 and VLP3, respectively, for Lake Annecy). There was only a weak relationship between the picocyanobacteria and the heterotrophic bacteria in Lake Annecy ($r = 0.48$, $n = 78$, $P < 0.01$).

Temporal dynamics of bacteria and viruses

The total viral abundance in Lake Bourget varied between 4.2×10^7 (Jan) and 2.1×10^8 VLP mL^{-1} (May), which was, on average, 1.6 times higher than in Lake Annecy [2.7×10^7 (Jun) to 8.4×10^7 VLP mL^{-1} (Nov)]. VLP1 was the dominant group whatever the lake or the period considered and accounted, on average, for 94% of virus abundance. In comparison, the contributions by VLP2 and VLP3 to total viral abundance were 5 and 1%, respectively (Table 1). Bacterial concentrations varied from 1.4×10^6 to 4.5×10^6 cells mL^{-1} in Lake Bourget and were about 1.5 times higher than in Lake Annecy (i.e. between 5.8×10^5 to 3.4×10^6 cells mL^{-1}). Over the year, picocyanobacterial abundance was about 2.4 times higher in Lake Annecy (with concentrations varying between 4.5×10^4 and 1.6×10^5 cells mL^{-1}) than in Lake Bourget (2.16×10^3 to 1.6×10^5 cells mL^{-1}). Nevertheless, in terms of relative abundance, picocyanobacteria accounted for <1.4% of the bacterioplankton in Lake Bourget but reached 5% in Lake Annecy.

Three types of cyanobacteria (unicellular, colonial and filamentous forms) could be observed microscopically, and the abundance of each group varied over time (Fig. 2; Table 1). The unicellular cyanobacteria detected by microscopy were mainly *Synechococcus*-like or *Synechocystis*-like spp. (mainly *Synechocystis parvula* and *Synechocystis aquatilis*), which were nine times more abundant in Lake Bourget than in Lake Annecy. Colonial cyanobacteria in the colonial form (mainly *Aphanocapsa* and *Chroococcus* spp.) were on average four times more abundant in Lake Annecy than in Lake Bourget. They outnumbered the filamentous and/or unicellular cyanobacterial forms in Lake Annecy throughout the year, except in June, while this was true only in summer in Lake Bourget. The mean abundance of filamentous cyanobacteria was 7.5-fold higher in Lake Bourget (mainly represented by *Aphanizomenon flos aquae*, *Pseudanabaena catenata* and *Pseudanabaena limntica*) than in Lake Annecy (represented mainly by *Pseudanabaena limnetica*).

Transmission electron microscopy revealed that the myoviruses and siphoviruses were two important groups in Lake Bourget, accounting for up to 56% (Aug) and 59% (Jun) of the bacteriophage morphotypes, respectively (Table 1). It is noteworthy, however, that in Lake Annecy, the podoviruses occasionally dominated the bacteriophage assemblage (e.g. 46% in April).

Assemblage structure and phylogeny of T4-like myoviruses

The number of *g23* DGGE bands varied between 28 (January) and 40 (November) in Lake Annecy and between 32 (July) and 40 (September) in Lake Bourget (Fig. 3). For Lake Annecy, we obtained 56 different bands of which *c.* 11% were detected in all samples. In Lake Bourget, we obtained 60 different bands of which only

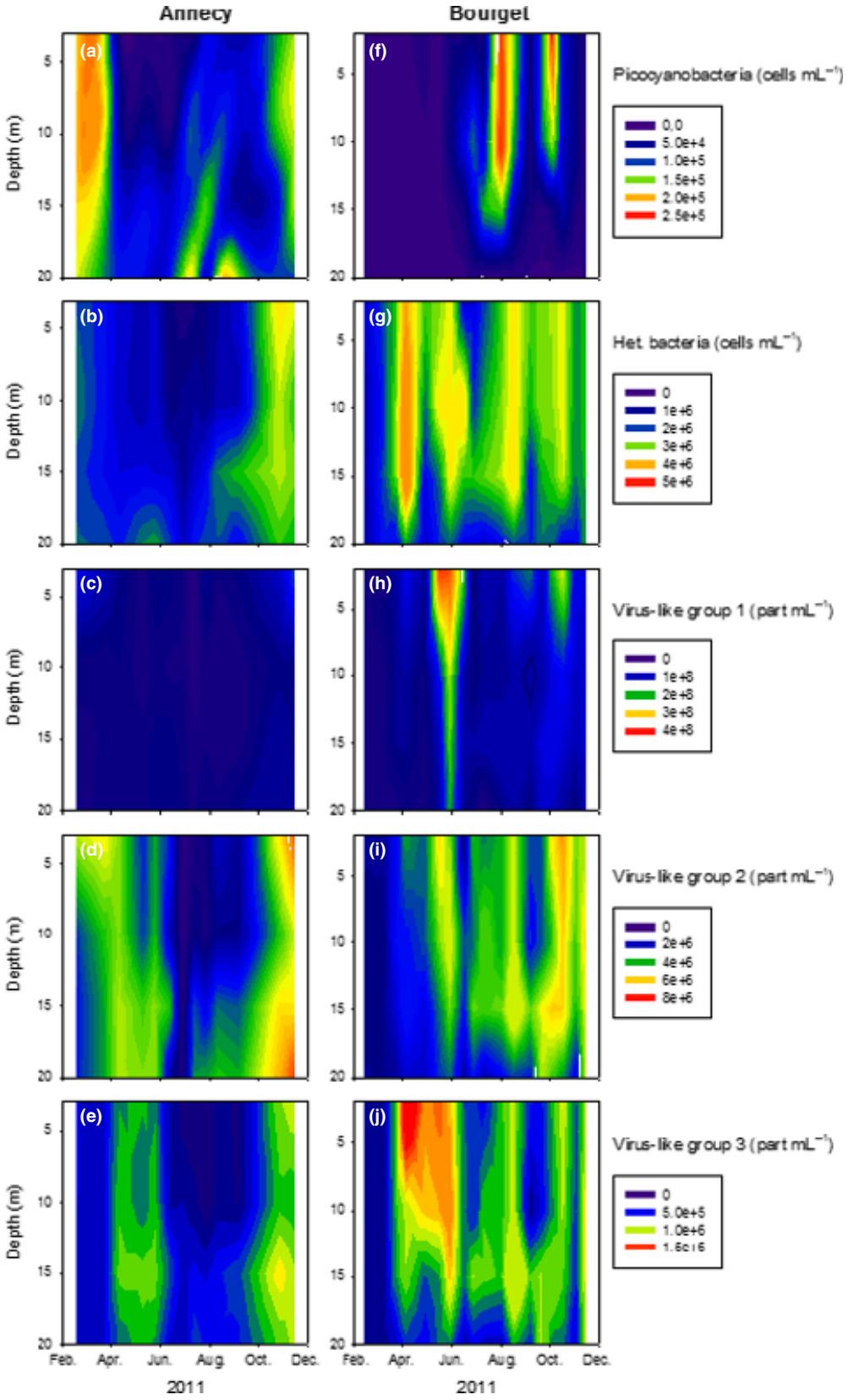

Fig. 1 Development of picocyanobacteria (a, f), heterotrophic bacteria (b, g) and virus-like particles [VLP 1 (c, h), VLP2 (d, i) and VLP3 (e, j)] in surface waters (0–20 m) of lakes Annecy (left) and Bourget (right) throughout 2011.

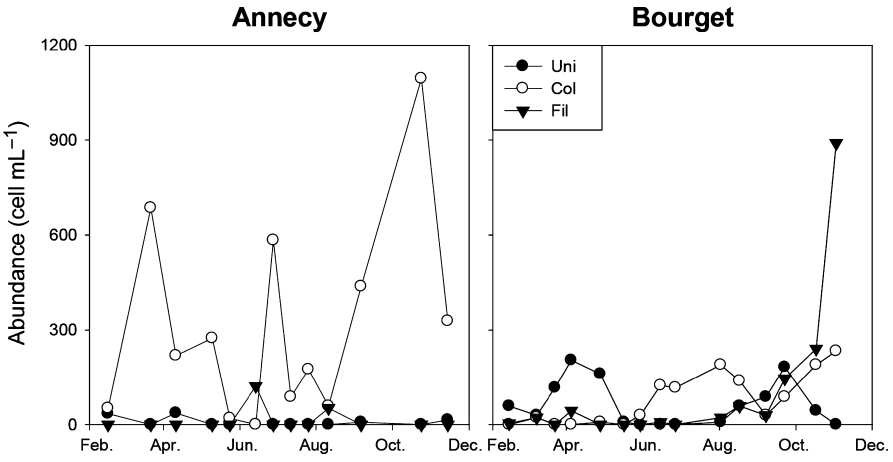

Fig. 2 Cyanobacterial type (unicellular, colonial or filamentous) abundances in lakes Annecy and Bourget in 2011.

17% were present in all samples. For both lakes, a total of 74 unique DGGE bands were obtained, of which 42 were common to both systems, and the band pattern varied over time between lakes. Only one band persisted in both lakes throughout the year. The DGGE banding pattern analysis (Fig. 3), with 60% similarity, revealed that the T4-like myovirus assemblage in Lake Bourget clustered into three distinct groups: winter to early spring, mid-spring to summer and autumn. For Lake Annecy, samples were grouped into two clusters: from late winter to early summer and from mid-summer to autumn, with 40% similarity between the two.

We excised the representatives of each DGGE band type for DNA purification, cloning and sequencing. Sixty-eight of 74 bands (92%) resulted in clones, and the other bands (characterised by weak luminosity or fluorescent density) failed, probably due to the low DNA concentration. We obtained 190 non-redundant sequences, varying from 388 to 541 bp, with nucleotide similarity varying between 31.6 and 99.8% and amino acid similarity varying between 42.9 and 100%. Among the eight sequenced clones analysed for each band, between one and five different sequences were obtained, with nucleotide dissimilarity reaching up to 61.2%. Ten of the 68 *g23* bands (15%) contained a single sequence. Therefore, a DGGE band could only be considered as a subpopulation of T4-like myovirus in most cases.

The phylogenetic analysis (Fig. 4) of these sequences, with both culture representatives of T4-like myoviruses and environmental sequences of different defined clusters from other studies (Filée *et al.*, 2005; Butina *et al.*, 2010, 2013; Jamindar *et al.*, 2012; Zheng *et al.*, 2013), showed that all the sequences obtained were more closely related to Exo T-evens (cyanophages and SAR11 myophages) than the others (T-evens, Pseudo T-evens and Schizo T-evens). For those sequences having phylogenetic proximity to Exo T-evens, we could discriminate 22 clusters, among which two were the culture-containing clusters (clusters I and II containing cyanophages and SAR11 phage HTVC008M). The other sequences assembled into environmental-sequences-only clusters, of which the potential hosts of these viruses are still unknown, due to the absence of cultured representatives to support their phylogenetic affiliation. Among these environmental-sequences-only clusters, four were initially identified by Filée *et al.* (2005) as Marine 2–5, four defined by Jaminder *et al.* (2013) as Clades II–V, one was the cluster B10 previously defined by Butina *et al.* (2010) and we redefined the other eleven clusters as cluster numbers 1–11. Our *g23* sequences were distributed in cluster numbers 1–11, cluster B10, clades II–V and culture-containing cluster II, but none was in the cluster Marine 2–5. Six clusters (clusters number 1, 2, 3, 4, 8 and 9) were composed of *g23* sequences from peri-alpine lakes exclusively.

When the *g23* sequences of each lake were pooled, the UniFrac analysis revealed significant differences between Lake Annecy and Lake Bourget ($P = 0.03$).

Dynamic patterns of individual T4-like myoviruses

We examined the dynamics of individual constituents of T4-like myoviruses (corresponding to DGGE bands) for each sample obtained throughout the year. We detected variability in intensity in nearly all the bands. The relative abundance of a band could increase or decrease up to 17-fold between two sampling dates, that is, in just 15 days. As an example, the relative abundance of band b9 increased from 0.7 to 11.7% between June and July in Lake Annecy (Fig. 5). Some other bands also switched from undetectable to dominant, and *vice versa*. For instance, band b27 'boomed' in July–August in Lake

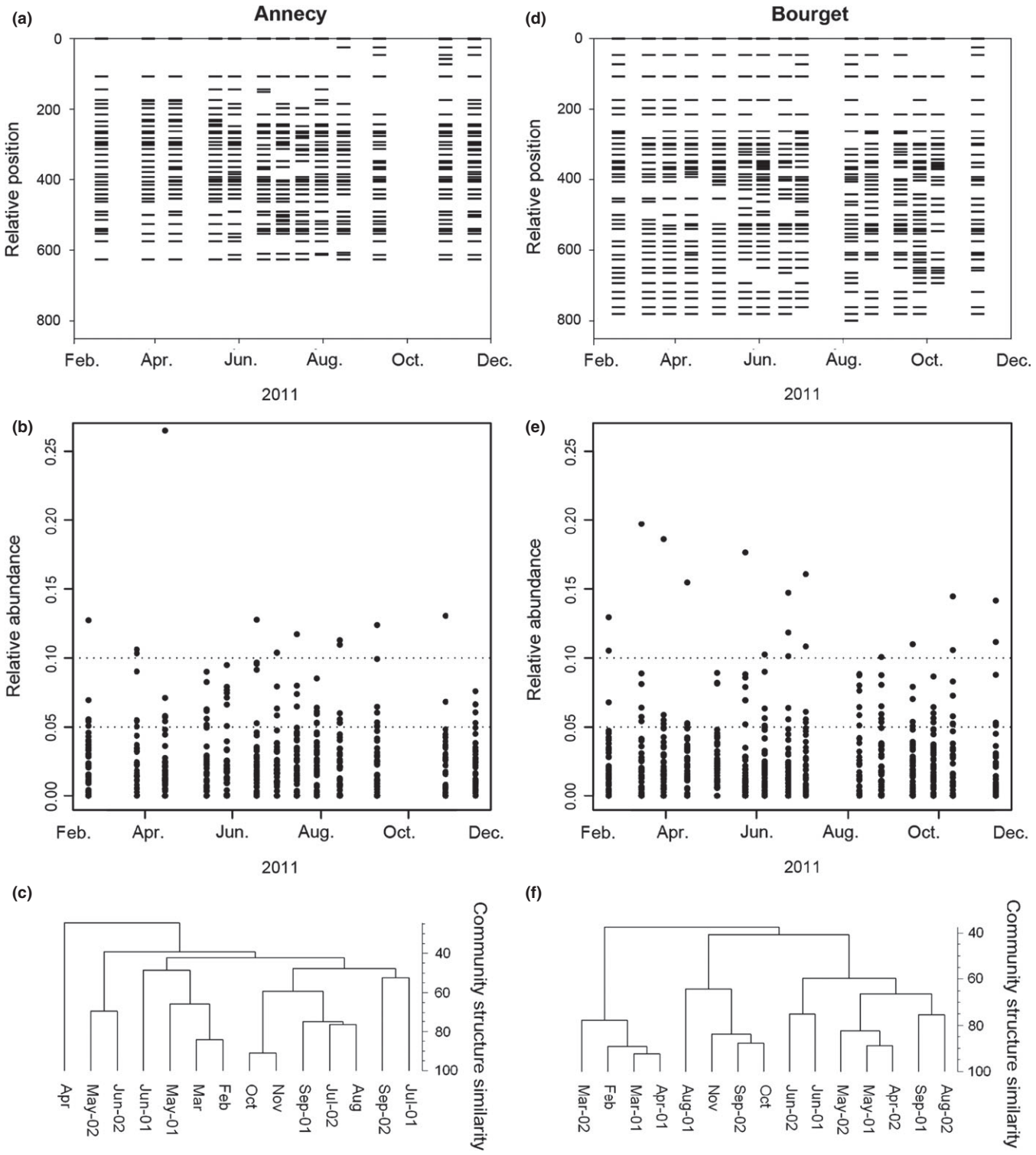

Fig. 3 Schematic outline of the presence/absence (a, d) and the relative abundance (b, e) of the *g23* DGGE bands in the two lakes during 2011. Bands with relative abundances of <5, 5–10 and ≥10% represented 90.4–91.2, 6.8–8.1 and 1.5–2% of the total in the two lakes. The cluster analyses of the DGGE fingerprinting patterns (c, f) were obtained using UPGMA clustering of the Pearson similarity based on both the presence/absence and the relative abundance of the bands. Lanes with no bands correspond to samples from the other lake which served as references and which were removed.

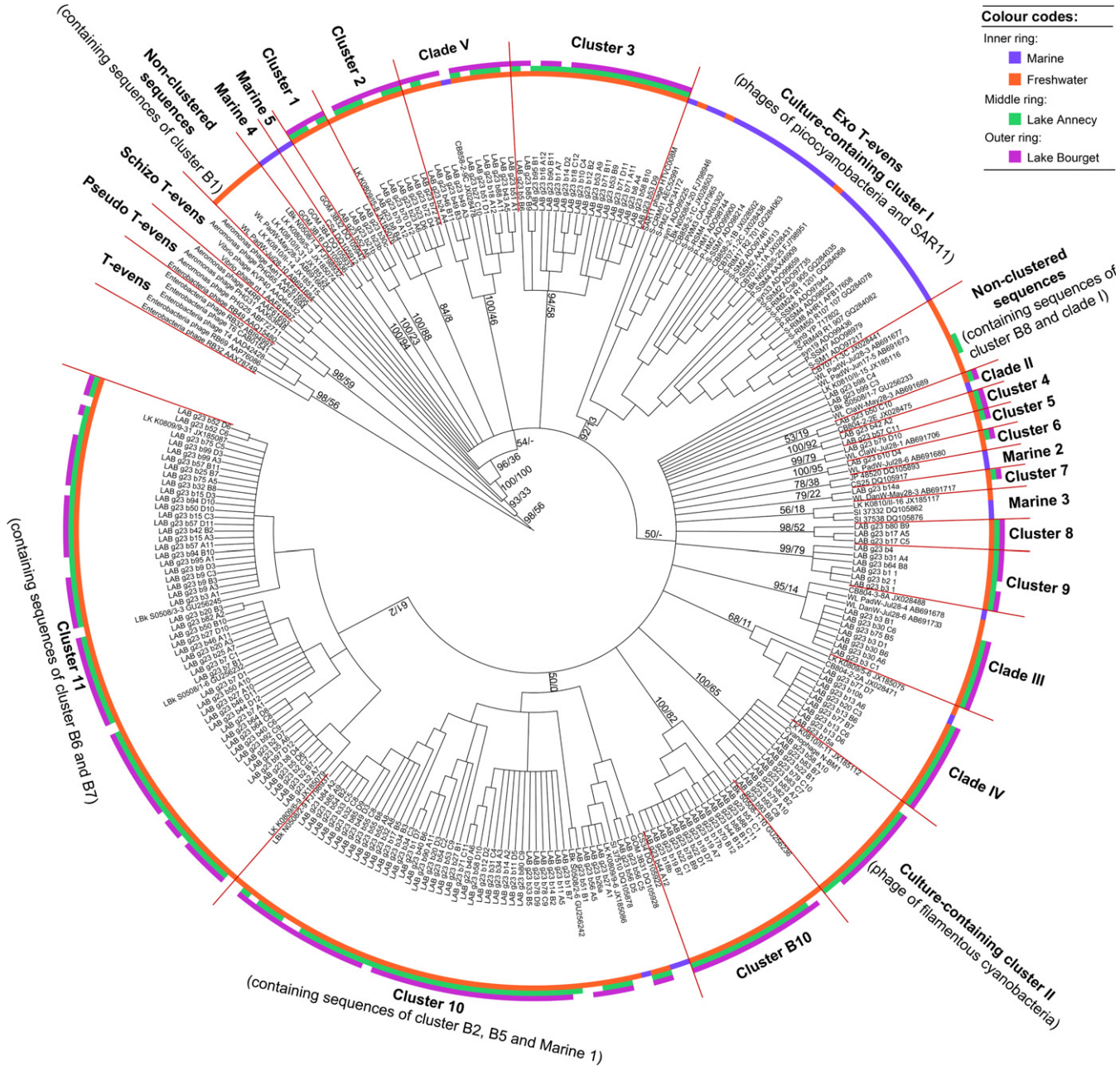

Fig. 4 Bayesian phylogenetic tree based on aligning 100 homologous amino acid positions of the major capsid protein gp23 from 279 gene *g23* sequences. Sequences for T4-like myovirus isolates from lakes Annecy and Bourget and other different aquatic environments (e.g. representative sequences of different defined clusters from other studies). Values at nodes as shown on the main branches are the Bayesian inference (BI) posterior probabilities and the maximum-likelihood (ML) bootstrap values in the order of BI/ML. The inner coloured ring distinguishes among sequences of marine (violet) and freshwater (orange) origin. The middle and outer rings indicate whether the sequences were detected in Lake Annecy (green) or Lake Bourget (purple). Each cluster was identified/assigned by following classical criteria: (i) the branch was supported by BI and ML values; (ii) the value for BI posterior probabilities was ≥50.

Annecy yet disappeared after August (Fig. 5). Overall, we observed three different dynamical patterns for the different *g23* bands: (i) a 'boom-bust' pattern (a dramatic increase or decrease with a sharp peak, either episodic or very occasional); (ii) relatively stable fluctuations (within relatively narrow limits); (iii) only a gradual progression through time (including seasonal variation, monotonic increase or decrease). Often, different pat-

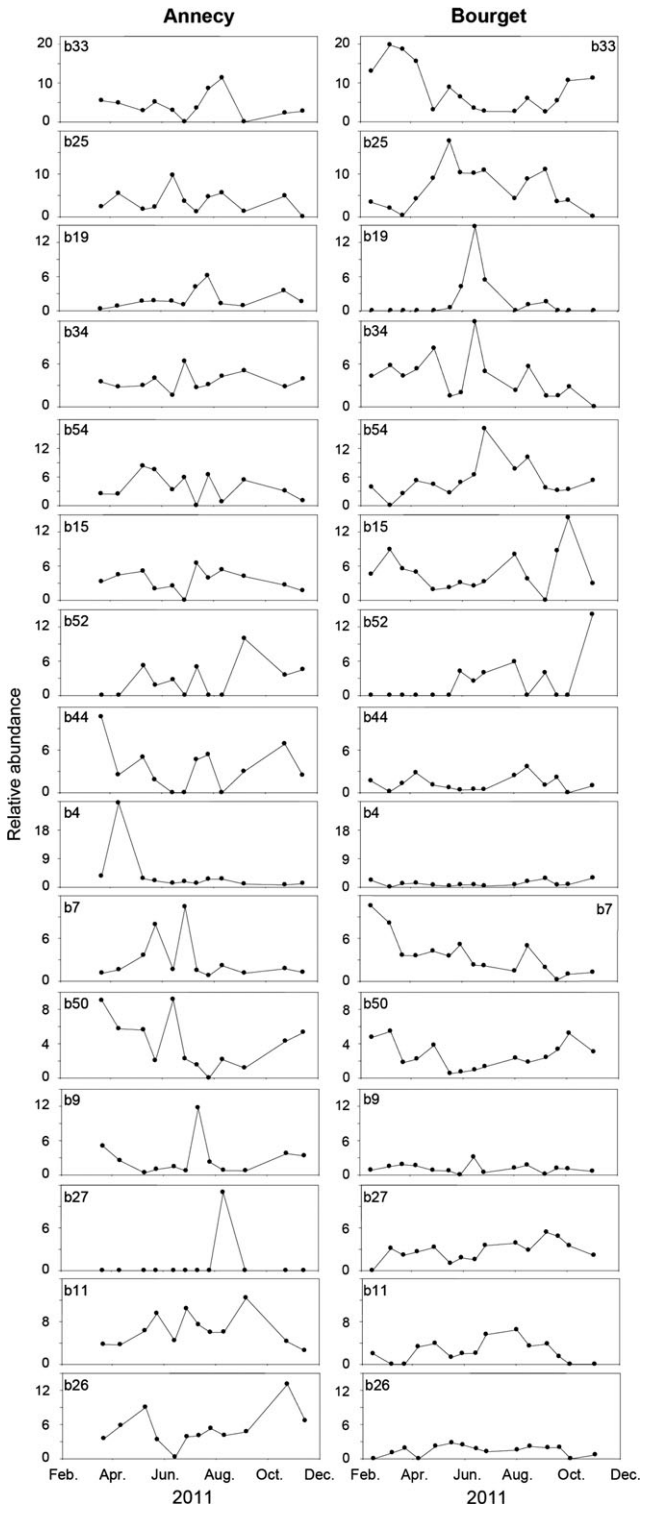

Fig. 5 Dynamic patterns of the dominant T4-like myovirus DGGE bands in 2011 in the two lakes. The *Y*-axis indicates the relative abundance for each band, that is, the normalised per cent contribution of each band to the total T4-like myovirus assemblage in the sample.

terns in the two lakes could be observed for a selected band (e.g. b50). Most bands displayed gradual progressions and relatively stable fluctuations, however.

The examination of each monthly sample revealed that the myophage assemblage was dominated by only a few bands. The most abundant bands accounted for up to 26.5% of the total T4-like myovirus abundance in Lake Annecy (e.g. b4 in April) and up to 18.6% in Lake Bourget (e.g. b33 in March). We found that the identity of the dominant bands changed with time and differed between lakes (Fig. 5). When examining 15 of these dominant bands we found that (i) some of them consisted of persistent bands (detectable throughout the year; e.g. b33 in Lake Bourget), while others were detected only at some times of the year (e.g. b52, b34, etc.) or even just once (e.g. b27 in Lake Annecy in August); (ii) most of them dominated only once in our samples, while a few others reoccurred as dominants (e.g. b7 in Lake Annecy and b25 in Lake Bourget); (iii) some bands displayed seasonal dynamics, but the patterns were different between the lakes ($P < 0.05$); (iv) significant relationships at $P < 0.05$ could be found between the dynamics of some bands and potential hosts [e.g. in Lake Annecy, between b26 and the heterotrophic bacteria ($r = 0.58$) or the colonial cyanobacteria ($r = 0.67$), and between the picocyanobacteria and b44 ($r = 0.62$), b54 ($r = -0.69$) and b9 ($r = 0.64$); in Lake Bourget, between the picocyanobacteria and b10 ($r = -0.51$) or b15 ($r = 0.51$), between the filamentous cyanobacteria and b52 ($r = 0.62$), b53 ($r = 0.62$) and b7 ($r = -0.56$) and between the colonial cyanobacteria and b52 ($r = 0.73$), b53 ($r = 0.57$) and b4 ($r = 0.53$)].

Viral assemblage structure in relation to biotic and abiotic parameters

We examined the relationships between the *g23* assemblage and environmental and biological variables using canonical correspondence analysis (CCA) and then tested the significance of the relationships using Monte Carlo permutations (999 iterations). The same 11 parameters explained 37.9 and 41.2% of the variance in the *g23* T4-like myovirus assemblages in Lake Annecy and Lake Bourget, respectively. These variables were the abundances of the different viral groups, bacteria and cyanobacteria, the P_{total} and the proportion of the myo- and siphoviruses. It is noteworthy, however, that each

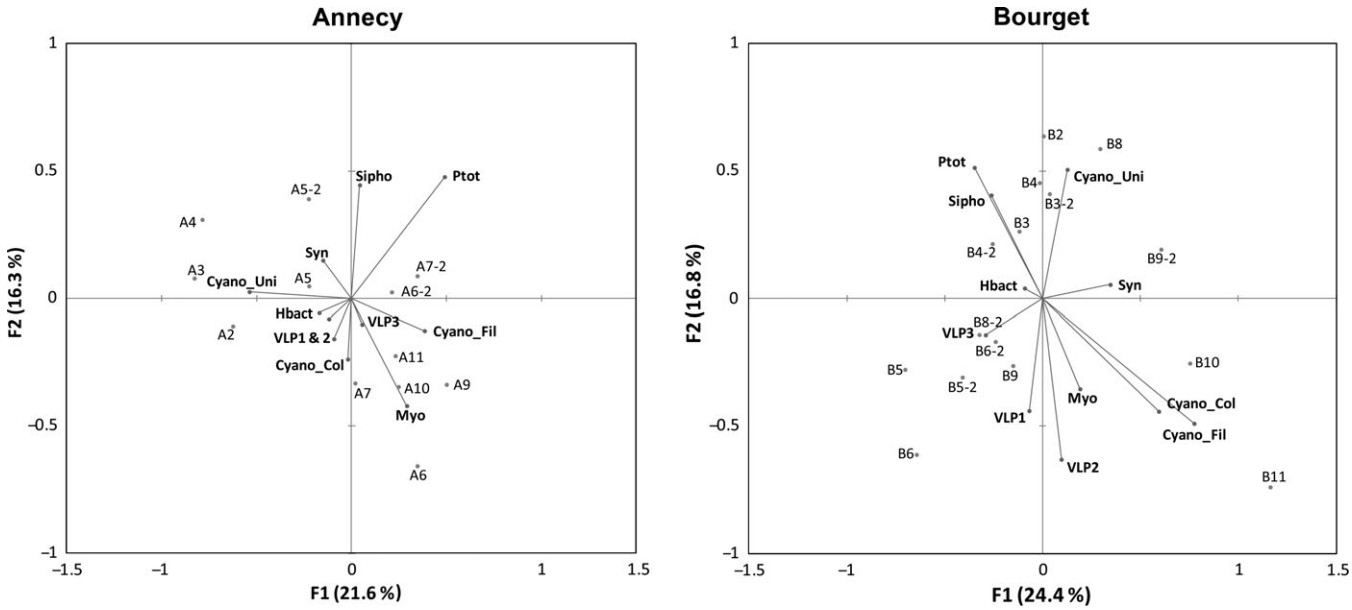

Fig. 6 Canonical correspondence analysis using both environmental parameters and biological counts as constrained variables to explain temporal changes in T4-like myovirus assemblage structure. The sample codes give the lake (A = Annecy, B = Bourget) and the month of sampling (2–11). The '-2' designates the second sampling occasion. So, for example, A6 and A6-2 are the first and second samples taken in June from Lake Annecy. B6 is the first sample taken in June from Lake Bourget.

parameter could play a role at different periods of the year between the lakes (Fig. 6).

We carried out a Pearson's correlation analysis to identify relationships between viruses and potential bacterial hosts in terms of abundance. The abundance of VLPs was correlated with the heterotrophic bacteria, but not with the picocyanobacteria (Syn). Also, the relationship between VLPs and the heterotrophic bacteria was different between the two lakes. Indeed, bacterial abundance was related to VLP1, VLP2 and VLP3 abundances in Lake Annecy (all $P < 0.05$), but only with VLP3 in Lake Bourget ($r = 0.6$, $n = 15$, $P < 0.05$). The myoviruses, as detected by TEM, were also correlated with the heterotrophic bacteria ($r = 0.56$, $n = 13$, $P < 0.05$) and the colonial forms of cyanobacteria ($r = 0.62$, $n = 13$, $P < 0.05$), but only in Lake Annecy. There were no significant correlations between podo- or siphoviruses and the abundance of prokaryotes of any kind.

Discussion

Freshwater viruses have received much less attention than marine viruses, especially in terms of viral diversity and assemblage composition (Middelboe, Jacquet & Weinbauer, 2008). To the best of our knowledge, this is the first study of fresh waters dealing with the structure and dynamics of the T4-like myovirus assemblage. We compared two lakes in the same ecoregion (i.e. with the same general environmental constraints) but of different trophic status. Together with the measurements of the virio- and bacterioplankton abundances (including the cyanobacteria) and a variety of other biological and abiotic parameters, our results suggest that the seed bank model for T4-like myophages can be applied to fresh waters such as these peri-alpine lakes.

T4-like myovirus diversity and potential hosts

The dynamics of T4-like myovirus diversity, based on the *g23* MCP gene, differed between lakes, reflecting their distinct trophic status and potential host communities. We found more different DGGE band types in Lake Bourget than in the more oligotrophic Lake Annecy, suggesting that the oligomesotrophic ecosystem may sustain higher virus diversity (or at least a greater number of dominant groups). This observation is consistent with previous studies, in which we reported a greater number of taxa for the potential bacterial hosts in Lake Bourget (Zhong, Berdjeb & Jacquet, 2013). We also found that *g23* composition was significantly different between the two lakes, mainly because at least eight DGGE bands between the relative positions 600 and 800 in the gel in Lake Annecy (Fig. 3).

Although we obtained many *g23* sequences from the two lakes, only a few (11 of 190 sequences, i.e. <6%) clustered together with cultured viruses (Fig. 4). Most of the sequences were distributed in 17 different environmental-sequence-only clusters, for which potential hosts are unknown, and genome representatives being absent (particularly myophages infecting Proteobacteria) to support their phylogenetic affiliation. It is interesting to note that the recently isolated phage HTVC008M, infecting the most abundant microbial component in the oceans (SAR11; Zhao *et al*., 2013), also has a *g23* homologue to that of T4. Our phylogenetic analysis, based on either partial (Fig. 4) or entire coding sequences (Fig. S1) of *g23*, and consistent with that of Zhao *et al.* (2013) who used 35 concatenated proteins, showed that it is grouped with phages infecting picocyanobacteria (i.e. members of *Prochlorococcus* and *Synechococcus* in culture-containing cluster I of the Exo T-evens). However, none of our *g23* sequences fell into cluster I, although note that there are only a few available sequences obtained from Lake Baikal and the Chesapeake Bay (Fig. 4). This is inconsistent, however, with: (i) our finding of diverse and numerous *Synechococcus* myoviruses in the two alpine lakes when using *g20* as a proxy (Zhong & Jacquet, 2013; Zhong *et al*., 2013); (ii) the CCA which showed that *Synechococcus* is related to the myophage structure in both lakes; and (iii) the significant relationships found between the dynamics of picocyanobacterial abundance and that of some major bands. Considering the fact that most viruses in culture-containing cluster I have been isolated from the marine environment and that freshwater viruses can be genetically different, we hypothesise that *Synechococcus* myoviruses are present in some of these environmental-sequence-only clusters. Therefore, our observations, in addition to past observations, suggest that the known cyanophage diversity from cultures is underrepresented and that most of the obtained *g23* sequences using this primer set belong to phages infecting bacterial phyla.

Approximately 6% of the *g23* sequences obtained from lakes Annecy and Bourget clustered together with myoviruses infecting the filamentous cyanobacteria N-BM1 [isolated from *Nodularia spumigena* from the Baltic Sea (Jenkins & Hayes, 2006)] in culture-containing cluster II. This result is consistent with the CCA, showing that filamentous cyanobacteria could explain T4-like myovirus structure in autumn in both lakes, and also with the significant relationships between the dynamics of filamentous cyanobacteria abundance and that of some major bands. However, our microscopy observations did not reveal *Nodularia* spp. in the peri-alpine lakes in 2011, but only two closely related genera (*Aphanizomenon* and *Pseudanabaena*) in the same family *Nostocaceae*. *Pseudanabaena limnetica* could be particularly important, as it was detected in both lakes and, in autumn, accounted for up to 47% of the cyanobacterial cell counts in Lake Annecy and 17.5% in Lake Bourget. Note that similar *Nodularia*-phage-like-*g23* sequences have also been detected in Lake Kotokel, in which *Anabaena* spp. are important and could thus be a potential host (Butina *et al*., 2013). It is noteworthy, however, that these studies revealed that *g23* sequences are different between T4-like myoviruses infecting pico- and filamentous cyanobacteria. This may reflect the difference in capsid morphology, capsid lattice arrangements, accessory protein content interactions and/or evolutionary histories among these phages (Comeau & Krisch, 2008).

Recently, Nakayama *et al.* (2009) reported that the siphoviruses infecting *Novosphingobium* and *Sphingobium* (with long non-contractile tails and a genome size range of 40–160 kb) isolated from a rice paddy field possessed *g23* homologues to T4-like myoviruses and could be amplified using MZIA1bis/A6. A similar result was reported for siphoviruses infecting *Nodularia* (e.g. N-BS2, N-BS3 and N-BS4), while their *g23* sequences were amplified by CAP1/2 [which target the same regions as MZIA1bis/A6 (Jenkins & Hayes, 2006)]. These siphoviruses have been identified using TEM, but no genome sequences are available. Such a contradiction between tail morphologies and capsid genes between myoviruses and siphoviruses is puzzling and may suggest that the TEM criteria for discriminating between the two virus types, based on tail features, may be insufficient in these cases. It might be possible that (i) some siphoviruses possess T4-like MCP; or (ii) some T4-like myoviruses possess tails that look similar to those of some siphoviruses. In fact, we can find MCPs in siphoviruses (e.g. S-CBS1, S-CBS2, S-CBS3, S-CBS4 and P-SS2; Sullivan *et al*., 2009; Huang *et al*., 2012), but they are not homologues of T4-like MCP. It is thus crucial to obtain genomes of *Sphingobium* and *Nodularia* siphoviruses. It remains possible that MZIA1bis/A6 can target *g23* sequences of some 'siphoviruses', as shown and proposed by Nakayama *et al.* (2009). Indeed, we observed a large proportion of siphoviruses in both lakes using TEM, and CCA revealed they could be related to *g23* diversity (Fig. 6). We tentatively introduced *g23* sequences of *Sphingobium* 'siphoviruses' into the phylogenetic analysis and found that they grouped with some of our sequences and fell into clusters 1, 10 and clade III (data not shown). This could be consistent with the fact that members of the *Sphingomonadaceae* are important

components in surface waters of alpine lakes (Debroas *et al*., 2009; Humbert *et al*., 2009) and that their phages (either sipho/myo/podo) have been isolated from freshwater lakes (Wolf *et al*., 2003; Jost & Wiese, 2013).

Dynamics of T4-like myoviruses

The assemblage structure of T4-like myoviruses displayed a seasonal pattern and could remain similar for only 15 days to up to 5 months (Fig. 3). Shifts operating at a short time scale (i.e. 15 days) were observed between August and September in Lake Bourget and between April and May in Lake Annecy. Such seasonal transitions were associated with 36% of bands in Lake Annecy and 50% in Lake Bourget. T4-like myophage dynamics differed significantly between the two ecosystems. We observed the following: (i) different identities of the most abundant bands across the year between the lakes; (ii) a higher proportion of bands persisting throughout the year in Lake Bourget (17%) than in Lake Annecy (11%); and (iii) the assemblages of Lake Bourget were more variable than those in Lake Annecy in terms of band numbers, relative abundance, proportions contributing to assemblage shifts, the frequencies of changes, etc. Together, these results confirmed that Lake Bourget is characterised by a more variable assemblage.

Changes observed in band presence/absence and dominance agree with the 'seed bank' model of Breitbart & Rohwer (2005), in which only a few members of the viral assemblage are active and abundant at any given time and most of them are rather rare and/or inactive, forming a potential 'seed bank' for recruitment. Inactive viruses may become abundant depending on host availability, whereas the previously abundant and active fraction can enter into the bank through viral lysis and decay and resulting in a change in identity of the dominants. This viral activity (i.e. viral production versus decay) was probably influenced by the abundance and/or growth of the bacterial hosts (Short, Rusanova & Short, 2011), but also by environmental changes. Such a model is rather rare for freshwater ecosystems.

We examined the patterns for 15 dominant bands during 2011 (Fig. 5). Most of the bands displayed temporal variation, gradual progression or mixed patterns, suggesting complex virus–host interactions (viral production, decay and lysis) in response to host and environmental changes, as well as possible cellular resistance phenomena to viral infection (e.g. b11). We detected that several bands display 'boom-and-bust' patterns (with sharp increases/decreases), responding in a 'kill-the-winner' manner, in which viruses control the most rapidly growing *r*-selected hosts allowing for the coexistence of less competitive populations and sustaining bacterial diversity (Thingstad & Lignell, 1997; Weinbauer & Rassoulzadegan, 2004; Suttle, 2007; Pagarete *et al*., 2013; Våge, Storesund & Thingstad, 2013). However, most bands with 'boom-and-bust' patterns were characterised by monotonic peaks (e.g. b4, b9 and b27 in Lake Annecy), suggesting they were opportunistic viruses moving from seed bank (minority) to a dominant fraction during transiently favourable conditions. In contrast, only a few phages showed multiple sharp peaks in their dynamical patterns (e.g. b7 in Lake Annecy), suggesting resilience/resistance processes. Positive and negative relationships found between the dynamics of some of these bands and abundances of either the heterotrophic bacteria, picocyanobacteria, colonial or filamentous cyanobacteria reinforced the idea of the existence of some links between some of these viruses and specific hosts. However, we should keep in mind that one band can contain several sequences (i.e. various genotypes), so that the real dynamical pattern of an individual myophage could be obscured, and caution must be taken when interpreting such results, as explained below.

Biotic and abiotic factors affecting T4-like myovirus assemblage

The CCAs revealed biotic environmental variables explained between 37.9% and 41.2% of the variance of T4-like myovirus community structures. Potential bacterial hosts (i.e. the heterotrophic bacteria, *Synechococcus* spp., the colonial and filamentous cyanobacteria), different viral groups detected by FCM and TEM (VLP1, VLP2, VLP3, myo- and siphovirus) and total phosphorus were significant variables (i.e. $P < 0.05$) mediating the relationships between the virus assemblage, host abundance and environmental factors. In general, the number and relative contribution of the variables shaping the T4-like myovirus assemblage structure were not dramatically different between the two lakes. Rather, these variables acted differently on each viral group at different times of the year in each lake. This may explain a large portion of the temporal variability in viral community structure in each lake.

We found some relationships between the abundances of VLP1, VLP2, VLP3, myo- or siphoviruses with the T4-like myophage assemblage, suggesting that most T4 phages were probably counted and also that the morphological classification seems generally valid, although not perfectly so (Comeau *et al*., 2012). Note, however, that our results and others using microscopy suggest

that siphoviruses may be part of the T4-like myovirus assemblage. Flow cytometry distinguished three groups based mainly on their dye–nucleic acid complex fluorescence. VLP1 are likely to be associated with small-genome-size bacteriophages (e.g. <80 kb) since this group accounted for >94% of total virus abundance and given that the small-genome-sized viruses dominate the virioplankton (Sandaa, 2008). If true, VLP2 and VLP3 may thus be associated with viruses of larger genome sizes (>80 kb). As T4-like myoviruses (including cyanomyophages) have, on average, a genome size ranging between 150 and 250 kb, VLP2 and/or VLP3 are probably also viruses that we can associate with T4-like myophages. Our results suggest this hypothesis may be valid since the heterotrophic bacterial abundance was more related to VLP2 and VLP3 than VLP1.

The fact that we found some correlations between the abundance of *Synechococcus* spp., cyanobacteria or heterotrophic prokaryote with the T4-like myovirus assemblage was not surprising, as they are well-known potential hosts for these viruses. Currently, known T4-like myoviruses have been found in bacterial families of the *Proteobacteria* and *Cyanobacteria* (Petrov *et al*., 2010), and these two groups are indeed important in these lakes (Debroas *et al*., 2009; Humbert *et al*., 2009). *Synechococcus* abundance was correlated with changes in the myoviral assemblage, as was also recently observed in a Norwegian Atlantic fjord (Pagarete *et al*., 2013). Whereas we did not distinguish groups within the heterotrophic bacteria, the cyanobacteria were separated into three morphological categories (unicellular, colonial and filamentous) and we found that each group could be related to the T4-like myophage assemblage. In particular, we observed that the colonial cyanobacteria could explain the composition of the T4-like myovirus assemblage in July in Lake Annecy and in late autumn in Lake Bourget (Fig. 6). *Chroococcus minutes* was a good candidate as a potential host for myoviruses, since at these times, it sometimes accounted for up to 100% of the cyanobacterial cells identified microscopically (Fig. 2). In Lake Bourget, *Aphanocapsa planctonica* was another good candidate as a host in late autumn, since it accounted for up to 34% of cells identified at the time. Interestingly in relation to these speculations, we found that *C. minutes* was correlated with b99 of cluster 11 ($P < 0.05$) in Lake Annecy and *A. planctonica* to b17 of clusters 8 and 10 ($P < 0.01$).

Among abiotic factors, only total phosphorus seemed to be related to the assemblage composition of T4-like myoviruses. This is perhaps because P is often a limiting nutrient for both autotrophs and heterotrophs in oligotrophic lakes such as Lake Annecy and is seasonally depleted even in Lake Bourget. Our results thus suggested that P_{total} explained variation in the assemblage of T4-like myoviruses in late summer in Lake Annecy and in spring in Lake Bourget. This was likely to be related to the rapid growth of heterotrophic bacteria following proliferation of the phytoplankton and nutrient depletion.

Methodological aspects

Our results must be treated with caution. Firstly, we used integrated water samples (i.e. water mixed from surface to 20 m) to infer myophage diversity and assemblage structure. However, FCM analysis performed on discrete samples clearly showed that picocyanobacteria, heterotrophic bacteria and viruses displayed vertical heterogeneity in this upper lit layer, so we are aware that our sampling strategy (mixing different water layers) could have masked vertical variations in diversity and subtle interactions between viruses and hosts.

To concentrate viruses, we employed ultrafiltration, while Hurwitz *et al.* (2013) have recently revealed, by comparing the efficacy of ultrafiltration versus $FeCl_3$ precipitation, that the former probably leads to an underestimation of viral genotypic diversity. They showed that this was especially true of the *Phycodnaviridae* and the podoviruses. As we used the 'classical' ultrafiltration method, it is thus possible that we underestimated the myoviral genotypic diversity, although this group seems to be less sensitive to the method (Hurwitz *et al*., 2013).

Using a fingerprinting approach, our analysis provides only an approximation of minimum richness for a single sample, since it is well known that this method is not suited to minor groups whose rarity may make them undetectable. We are also aware that the DGGE bands probably contain several genotypes, that is, a single band may contain multiple sequences whose composition and relative abundance may vary with time. Thus, we have probably missed some sequences or genotypes occurring at other periods of the year because we sequenced only one or two bands for each band type. Moreover, the occurrence of a number of sequences within a single band may have biased our interpretation of assemblage structure, dynamic patterns of individual bands and the CCA. In this study, 85% of DGGE band types contained several (usually two) sequences, and nucleotide dissimilarity ranged up to 62%. Therefore, one band may refer to a subpopulation and its dynamic pattern could be a collective contribution from different T4-like myoviruses (if the multiple *g23* sequences of a band belong to viruses of different hosts which display

distinct temporal dynamics). The dynamical pattern of T4-like virus assemblages (both the whole assemblage and individual constituents) revealed by DGGE is thus just an index of banding/subpopulation patterns (not the sequence pattern).

By comparison with other fingerprinting methods (e.g. T-RFLP) used to evaluate the dynamic patterns of individual OTUs of T4-like myophages in seawater (Chow & Fuhrman, 2012; Needham *et al*., 2013; Pagarete *et al*., 2013), it seems that T-RFLP provides (in these cases) a better resolution in the separation of *g23* amplicons than DGGE. Indeed, a higher number of bands/OTUs per sample was detected (42 OTUs per sample in Norwegian fjords, 60 OTUs in eastern Pacific Ocean coastal seawater, 112 OTUs in SPOT seawater from the eastern Pacific Ocean) than in our study (34 bands per sample). This is probably due to the fact that T-RFLP has a higher sensitivity in discerning OTUs/bands than DGGE, as the former employs a laser to read data and is thus less restricted in the detection of dominant groups. Note that T-RFLP resolution also has some limitations, however, typically with the digestion that depends on the choice of the restriction enzymes for the target sequences. However, we did not detect such a bias after examining our *g23* sequences for the potential digestion site of RsaI/HincII [corresponding to enzymes used in studies of Chow & Fuhrman (2012), Pagarete *et al.* (2013) and Needham *et al.* (2013)] since all of them contained it and could have been digested.

Finally, light microscopy has limitations in detecting and identifying small cells (e.g. of size <5 μm, typically), and samples are generally only partially analysed. This explains the underestimation of unicellular cyanobacterial members of *Synechococcus* spp. using microscopy compared to counts obtained using FCM. The unicellular cyanobacteria detected by microscopy were identified more as *Synechocystis*-like than *Synechococcus*-like cells, but both scarce (Fig. 2). However, both the colonial and filamentous forms of the cyanobacteria were counted efficiently using microscopy.

In conclusion, our results, while not perfect, reveal that T4-like myoviruses are diverse in peri-alpine lakes, with an assemblage composition and component dynamics that can be very different between two neighbouring systems. We found that variation in the viral assemblage depended on a variety of individual constituents, apparently responding to host abundance and to environmental change. Dynamical patterns of the most abundant DGGE bands supported the seed bank model. It would be interesting in future to (i) identify unambiguously the hosts (unicellular, colonial or filamentous cyanobacteria and/or eubacteria) of the individual constituents; (ii) link individual T4-like myoviruses to the dynamics of particular bacteria; and (iii) carry out a multi-year analysis to examine whether there is regular seasonality in the occurrence of phages, as there is for other components of the plankton.

Acknowledgments

This work was supported by a fellowship from the Rhône-Alpes region (France) awarded to XZ. We thank Frédéric Rimet for the cyanobacteria counts and Angia Siram Pradeep Ram for the viral morphotype counts. We thank François Keck for figure support and Susan Lemprière for correcting the English and for her critical reading of the manuscript before its initial submission and after its revision. Environmental data were obtained from INRA CARRTEL observatory databasis referred to as IS SOERE GLACPE.

References

Bellas C.M. & Anesio A.M. (2013) High diversity and potential origins of T4-type bacteriophages on the surface of Arctic glaciers. *Extremophiles*, **17**, 861–870.

Berdjeb L., Pollet T., Chardon C. & Jacquet S. (2013) Spatio-temporal changes in the structure of archaeal communities in two deep freshwater lakes. *FEMS Microbiology Ecology*, **86**, 215–230.

Breitbart M. & Rohwer F. (2005) Here a virus, there a virus, everywhere the same virus? *Trends in Microbiology*, **13**, 278–284.

Breitbart M., Salamon P., Andresen B., Mahaffy J.M., Segall A.M., Mead D. *et al.* (2002) Genomic analysis of uncultured marine viral communities. *Proceedings of the National Academy of Sciences U.S.A.*, **99**, 14250–14255.

Butina T.V., Belykh O.I., Maksimenko S.Y. & Belikov S.I. (2010) Phylogenetic diversity of T4-like bacteriophages in Lake Baikal, East Siberia. *FEMS Microbiology Letters*, **309**, 122–129.

Butina T.V., Belykh O.I., Potapov S.A. & Sorokovikova E.G. (2013) Diversity of the major capsid genes (g23) of T4-like bacteriophages in the eutrophic Lake Kotokel in East Siberia, Russia. *Archives of Microbiology*, **195**, 513–520.

Castresana J. (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*, **17**, 540–552.

Chow C.-E.T. & Fuhrman J.A. (2012) Seasonality and monthly dynamics of marine myovirus communities. *Environmental Microbiology*, **14**, 2171–2183.

Comeau A.M. & Krisch H.M. (2008) The capsid of the T4 phage superfamily, the evolution, diversity, and structure of some of the most prevalent proteins in the biosphere. *Molecular Biology and Evolution*, **25**, 1321–1332.

Comeau A.M., Tremblay D., Moineau S., Rattei T., Kushkina A.I., Tovkach F.I. *et al.* (2012) Phage morphology recapitulates phylogeny, the comparative genomics of a new group of myoviruses. *PLoS ONE*, **7**, e40102.

Debroas D., Humbert J.-F., Enault F., Bronner G., Faubladier M. & Cornillot E. (2009) Metagenomic approach studying the taxonomic and functional diversity of the bacterial community in a mesotrophic lake (Lac du Bourget-France). *Environmental Microbiology*, **11**, 2412–2424.

DeLong E.F., Preston C.M., Mincer T., Rich V., Hallam S.J., Frigaard N.U. *et al.* (2006) Community genomics among stratified microbial assemblages in the ocean's interior. *Science*, **31**, 496–503.

Desplats C. & Krisch H.M. (2003) The diversity and evolution of the T4-type bacteriophages. *Research in Microbiology*, **154**, 259–267.

Druart J.C. & Rimet F. (2008) *Protocoles d'analyse du phytoplancton de l'INRA, prélèvement, dénombrement et biovolumes*. INRA, Thonon les Bains, pp. 1–200.

Duhamel S., Domaizon-Pialat I., Personnic S. & Jacquet S. (2006) Assessing the microbial community dynamics and the role of bacteriophages in bacterial mortality in Lake Geneva. *The Journal of Water Science*, **19**, 115–126.

Fancello L., Trape S., Robert C., Boyer M., Popgeorgiev N., Raoult D. *et al.* (2013) Viruses in the desert, a metagenomic survey of viral communities in four perennial ponds of the Mauritanian Sahara. *The ISME Journal*, **7**, 359–369.

Filée J., Tétart F., Suttle C.A. & Krisch H.M. (2005) Marine T4-type bacteriophages, a ubiquitous component of the dark matter of the biosphere. *Proceedings of the National Academy of Sciences U.S.A.*, **102**, 12471–12476.

Fujihara S., Murase J., Tun C.C., Matsuyama T., Ikenaga M., Asakawa S. *et al.* (2010) Low diversity of T4-type bacteriophages in applied rice straw, plant residues and rice roots in Japanese rice soils, estimation from major capsid gene (g23) composition. *Soil Science and Plant Nutrition*, **56**, 800–812.

Fujii T., Nakayama N., Nishida M., Sekiya H., Kato N., Asakawa S. *et al.* (2008) Novel capsid genes (g23) of T4-type bacteriophages in a Japanese paddy field. *Soil Biology and Biochemistry*, **40**, 1049–1058.

Guindon S. & Gascuel O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, **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 Symposium Series*, **41**, 95–98.

Huang H.Z., Cheng K., Xu M. & Zhao Y.J. (2011) Genetic diversity of T4 virioplankton, inferred from g23 gene, in Wuhan Donghu Lake. *China Environmental Science*, **31**, 443–447.

Huang S., Wang K., Jiao N. & Chen F. (2012) Genome sequences of siphoviruses infecting marine Synechococcus unveil a diverse cyanophage group and extensive phage-host genetic exchanges. *Environmental Microbiology*, **14**, 540–558.

Humbert J.-F., Dorigo U., Cecchi P., Le Berre B., Debroas D. & Bouvy M. (2009) Comparison of the structure and composition of bacterial communities from temperate and tropical freshwater ecosystems. *Environmental Microbiology*, **11**, 2339–2350.

Hurwitz B.L., Deng L., Polos B.T. & Sullivan M.B. (2013) Evaluation of methods to concentrate and purify ocean virus communities through comparative, replicated metagenomics. *Environmental Microbiology*, **15**, 1428–1440.

Jacquet S., Domaizon I., Personnic S., Duhamel S., Pradeep Ram A.S. *et al.* (2005) Estimates of protozoan and virus-mediated mortality of bacterioplankton in Lake Bourget (France). *Freshwater Biology*, **50**, 627–645.

Jacquet S., Domaizon I., Personnic S. & Sime-Ngando T. (2007) Do small grazers influence viral-induced bacterial mortality in Lake Bourget? *Fundamental and Applied Limnology*, **170**, 125–132.

Jamindar S., Polson S.W., Srinivasiah S., Waidner L. & Wommack K.E. (2012) Evaluation of two approaches for assessing the genetic similarity of virioplankton populations as defined by genome size. *Applied and Environmental Microbiology*, **78**, 8773–8783.

Jenkins C.A. & Hayes P.K. (2006) Diversity of cyanophages infecting the heterocystous filamentous cyanobacterium Nodularia isolated from the brackish Baltic Sea. *Journal of the Marine Biological Association of the United Kingdom*, **86**, 529–536.

Jia Z., Ishihara R., Nakajima Y., Asakawa S. & Kimura M. (2007) Molecular characterization of T4-type bacteriophages in a rice field. *Environmental Microbiology*, **9**, 1091–1096.

Jost G. & Wiese J. (2013) Temporal variations in the concentrations of bacteria and their lytic phages, an example of an indigenous phage host system in Lake Plußsee, Germany. *Fundamental and Applied Limnology*, **182**, 183–190.

Katoh K., Misawa K., Kuma K. & Miyata T. (2002) MAFFT, a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, **30**, 3059–3066.

King A.M.Q., Adams M.J., Carstens E.B. & Lefkowitz E.J. (2011) *Virus Taxonomy, Classification and Nomenclature of Viruses, Ninth Report of the International Committee on Taxonomy of Viruses*. Elsevier Academic Press, San Diego, CA.

Li Y., Watanabe T., Murase J., Asakawa S. & Kimura M. (2012) Identification of the major capsid gene (g23) of T4-type bacteriophages that assimilate substrates from root cap cells under aerobic and anaerobic soil conditions using a DNA–SIP approach. *Soil Biology and Biochemistry*, **63**, 97–105.

Liu J., Wang G., Wang Q., Liu J., Jin J. & Liu X. (2012) Phylogenetic diversity and assemblage of major capsid genes (g23) of T4-type bacteriophages in paddy field soils during rice growth season in Northeast China. *Soil Science and Plant Nutrition*, **58**, 435–444.

Liu J., Wang G., Zheng C., Yuan X., Jin J. & Liu X. (2011) Specific assemblages of major capsid genes (g23) of T4-type bacteriophages isolated from upland black soils in Northeast China. *Soil Biology and Biochemistry*, **43**, 1980–1984.

López-Bueno A., Tamames J., Velázquez D., Moya A., Quesada A. & Alcamí A. (2009) High diversity of the viral community from an Antarctic lake. *Science*, **326**, 858–861.

Lozupone C. & Knight R. (2005) UniFrac, a new phylogenetic method for comparing microbial communities. *Applied and Environmental Microbiology*, **71**, 8228–8235.

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

Martis R., Zurfluh K., Hagens S., Pianezzi J., Klumpp J. & Loessner M.J. (2013) Long tail fibres of the novel broad-host-range T-even bacteriophage S16 specifically recognize *Salmonella* OmpC. *Molecular Microbiology*, **87**, 818–834.

Middelboe M., Jacquet S. & Weinbauer M.G. (2008) Viruses in freshwater ecosystems: an introduction to the exploration of viruses in new habitats. *Freshwater Biology*, **53**, 1069–1075.

Nakayama N., Tsuge T., Asakawa S. & Kimura M. (2009) Morphology, host range and phylogenetic diversity of Sphingomonas phages in the floodwater of a Japanese paddy field. *Soil Science and Plant Nutrition*, **55**, 53–64.

Needham D.M., Chow C.-E.T., Cram J.A., Sachdeva R., Parada A. & Fuhrman J.A. (2013) Short-term observations of marine bacterial and viral communities, patterns, connections and resilience. *The ISME Journal*, **7**, 1274–1285.

Pagarete A., Chow C.-E.T., Johannessen T., Fuhrman J.A., Thingstad T.F. & Sandaa R.A. (2013) Strong seasonality and inter-annual recurrence in marine myovirus communities. *Applied and Environmental Microbiology*, **79**, 6253–6259.

Parvathi A., Zhong X. & Jacquet S. (2012) Dynamics of various viral groups infecting autotrophic plankton in Lake Geneva. *Advances in Oceanography and Limnology*, **3**, 171–191.

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

Personnic S., Domaizon I., Sime-Ngando T. & Jacquet S. (2009b) Seasonal variations of microbial abundances and virus- vs flagellate-induced mortality of picoplankton in three peri-alpine lakes. *Journal of Plankton Research*, **31**, 1161–1177.

Petrov V.M., Ratnayaka S., Nolan J.M., Miller E.S. & Karam J.D. (2010) Genomes of the T4-related bacteriophages as windows on microbial genome evolution. *Virology Journal*, **7**, 292.

Pradeep-Ram A.S., Arnous B., Danger M., Carrias J.-F., Lacroix G. & Sime-Ngando T. (2010) High and differential viral infection rates within bacterial 'morphopopulations' in a shallow sand pit lake (Lac de Créteil, France). *FEMS Microbiology Ecology*, **74**, 83–92.

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. *Systematic Biology*, **61**, 539–542.

Sandaa R.A. (2008) Burden or benefit-viral–host interaction in the marine environment. *Research in Microbiology*, **159**, 374–381.

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

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

Sime-Ngando T., Colombet J., Personnic S., Domaizon I., Dorigo U., Perney P. *et al.* (2008) Short-term variations in abundances and potential activities of viruses, bacteria and nanoprotists in Lake Bourget (France). *Ecological Research*, **23**, 851–861.

Sullivan M.B., Krastins B., Hughes J.L., Kelly L., Chase M., Sarracino D. *et al.* (2009) The genome and structural proteome of an ocean siphovirus: a new window into the cyanobacterial 'mobilome'. *Environmental Microbiology*, **11**, 2935–2951.

Suttle C.A. (2007) Marine viruses-major players in the global ecosystem. *Nature Reviews Microbiology*, **5**, 801–812.

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

Thingstad T. (2000) Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. *Limnology and Oceanography*, **45**, 1320–1328.

Thingstad T.F. & Lignell R. (1997) Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand. *Aquatic Microbial Ecology*, **13**, 19–27.

Thomas R., Berdjeb L., Sime-Ngando T. & Jacquet S. (2011) Viral abundance, production, decay rates and life strategies (lysogeny versus lysis) in Lake Bourget (France). *Environmental Microbiology*, **13**, 616–630.

Våge S., Storesund J.E. & Thingstad T.F. (2013) SAR11 viruses and defensive host strains. *Nature*, **499**, E3–E4.

Wang G., Hayashi M., Saito M., Tsuchiya K., Asakawa S. & Kimura M. (2009b) Survey of major capsid genes (g23) of T4-type bacteriophages in Japanese paddy field soils. *Soil Biology and Biochemistry*, **41**, 13–20.

Wang G., Jin J., Asakawa S. & Kimura M. (2009a) Survey of major capsid genes (g23) of T4-type bacteriophages in

rice fields in Northeast China. *Soil Biology and Biochemistry*, **41**, 423–427.

Wang G., Yu Z., Liu J., Jin J., Liu X. & Kimura M. (2011) Molecular analysis of the major capsid genes (g23) of T4-type bacteriophages in an upland black soil in Northeast China. *Biology and Fertility of Soils*, **47**, 273–282.

Weinbauer M.G. (2004) Ecology of prokaryotic viruses. *FEMS Microbiology Reviews*, **28**, 127–181.

Weinbauer M.G. & Rassoulzadegan F. (2004) Are viruses driving microbial diversification and diversity? *Environmental Microbiology*, **6**, 1–11.

Wilhelm S.W. & Matteson A.R. (2008) Freshwater and marine virioplankton, a brief overview of commonalities and differences. *Freshwater Biology*, **53**, 1076–1089.

Wolf A., Wiese J., Jost G. & Witzel K.P. (2003) Wide geographic distribution of bacteriophages that lyse the same indigenous freshwater isolate (*Sphingomonas* sp. strain B18). *Applied and Environmental Microbiology*, **69**, 2395–2398.

Yooseph S., Sutton G., Rusch D.B., Halpern A.L., Williamson S.J., Remington K. *et al.* (2007) The Sorcerer II Global Ocean Sampling expedition, expanding the universe of protein families. *PLoS Biology*, **5**, e16.

Zhao Y., Temperton B., Thrash J.C., Schwalbach M.S., Vergin K.L., Landry Z.C. *et al.* (2013) Abundant SAR11 viruses in the ocean. *Nature*, **494**, 357–360.

Zheng C., Wang G., Liu J., Song C., Gao H. & Liu X. (2013) Characterization of the major capsid genes (g23) of T4-type bacteriophages in the wetlands of Northeast China. *Microbial Ecology*, **65**, 616–625.

Zhong X., Berdjeb L. & Jacquet S. (2013) Temporal dynamics and structure of picocyanobacteria and cyanomyoviruses in two large and deep peri-alpine lakes. *FEMS Microbiology Ecology*, **86**, 312–326.

Zhong X., Colombet J., Pradeep-Ram A.S. & Jacquet S. (2014) Diversity of genome size, morphology, abundance and functional role of the virioplankton in Lakes Annecy and Bourget over a one-year period. *Microbial Ecology*, **67**, 66–82.

Zhong X. & Jacquet S. (2013) Prevalence of viral photosynthetic and capsid protein genes from cyanophages in two large and deep peri-alpine lakes. *Applied and Environmental Microbiology*, **79**, 7169–7178.

Zhong X. & Jacquet S. (2014) Contrasting diversity of phycodnavirus signature genes in two large and deep western European lakes. *Environmental Microbiology*, **16**, 759–773.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1. Bayesian phylogenetic tree based on aligning 260 homologous amino acid positions of the complete protein sequences of the *g23* gene from 30 T4-like phage isolates.

(Manuscript accepted 27 February 2014)