MOLECULAR ECOLOGY

Molecular Ecology (2015) **24**, 4812–4825 doi: 10.1111/mec.13321

FROM THE COVER

Ecological connectivity shapes quasispecies structure of RNA viruses in an Antarctic lake

A. LÓPEZ-BUENO, A. RASTROJO, R. PEIRÓ, M. ARENAS and A. ALCAMÍ

Department of Virology and Microbiology, Centro de Biología Molecular 'Severo Ochoa' (Consejo Superior de Investigaciones Científicas-Universidad Autónoma de Madrid), Nicolás Cabrera 1, Cantoblanco 28049, Madrid, Spain

Abstract

RNA viruses exist as complex mixtures of genotypes, known as quasispecies, where the evolution potential resides in the whole community of related genotypes. Quasispecies structure and dynamics have been studied in detail for virus infecting animals and plants but remain unexplored for those infecting micro-organisms in environmental samples. We report the first metagenomic study of RNA viruses in an Antarctic lake (Lake Limnopolar, Livingston Island). Similar to low-latitude aquatic environments, this lake harbours an RNA virome dominated by positive single-strand RNA viruses from the order *Picornavirales* probably infecting micro-organisms. Antarctic picorna-like virus 1 (APLV1), one of the most abundant viruses in the lake, does not incorporate any mutation in the consensus sequence from 2006 to 2010 and shows stable quasispecies with low-complexity indexes. By contrast, APLV2-APLV3 are detected in the lake water exclusively in summer samples and are major constituents of surrounding cyanobacterial mats. Their quasispecies exhibit low complexity in cyanobacterial mat, but their run-off-mediated transfer to the lake results in a remarkable increase of complexity that may reflect the convergence of different viral quasispecies from the catchment area or replication in a more diverse host community. This is the first example of viral quasispecies from natural aquatic ecosystems and points to ecological connectivity as a modulating factor of quasispecies complexity.

Keywords: Antarctica, ecological connectivity, metagenomics, *Picornavirales*, quasispecies, RNA virus

Received 21 April 2015; revision received 8 July 2015; accepted 10 July 2015

Introduction

Viral infection is one of the major selective pressures shaping the composition of microbial ecosystems (Suttle 2007; Rodriguez-Valera *et al.* 2009). Metagenomic analyses have revealed the enormous diversity of DNA viruses in nature (Angly *et al.* 2006; López-Bueno *et al.* 2009) and their potential to manipulate host genomes by lateral gene transfer (Sharon *et al.* 2009; Anantharaman *et al.* 2014). RNA viral communities have received less attention although their abundance can exceed that of DNA viruses in the oceans (Steward *et al.* 2013). Metagenomic surveys of RNA viruses in seawater (Culley *et al.* 2006; Steward *et al.* 2013), freshwater (Djikeng *et al.* 2009), reclaimed water (Rosario *et al.* 2009b), untreated sewage (Cantalupo *et al.* 2011; Ng *et al.* 2012) and hot springs (Bolduc *et al.* 2012) have shown a recurrent dominance of positive-sense single-strand RNA (+ssRNA) viruses, mainly belonging to the order *Picornavirales*. Unlike most available DNA viromes in seawater, which are composed predominantly of bacteriophages, marine RNA viruses seem to infect unicellular eukaryotes (Lang *et al.* 2009; Culley *et al.* 2014).

Only a few RNA viruses have been isolated from eukaryotic micro-organisms in aquatic ecosystems. Most of them are +ssRNA viruses that define a protistan clade within the *Picornavirales* order. This includes a virus infecting a marine algae (*Marnaviridae* family) (Lang *et al.* 2004), four diatom viruses (unclassified genus *Bacillarnavirus*) (Nagasaki *et al.* 2004; Shirai *et al.* 2008; Tomaru *et al.* 2009, 2012) and a fungoid protist

Correspondence: Antonio Alcamí, Fax +34 911964420;
E-mail: aalcami@cbm.csic.es

virus (unclassified genus *Labyrnavirus*) (Takao *et al.* 2005). Phylogenetically related viruses were identified in seawater by PCR amplification and Sanger sequencing (Culley *et al.* 2003, 2007) and up to six complete or nearly full-length genomes ascribed to the protistan clade were assembled from seawater metagenomic reads (Culley *et al.* 2006, 2014). On the contrary, only a single dsRNA virus of the *Reoviridae* family has been isolated from single-celled eukaryotic micro-organisms (Brussaard *et al.* 2004).

The protistan clade of *Picornavirales* was one of the most abundant viral groups in a metagenomic study of RNA viruses in freshwater (Djikeng *et al.* 2009), only outnumbered by *Dicistroviridae*, a family comprised of viruses infecting insects and crustaceans. Probably due to the connection with terrestrial environments, this lake encompasses a wide range of RNA viruses that infect insects, plants or animals. Metagenomic studies of RNA viruses in mammal faeces (Phan *et al.* 2011) and in untreated sewage water (Cantalupo *et al.* 2011; Ng *et al.* 2012) revealed higher diversity of RNA viral families than in natural aquatic ecosystems.

One of the best-studied limnological areas in maritime Antarctica is Byers Peninsula in Livingston Island (Toro *et al.* 2007; Villaescusa *et al.* 2010). The microbial dominated food webs of the numerous lakes in Byers Peninsula are poorly influenced by maritime fauna and provide a natural laboratory to study aquatic freshwater ecosystems (Laybourn-Parry 2009). Lake Limnopolar is an oligotrophic freshwater body located in the central plateau of the peninsula and harbours a wide range of bacteria, algae, rotifers and protozoa, and a macrozooplankton community restricted to the copepod *Boeckella poppei* and the fairy shrimp *Branchinecta gainii* (Toro *et al.* 2007). This lake remains ice-covered for at least 9 months per year under extreme weather conditions. Ice melting during the spring uncovers the lake water and the surrounding terrestrial environments in the catchment area such as cyanobacterial mats, lichens and moss carpets (Velázquez *et al.* 2013). Cyanobacterial mats show a widespread distribution in Antarctica and are complex ecosystems that often dominate total biomass and productivity (Laybourn-Parry & Pearce 2007). Seasonal ice melting is the most important physical shift affecting the lake biota due to the intense light irradiation, loss of water stratification and high rates of allochthonous nutrients input by run-off from terrestrial ecosystems of the catchment area (Camacho 2006).

Metagenomic analysis of DNA viruses in Lake Limnopolar reported a remarkable high diversity of viral families dominated by small circular genomes, many of them far-related to known eukaryotic ssDNA viruses (López-Bueno *et al.* 2009). Similar viruses were found in other aquatic environments (Rosario *et al.* 2009a, 2012; Labonte & Suttle 2013). In Lake Limnopolar, ice melting seems to have an important effect in the viral assemblage since the dominating ssDNA virus population in spring was replaced in summer by large dsDNA viruses (López-Bueno *et al.* 2009).

RNA viruses exist as populations of genetically distinct but closely related genomes known as quasispecies. This is due to their high mutation rates caused by error-prone viral RNA polymerases, high recombination rates, short generation times and large population sizes (Domingo *et al.* 2006). Complex mixtures of mutants provide genetic plasticity to RNA viruses that can be advantageous in changing environments. Quasispecies dynamics have been monitored under certain experimental conditions either in vitro (Domingo *et al.* 2012), following controlled animal host inoculation (Sanz-Ramos *et al.* 2008), and in natural infections of human and animal hosts (Vignuzzi *et al.* 2006; Bull *et al.* 2011). With the exception of the pioneer studies with Qβ and MS2 RNA bacteriophages (Horiuchi 1975), concepts related to quasispecies dynamics have been established in studies of animal and plant RNA viruses (Roossinck & Schneider 2006; Domingo *et al.* 2012) and also in some small DNA viruses (López-Bueno *et al.* 2006; Sarker *et al.* 2014). To our knowledge, the quasispecies nature of viruses infecting simpler eukaryotic organisms in natural settings remains unexplored. The advent of next-generation sequencing (NGS) technologies has provided an invaluable tool for addressing the complexity of quasispecies in unprecedented detail (Hoffmann *et al.* 2007; Bouquet *et al.* 2012) and could be applied to viruses highly represented in metagenomes.

Here we report the first metagenomic analysis of the RNA viral community from an Antarctic freshwater lake and the changes of the virome over a 4-year period. The seasonal and spatial distribution of four putative new members of the order *Picornavirales* within the catchment area is analysed. These viruses exhibit distinct patterns of quasispecies complexity that change when they are transported from cyanobacterial mats of the watershed to the lake water. These results highlight the influence of ecological connectivity on RNA virus quasispecies structure in nature.

Materials and methods

Samples

Lake Limnopolar is located in the inland plateau of the Byers Peninsula, at the western end of Livingston Island (62°34′35″ to 62°40′35″S and 60°54′14″ to 61°13′07″W), South Shetland Islands, Antarctica. The lake was sampled on the 27th of November 2006 when it was covered by a thick layer of ice (spring sample).

Two other water samples were collected on the 22nd of January 2007 and the 1st of February 2010 once the ice cover and almost the entire bulk of snow in the watershed had melted (summer samples) (Table S1, Supporting information). Five separated purple-pigmented microbial mats growing in the catchment of Lake Limnopolar were collected on the 2nd of January 2008 and stored at −20 °C until further processing.

Virus purification, random amplification of RNA viral genomes and sequencing

Cyanobacterial mat samples (2.5 g) were homogenized in SM buffer (50 mM Tris-HCl pH 7.5, 100 mM NaCl and 8 mM $MgSO_4.7H_2O$) by three cycles of vigorous vortex and sonication in a water bath during 20 and 10 s, respectively, then centrifuged at 3000 ***g*** for 5 min. This process was repeated twice. The resulting supernatants were combined and centrifuged at 8000 ***g*** for 1 h and then filtered through a 0.45-μm syringe filter (Millex, Durapore PVDF) to remove cellular organisms. The resulting viral fractions, as well as those from lake water samples, were purified as described previously (López-Bueno *et al.* 2009). In the case of the water sample collected in 2010, 0.45 μm filtration was carried out by TFF using two 0.093 m^2 polyethersulfone filter cassettes (Pall) and nuclease treatment of purified viral particles included 100 U/mL of nuclease S7 (Roche). RNA viral genomes were purified with Trizol-LS (Invitrogen) followed by DNAseI RNAse-free (Roche) treatment before randomly amplification by sequence independent single primer amplification (SISPA) (Victoria *et al.* 2008; Djikeng *et al.* 2009; Culley *et al.* 2010). Briefly, Superscript II or III and the Klenow fragment (3′–>5′exo-) enzyme (NEBiolabs) were used to convert RNA into dsDNA using 60 pmol of pseudo-degenerated primers FR26-RV (5′-GCCGGAGCTCTGCAGATA TCNNNNNN-3) for the water sample collected in 2007 and primer A (5′-GTTTCCCAGTCACGATANN NNNNNNN-3) for samples collected in 2006 and 2010. After 40 cycles of PCR amplification with FastStart high fidelity polymerase (Roche), DNA fragments between 500 and 3000 bp were gel-extracted with QIAquick Gel Extraction kit (Qiagen) and sequenced in the 454 GS FLX titanium platforms (Roche-454) from LifeSequencing (Valencia, Spain) or from Parque Científico de Madrid (Spain).

Quality filtration, assembly and taxonomic binning

Primer sequences were removed from 5′ and 3′ ends using the adaptor-clipping tool of BIOPIECES (http://code.google.com/p/biopieces/). PRINSEQ suite of tools (Schmieder & Edwards 2011) were used for quality filtering with the next parameters: -derep 124 -range_len 50-'virome mode + standard deviation' -lc_method entropy -lc_threshold 50 -ns_max_p 1 -ns_max_n 3 -trim_qual_left 20 -trim_qual_right 25 -trim_qual_type mean -trim_qual_window 2 -trim_qual_step 1 -min_qual_mean 25. Sequences were assembled *de novo* under strict parameters (97% of minimum identity along 90% of read length) with NEWBLER 2.5.3 (Roche) and CLC GENOMIC WORKBENCH 5 (http://www.clcbio.com/). The full-length genomes of four Antarctic RNA viruses were obtained using information from both assemblers and joining contigs by PCR amplification and Sanger sequencing (Table S2, Supporting information). The 3′ ends were PCR-amplified with specific and poly-T primers and sequenced with Sanger technology.

Taxonomic binning of reads and contigs was based on the best hit of BLASTX searches (E-value <0.001) against the GenBank nr protein database and TBLASTX searches (E-value <0.001) against reference viral genomes (4602 reference viral genomes downloaded from the GenBank on the 10th of April 2013) using in-house computational resources.

Phylogenetic analysis and sequence-signature categorization

To assess the phylogenetic affiliation of Antarctic *Picornavirales*, they were aligned (CLUSTAL OMEGA) with a comprehensive set of protein sequences containing the RdRp domain (pfam00680) from GenBank and some published aquatic viromes. The alignment was trimmed with JALVIEW (http://www.jalview.org/) in such a way that only the region encompassing the positions corresponding to amino acids 159 and 381 of encephalomyocarditis virus replication protein was maintained. Less informative regions were removed with TRIMAL v1.2 (Capella-Gutierrez *et al.* 2009). A maximum-likelihood phylogenetic tree was constructed with bootstrap 100, under the WAG substitution model in R (package PHANGORN), and drawn with DENDROSCOPE (Huson & Scornavacca 2012).

Dinucleotide composition bias over the expected frequencies was calculated using the COMPSEQ tool (EMBOSS package) for a set of 131 available complete viral RNA genomes. Principal component analysis (PCA) based on dinucleotide composition bias was calculated in R (package ADE4).

Single nucleotide variant calling

Three alternative methods for single nucleotide variant (SNV) calling were used. First, we applied the suite of tools of CLC GENOMIC WORKBENCH to further trim quality-filtered reads by setting a quality limit of 0.05.

Next, resulting sequences were aligned to reference genomes with a minimum identity of 90% along 90% of the read length. Then, we followed a conservative method for SNV calling using the Quality-based Variant Detection tool setting a minimum neighbourhood quality of 15, minimum central quality of 20 and filtering 454 homopolymer indels. Second, quality-filtered sequences were trimmed again with PRINSEQ in the last 15 bp of the 3′ end and were aligned against reference genomes using BOWTIE 2 (Langmead & Salzberg 2012) under strict parameters (–np 0 –n-ceil L,0,0.02 –rdg 0,6 –rfg 0,6 –mp 6,2 –score-min L,0,-0.2), and SNV calling was performed with the MPILEUP tool implemented in SAMTOOLS (Li *et al.* 2009) with and without specifying the integrated probabilistic realignment option. Third, we used *Shorah* (McElroy *et al.* 2013) to correct quality-filtered and trimmed reads under a local probabilistic clustering. Then, these corrected reads were aligned to reference genomes with BOWTIE 2 and SNV calling carried out with MPILEUP as above but with the probabilistic realignment tool disable. In the three methods, only nucleotide differences with respect to the reference sequence represented by at least two reads, with a minimum frequency of 1%, distributed on both strands of the DNA and in regions with a minimum coverage of 25× were validated as SNV.

Quasispecies complexity indexes

To determine quasispecies complexity, we developed a framework that utilizes validated SNV tables as input to estimate the Shannon index (*Sn*) and nucleotide diversity (π). The Shannon index was calculated following the next equation:

$$S = -\Sigma_{i=A,C,G,T}\Sigma_{k=1}^{n} P_{ik} \log_2 P_{ik}$$

where P_{ik} represents the probability of *i*th base in *k*th position. The resulting Shannon index was divided by the maximum theoretical entropy ($n \log_2 4$; where n is the number of analysed positions) to obtain a normalized Shannon index (*Sn norm*). To estimate nucleotide diversity (π) from validated lists of SNVs, we computed the average number of different nucleotide states (at every position) between any two sequences:

$$\pi = \frac{1}{n}\Sigma_{k=1} P(i \neq j; P_k)$$

where n is the total number of analysed positions and $P(i \neq j; P_k)$ is the proportion of different that *i*th and *j*th nucleotide states at a particular position k. Finally, mean diversity was calculated as the percentage of validated SNV positions divided by the total number of positions with coverage >25×.

Nonsynonymous to synonymous rates ratio (*dN*/*dS*) was estimated using KA/KS CALCULATOR (http://services.cbu.uib.no/tools/kaks) for a theoretical sequence with all validated SNVs collapsed. We also used the single likelihood ancestor counting (SLAC) method implemented in the HYPHY package (Kosakovsky Pond & Frost 2005) to estimate *dN*/*dS* from 100 simulated data sets of 500 sequences reflecting the observed frequencies of SNVs. Finally, we developed a framework that estimates *dN*/*dS* globally by calculating the expected and the observed nonsynonymous and synonymous substitutions for each codon accounting for the total number of evolutionary pathways between codon states (Nei & Gojobori 1986). The best-fit model of nucleotide substitution for the second and third methods was selected with JMODELTEST2 (http://code.google.com/p/jmodeltest2/).

Results

The RNA viromes from Lake Limnopolar are dominated by viruses related to Picornavirales

Water samples were collected from Lake Limnopolar (Livingston Island, Antarctica) during the spring of 2006, when the lake was covered by ice, and in the summers of 2007 and 2010 after ice melting and under high run-off regimes in the basin. RNA viral genomes were extracted from purified virus particles, randomly amplified and subjected to Roche-454 pyrosequencing (Table S3, Supporting information). Most of the assigned reads were related to known viruses based on the best BLAST hit (Fig. 1). The 2006 RNA virome sequences related to DNA viral families *Microviridae* and *Circoviridae* reflect a contamination with DNA viral genomes from the same sample since they exhibited 100% nucleotide identity with the previously reported DNA viromes (López-Bueno *et al.* 2009). The vast majority of the sequences from the three RNA viromes were related to viruses from the order *Picornavirales* as found in other aquatic environments (Culley *et al.* 2003, 2006, 2007; Djikeng *et al.* 2009; Rosario *et al.* 2009b; Cantalupo *et al.* 2011; Ng *et al.* 2012). *Dicistroviridae* was the most represented family of RNA viruses from 2006 to 2010. On the contrary, the genus *Bacillarnavirus* and the very closely related unclassified/environmental *Picornavirales* were almost restricted to summer samples collected in 2007 and 2010. In decreasing order, the next most abundant RNA viral families in Lake Limnopolar virome corresponded to *Secoviridae*, *Marnaviridae*, *Iflaviridae*, *Potyviridae*, *Picornaviridae* and *Tombusviridae*.

Sequencing reads were assembled into contigs under strict parameters. These contigs showed a viral family profile similar to that observed for metagenomic reads with most of them, including those with highest coverage, assigned to *Dicistroviridae*, *Bacillarnavirus* and

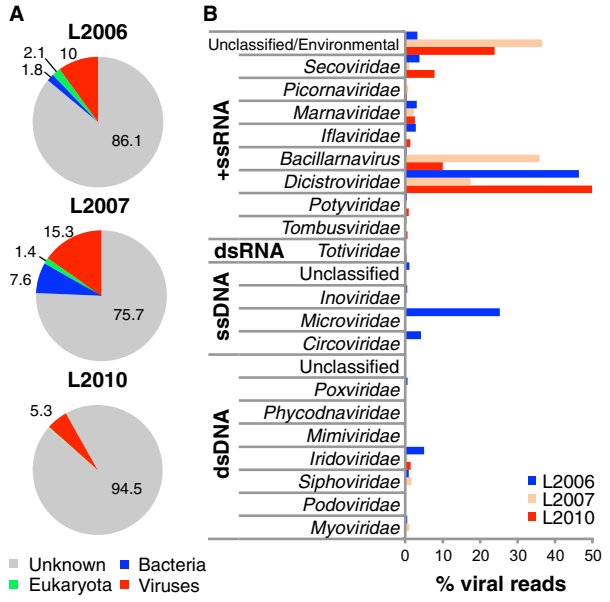

Fig. 1 Taxonomy overview of RNA viromes from Lake Limnopolar. BLASTX (E-value <0.001) comparison of metagenomic reads (>50 bp) against GenBank nr (A) and reference viral genomes (B) data sets. The percentage of reads assigned to different domains and viruses (A) and the percentage of viral sequences assigned to viral families (B) are shown. Viral families represented by <10 reads or 0.2% of the total viral reads assigned are omitted.

related environmental *Picornavirales* (Fig. S1 and Table S4, Supporting information). Interestingly, a higher proportion of contigs (27.0–49.3%) compared to reads (5.3–15.3%) were assigned to known viral families. Moreover, 41.8–66.6% of the metagenomic reads were assembled into contigs ascribed to known viral families. This indicated that many of the metagenomic reads classified as unknown by BLAST belong to highly diverse regions of viral genomes poorly related to those available in databases.

Nearly full-length genome sequence of four Antarctic picorna-like viruses

Among the contigs with highest coverage, four nearly full-length viral genomes or genomic segments could be reconstructed with lengths from 8644 to 9357 bp (Fig. S2, Supporting information). Since their predicted ORFs showed the best genetic similarity to members of the *Picornavirales* order (Table S5, Supporting information), we have tentatively named them as Antarctic picorna-like virus (APLV) 1–4. Consistently, all of them harbour a large ORF encoding characteristic *Picornavirales* nonstructural proteins with helicase (pfam00910) and RNA-dependent RNA polymerase (RdRp; pfam00680) domains. In addition, APLV2 and APLV4 also harbour a peptidase C3 domain (pfam00548), typically found in nonstructural proteins of *Picornavirales*. APLV1, APLV2 and APLV4 encode a structural protein with domains conserved among capsids of *Picornavirales* such as picornavirus capsid domain (Rhv_like; pfam00073) or cricket paralysis virus capsid domain (CRPV-C; pfam08762), the latter only found in APLV2 and APLV4. This structural protein type was not encoded by APLV3, suggesting that this virus bears a segmented genome. Attempts to identify a second genomic segment of APLV3 have been unsuccessful. These four new *Picornavirales* were genetically poorly related to each other since alignment of the most conserved domains revealed amino acid identity values <35% for the RdRp domain and <41% for the Rhv_like domain (Table S6, Supporting information).

A phylogenetic analysis based on the RdRp domain, the best molecular markers within the *Picornavirales* order (Koonin 1991), showed that APLV1 clusters with viruses of the *Dicistroviridae* family (Fig. 2), known to infect insects. This result is in agreement with its best BLAST similarity to Nedicistrovirus TNF-2012 (Table S5, Supporting information) (Ng *et al.* 2012). On the other hand, APLV2, APLV3 and APLV4 cluster with other *Picornavirales* known to infect micro-organisms such as diatoms (*Bacillarnavirus*), algae (*Marnaviridae*) or fungoid protist (*Labyrnavirus*), consistent with the microbial dominated nature of Lake Limnopolar. Methods to infer the host of APLVs based on dinucleotide composition (Kapoor *et al.* 2010) suggested that they infect microbes or insects, and not plants, but we could not successfully separate microbial and insect *Picornavirales* (Fig. S3, Supporting information).

We also monitored the presence and abundance of APLV1-APLV4 in Lake Limnopolar along a period of 3 years by mapping metagenomic reads against their genomes with CLC GENOMIC WORKBENCH (Table 1), although similar results were obtained with NEWBLER or BOWTIE 2 (not shown). APLV1 was the most abundant RNA virus in 2006, 2007 and 2010 viromes. Taking into consideration the clustering of APLV1 with viruses of the family *Dicistroviridae* (Fig. 2), these percentages fit well the number of sequencing reads assigned to this family (Fig. 1). APLV2 and APLV3 were also abundant in samples collected during the summers of 2007 and 2010 but not detected at all in the spring of 2006 when the lake was covered by a thick layer of ice. The seasonality of these two viruses was consistent with the summer increase of *Bacillarnavirus* and unclassified/environmental *Picornavirales* (Fig. 1). Comparison of the consensus sequence of APLV2 and APLV3 from 2007 to 2010 revealed 13 and 20 substitutions, respectively, whereas APLV1 showed a higher level of genetic stability since its consensus sequence remained unaltered

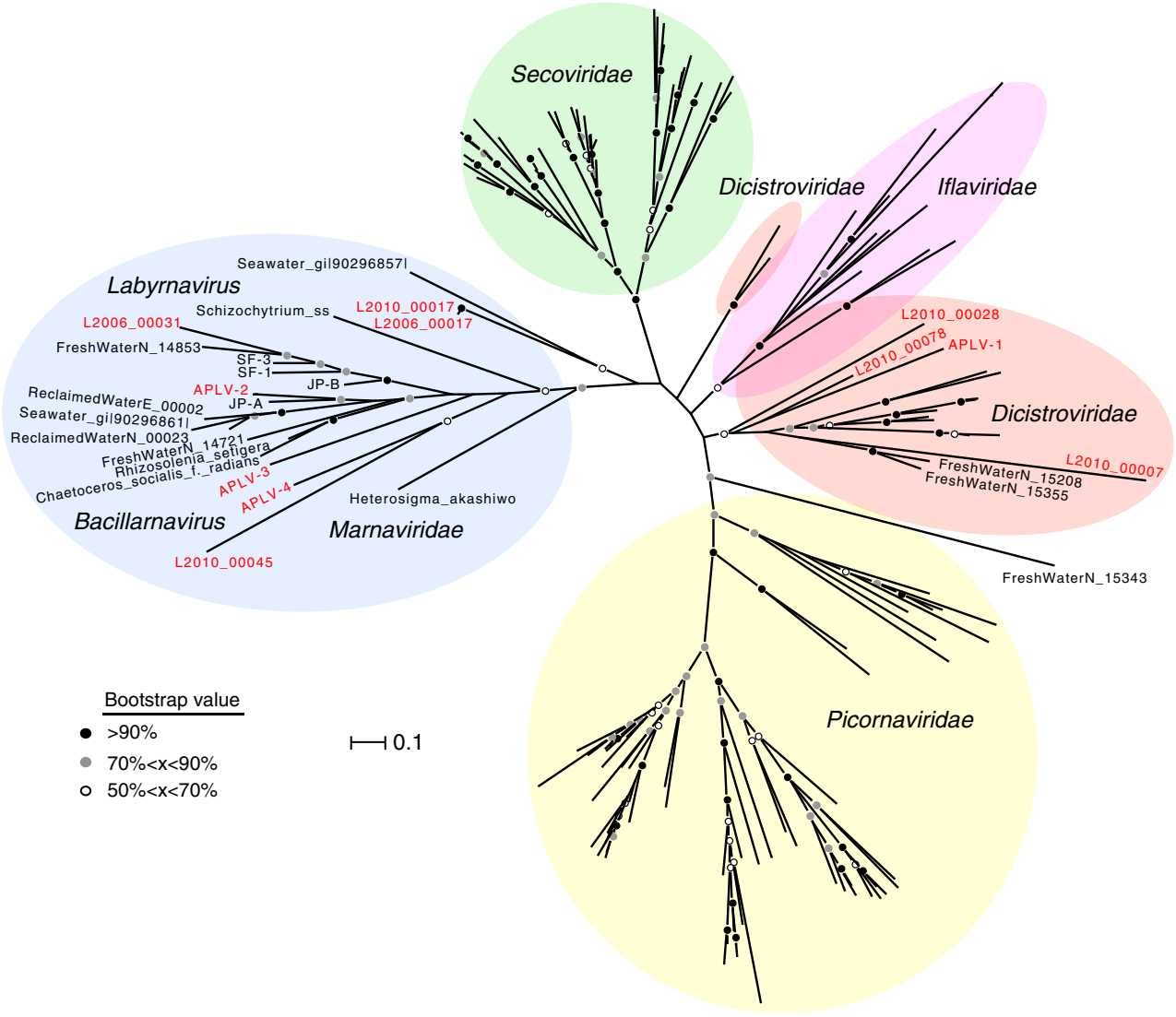

Fig. 2 Phylogenetic analysis of Antarctic picorna-like viruses (APLVs). Contigs and APLVs from RNA viromes of Lake Limnopolar containing the RdRp (pfam00680) (labelled in red colour) were aligned with a set of RdRp proteins of the order *Picornavirales* and aquatic environments. An unrooted maximum-likelihood tree with bootstrap values based on 100 replicates was constructed. The scale bar indicates the number of amino acid substitutions per residue.

from 2006 to 2010 (Fig. S4A, Supporting information). Finally, APLV4 was only detectable in the lake water in 2010 and constituted more than 10% of all metagenomic reads (Table 1).

Monitoring quasispecies structure of four RNA viruses in a natural setting

The abundance of a limited number of RNA viruses provided us with a set of new viruses sequenced to a high coverage (Table 1). We explored the quasispecies structure of these four viruses by calling SNVs. There are a number of caveats to consider when analysing virus quasispecies from deep-sequencing data such as the uneven coverage along viral genomes and the difficulty to distinguish between true changes and sequencing errors. To address these issues, we have followed a strict quality filtration process of the reads and a conservative approach for validation of SNVs, as described in Materials and Methods. APLV1, an abundant RNA virus in the lake viromes from 2006 to 2010, shows a quasispecies with low number of SNVs equally distributed along the genome, and only a few of them appeared in >10% of the sequences (Fig. 3A). Consistently, this virus exhibited low values of average SNV frequency, Shannon index and nucleotide diversity (Table 2). This is in agreement with the stability of its consensus sequence, the remarkable preservation of

most abundant SNVs (Fig. S4B,C, Supporting information) and the lack of statistical differences in SNV frequency, nucleotide diversity and Shannon index among APLV1 quasispecies assessed in 2006, 2007 and 2010 (Fig. S5, Supporting information). On the contrary, APLV2, APLV3 and, to a lesser extent, APLV4 harboured quasispecies with a higher number of SNVs, many of them with frequencies >10% (Fig. 3A). Correspondingly, the average SNV frequency, Shannon index and nucleotide diversity show values higher than those obtained for APLV1 (Table 2). Importantly, these radically different patterns of quasispecies complexity between APLV1 and the three other viruses were maintained in time from 2006 to 2010 and occur within the same viral community (APLV1 vs. APLV2-APLV3 in 2007 or 2010 samples), or between

Table 1 APLVs' abundance and distribution among samples

	L2006		L2007		L2010		CyaMat	
Virome	Reads (%)	Cover.(x)*	Reads (%)	Cover.(x)	Reads (%)	Cover.(x)	Reads (%)	Cover.(x)
APLV1	27 145 (71.0)	1162	10 654 (19.2)	480	59 918 (71.6)	2457	ND	ND
APLV2	ND†	ND	9983 (18.0)	399	2467 (2.9)	95	3054 (11.3)	103
APLV3	ND	ND	6716 (12.1)	281	2595 (3.1)	108	4829 (17.8)	167
APLV4	ND	ND	ND	ND	9021 (10.8)	340	ND	ND

APLV, Antarctic picorna-like virus.
*Coverage.
†Nondetected.

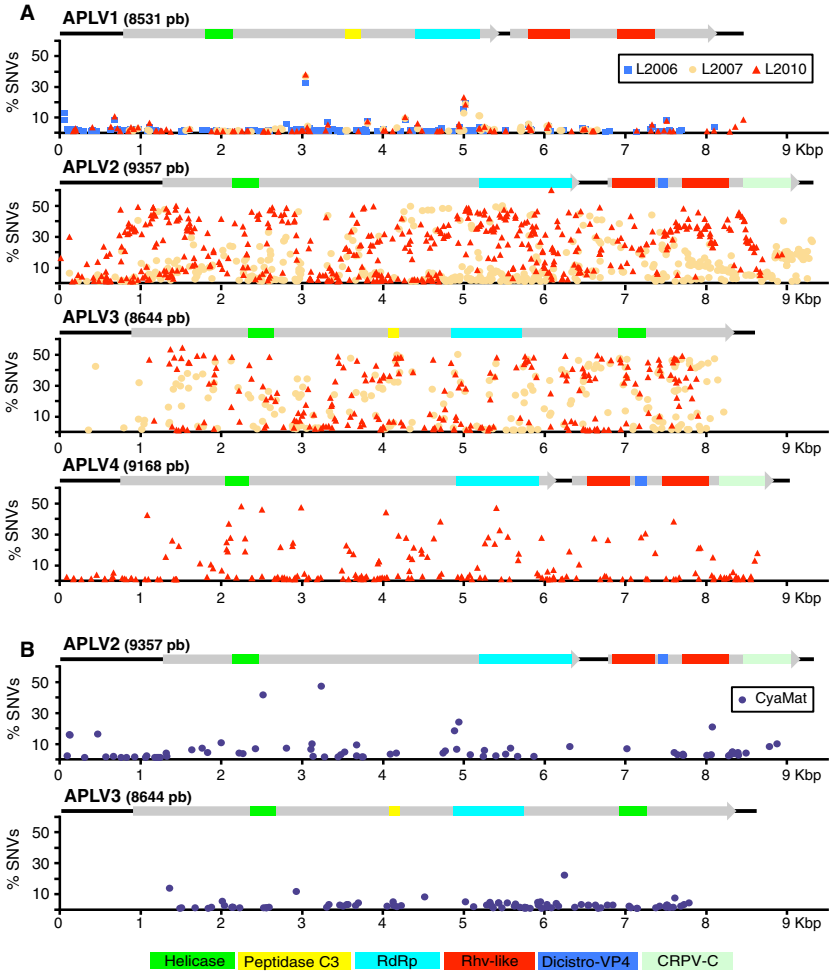

Fig. 3 Quasispecies of Antarctic picorna-like viruses (APLVs). The distribution of single nucleotide variant (SNV)s along four APLV genomes is shown. As indicated, each coloured dot represents SNVs with frequencies higher than 1% forming the quasispecies of a virus in a particular sample. Genome organization schemes are drawn above graphs to properly locate the SNVs. Quasispecies of viruses present in the Lake Limnopolar water in 2006, 2007 or 2010 (A), and in cyanobacterial mats (CyaMat) (B) are represented. SNV calling shown in this figure is based on the Quality-based Variant Detection tool of CLC GENOMIC WORKBENCH.

Table 2 Quasispecies complexity indexes and *dN/dS* estimates of APLVs

Virus		APLV1			APLV2			APLV3			APLV4
Sample		L2006	L2007	L2010	L2007	L2010	CyaMat	L2007	L2010	CyaMat	L2010
No. SNVs*		134	60	94	475	492	79	212	218	113	199
Coverage†		831	552	2190	447	90	122	234	124	200	318
Average quality		34.6	35.3	33.6	35.4	33.5	35.4	35.5	33.6	34.8	33.3
Length studied‡		8333	6497	8454	8877	8947	8056	7833	7197	7654	8758
Mean diversity§		1.61	0.92	1.11	5.35	5.50	0.98	2.71	3.03	1.48	2.27
SNV frequency¶		2.58	3.59	3.22	11.89	24.00	5.92	21.17	23.18	3.55	8.70
Ti/Tv**		11.18	6.50	6.23	4.22	3.82	5.58	4.17	3.11	5.65	2.62
Mutation frequency††		2,9E-04	2,3E-04	4,2E-04	5,9E-03	1,1E-02	5,6E-04	4,1E-03	6,7E-03	5,7E-04	1,5E-03
Sn‡‡		20.8	11.4	16.6	208.8	339.6	21.8	132.8	138.4	22.1	65.3
Sn norm§§		0.0012	0.0009	0.0010	0.0118	0.0190	0.0014	0.0085	0.0096	0.0014	0.0037
π (%)¶¶		0.0766	0.0585	0.0639	0.9641	1.7445	0.0974	0.7703	0.8780	0.0944	0.3012
dN/dS	Orf-1	0.41	0.30	0.35	0.05	0.07	0.32	0.08	0.12	0.23	0.18
(KAKS-CALC.)***	Orf-2	0.15	0.18	0.11	0.14	0.14	0.22				0.26
dN/dS	Orf-1	0.63	0.67	0.68	0.05	0.07	0.39	0.08	0.10	0.13	0.12
(HYPHY)†††	Orf-2	0.16	0.14	0.18	0.18	0.15	0.27				0.28
dN/dS	Orf-1	0.26	0.19	0.20	0.04	0.06	0.23	0.06	0.10	0.17	0.17
(In-house)‡‡‡	Orf-2	0.15	0.17	0.12	0.12	0.13	0.18				0.31

SNV, single nucleotide variant; APLV, Antarctic picorna-like virus.
*Number of SNVs >1%.
†Average coverage for SNV sites.
‡Length studied (coverage >25×; bp).
§% SNV sites.
¶Average percentage of SNVs.
**Nucleotide transitions/transversions ratio.
††Number of mutations divided by total number of nucleotides sequenced.
‡‡Shannon index.
§§Normalized Shannon index.
¶¶Nucleotide diversity.
***Global *dN/dS* estimated by KAKS CALCULATOR software.
†††Global estimated by HYPHY package for simulated reads.
‡‡‡Global dN/dS estimated by a dedicated framework (In-house).

viruses sequenced with the same coverage (APLV1 and APLV2 in 2007) ruling out the influence of putative experimental variations or sequence depth, respectively. Furthermore, these differences could be confirmed using several SNV calling methods (Fig. S6 and Table S7, Supporting information). These results show that RNA viruses in aquatic ecosystems can present very different quasispecies structures and that low-complexity quasispecies can be found in viruses with high ecological success.

Ecological connectivity affects RNA virus quasispecies entropy

Melting of the ice cover in Lake Limnopolar during the transition from spring to summer coincides with an important supply of meltwater and organic matter from terrestrial environments of the catchment area. The presence of APLV2 and APLV3 in the lake water from summer samples (2007 and 2010) but not from the spring sample (2006) might be explained by seasonal supply of biomass from the Limnopolar watershed through run-off. To assess this hypothesis, we checked the presence of APLV1-APLV4 in a virome obtained from a pool of five surrounding cyanobacterial mats, which is one of the most active terrestrial ecosystems in the catchment area of Lake Limnopolar (Velázquez *et al.* 2013). As shown in Table 1, APLV2 and APLV3 are highly abundant viruses in cyanobacterial mats, supporting the notion that viruses from other terrestrial environments can reach the lake every season. A completely different pattern was observed for APLV1. This virus was abundant in Lake Limnopolar irrespective of the season and was not detected in cyanobacterial mats. Five additional large contigs corresponding to putative *Picornavirales* were assembled from the cyanobacterial mat virome, but, similar to APLV2 and APLV3, only two of them were present in summer samples of Lake Limnopolar (Table S8, Supporting information).

Noteworthy, for a similar coverage, the quasispecies structure of APLV2 and APLV3 exhibited a remarkable shift in their complexity indicators in cyanobacterial mats vs. lake water. These viruses showed fewer SNVs represented at lower percentages in cyanobacterial mats, and consequently, the estimated mutation frequency, nucleotide diversity and Shannon index values were around 10 times lower than in the Lake Limnopolar water (Fig. 3B, Table 2, Fig. S5, Supporting information).

To gain insights into the evolutionary mechanisms of adaptation followed by these viruses in each environment, we estimated the ratio of nonsynonymous to synonymous substitution rates (*dN*/*dS*) along their ORFs. Global ORF *dN*/*dS*, estimated by three alternative methods (Table 2, Table S9, Supporting information), evidenced a general purifying selection in all the analysed data. However, some differences were observed when comparing APLV2 and APLV3 *dN*/*dS* estimates from cyanobacterial mats and lake water (Table 2, Fig. S5, Table S9, Supporting information). In particular, *dN*/*dS* was higher in viruses collected from cyanobacterial mats than in viruses collected from the lake water.

Discussion

Following previous efforts to characterize virus assemblages in Antarctica (López-Bueno *et al.* 2009), this report describes for the first time the community of RNA viruses in aquatic polar ecosystems and the changes occurring after a 4-year period. Consistent with RNA viromes from seawater and temperate freshwater lakes where one or two genotypes accounted for >60% of the metagenomic reads (Culley *et al.* 2003, 2006; Djikeng *et al.* 2009; Steward *et al.* 2013), a few +ssRNA viruses were found to be highly abundant in the Antarctic viromes reported here. Every step in the manipulation of a natural viral community is a potential source of bias (Wen *et al.* 2004; Willner *et al.* 2011); therefore, we should be cautious when interpreting metagenomes in quantitative terms. We have followed a conservative experimental approach that avoids the use of known sources of bias such as chloroform treatment (damages the lipid envelope surrounding the viral capsid of some viruses) and CsCl buoyant gradients (biased towards tailed phages). On the other hand, the PCR-based random amplification method (SISPA) might introduce a modest overamplification of the most abundant genomes (Karlsson *et al.* 2013). This might explain, at least in part, that up to 71, 49 and 88% of the metagenomic reads align to the four *Picornavirales* genomes reconstructed from the 2006, 2007 and 2010 samples, respectively. On the contrary, aquatic DNA viromes bear higher diversity irrespective of the dominance of bacteriophages (Angly *et al.* 2006) or putative small eukaryotic viruses like in temperate freshwater lakes (Roux *et al.* 2012) and even in Lake Limnopolar (López-Bueno *et al.* 2009).

Lake Limnopolar is dominated by micro-organisms (Toro *et al.* 2007). Correspondingly, BLAST and phylogenetic analyses suggest that RNA viruses in Lake Limnopolar belong to *Picornavirales* infecting Arthropoda or unicellular eukaryotes. The phylogenetic clustering of APLV1, with insect viruses (*Dicistroviridae*), suggests that this virus may infect Arthropoda inhabiting Lake Limnopolar (copepod *Boeckella poppei*, fairy shrimp *Branchinecta gainii* or chironomid *Parochlus steinenii*) (Toro *et al.* 2007). However, these putative hosts represent only a small portion of the eukaryote biomass in the water column, which is mainly composed of phytoplankton (Toro *et al.* 2007; Rochera *et al.* 2013). Thus, it seems unlikely that an arthropod could host the most abundant RNA virus in the lake. Moreover, the deep branching of APLV1 and the closest related Nedicistrovirus found in sewage (Ng *et al.* 2012) in phylogenetic trees suggest that they might belong to poorly characterized dicistroviruses infecting a nonarthropod host. Attempts to identify the APLV1 host based on dinucleotide composition failed due to the overlap between *Picornavirales* infecting insects and micro-organisms. Further research is required to accurately identify the host origin of many RNA viruses assigned to the *Dicistroviridae* family found in aquatic ecosystems (Culley *et al.* 2006; Djikeng *et al.* 2009; Rosario *et al.* 2009b; Cantalupo *et al.* 2011; Ng *et al.* 2012).

APLV2, APLV3 and APLV4 clustered in a phylogenetic tree with *Picornavirales* of the protistan clade. Consequently, it is probably that they infect diatoms, chrysophytes or chlorophytes, the most abundant eukaryotic micro-organisms in the Lake Limnopolar water (Toro *et al.* 2007). APLV2 and APLV3 were abundant viruses in summer samples of 2007 and 2010 but undetectable in the 2006 spring sample. Their temporal coincidence in cyanobacterial mats and in the water column of the lake during the summer season is accompanied by an intensive run-off from the catchment area suggesting an allochthonous origin for these viruses. The enormous productivity of terrestrial ecosystems such as cyanobacterial mats and moss in Byers Peninsula (Velázquez *et al.* 2013) and their influence on the Antarctic lakes is well documented (Camacho 2006). A deep-sequencing analysis of 18S rRNA genes from the same cyanobacterial mats surrounding Lake Limnopolar (D. Velázquez, A. López-Bueno, D. Aguirre de Cárcer, A. de los Ríos, A. Alcamí, A. Quesada. Manuscript submitted for publication) revealed a dominance of pennate diatoms (85%). Some of these diatoms might be the host of abundant Antarctic *Picornavirales* such as

APLV2 and APLV3. Indeed, the structural protein of APLV2 shows best similarity to that of *Asterionellopsis glacialis* RNA virus: the only available *Bacillarnavirus* that infect the pennate diatom (Table S5, Supporting information) (Tomaru *et al*. 2012).

Once in the lake, and although APLV2 and APLV3 represent a notable proportion of the viral assemblage in two nonconsecutive summers (2007 and 2010), they may persist in the water column only for a limited time period as suggested by their absence in the spring sample before ice melting. Conversely, APLV1 was not detected in surrounding cyanobacterial mats and occupies a dominant position in Lake Limnopolar irrespective of the season. It is conceivable that the seasonal lost of the ice cover in Antarctic lakes promotes shifts in the water biota including viral communities; however, a higher number of metagenomes are required to confirm the observed DNA (López-Bueno *et al.* 2009) and RNA viruses' (this study) seasonal changes. Interestingly, a similar shift was observed in RNA viromes from a temperate freshwater lake, which was proposed to be due to the supply of biomass from surrounding flora and fauna (Djikeng *et al*. 2009).

DNA viral assemblages of seawater have been shown to be composed mainly of tailed bacteriophages (Angly *et al.* 2006), but electron microscopy analysis of a seawater surface ecosystem revealed a dominance of small nontailed viruses (Brum *et al.* 2013). This inconsistency may be explained by a similar abundance of DNA and nontailed eukaryotic RNA viruses in seawater (Culley *et al.* 2006; Steward *et al.* 2013). Lake Limnopolar is dominated by eukaryotic RNA viruses (*Picornavirales*), as well as ssDNA viruses related to eukaryotic viral families. This is in agreement with the observed abundance of small nontailed particles by electron microscopy (López-Bueno *et al.* 2009). In freshwater ecosystems (Takacs & Priscu 1998; Rochera *et al.* 2013), as in surface seawater (Caron *et al*. 1995; Massana 2011), unicellular eukaryotes, particularly picoeukaryotes, account for the largest biomass of the phytoplankton that can even exceed cyanobacterial biomass of high-latitude oceans (Cuvelier *et al*. 2010). Microbial eukaryotic abundance, diversity and productivity, together with the high burst size of their RNA (Lang *et al.* 2009) and ssDNA (Tomaru *et al.* 2011) viruses, might explain the dominance of eukaryotic virus-related sequences in aquatic viromes (Culley *et al.* 2006; López-Bueno *et al.* 2009; Roux *et al.* 2012; Steward *et al*. 2013).

The deep coverage of the APLV1-APLV4 genome sequences allowed us to explore, for the first time, the quasispecies structure of viruses in natural aquatic ecosystems. Since we were comparing closely related genomes that differ among them in a far lower number of nucleotide mutations (0.9–5.5% considering a theoretical sequence with all validated SNVs) than different APLVs of the same habitat (22.4–34.4% aa identity found among the conserved RpRd domain; Table S6, Supporting information), we adopted the term quasispecies as presently used in experimental virology (Domingo *et al*. 2012). The quasispecies of APLVs along 4 years unveiled two distinct patterns of complexity. APLV1, the most abundant RNA virus in Lake Limnopolar from 2006 to 2010, exhibits low values of normalized Shannon entropy and nucleotide diversity. These nucleotide diversity levels are lower than those found in some animal RNA viruses such as hepatitis C virus (Sakai *et al.* 1999; Gregori *et al.* 2013), but are in the range of those estimated for zoonotic viruses, such as hepatitis E virus (Bouquet *et al.* 2012) or West Nile virus (Jerzak *et al.* 2005). The low level of complexity of the APLV1 quasispecies is concomitant with its remarkable genetic stability based on identical consensus sequences from 2006 to 2010 and a fine preservation of SNV spectra. Similar quasispecies stability has been previously reported for a human strain of hepatitis E during the infection of pigs, suggesting that this virus is well adapted to both hosts (Bouquet *et al.* 2012). In Lake Limnopolar, APLV1 genetic stability is suggestive of a perfect adjustment of the quasispecies to a stable host or several hosts. This might be supported by the detected signatures of strong purifying selection. High quasispecies complexity represents a powerful strategy to survive under a continuously changing environment (Vignuzzi *et al.* 2006). However, one of the most ecologically successful viruses (APLV1) in Lake Limnopolar exhibits low quasispecies complexity indexes. This may be possible in the context of a highly conserved environment, as it has been reported for Lake Limnopolar during the last 5000 years (Toro *et al.* 2013).

Contrary to the low complexity and stability of APLV1 quasispecies, APLV2-APLV4 showed complexity indexes 10 times higher in the lake water. Comparing the quasispecies of APLV2-APLV3 among seasons and ecosystems, these viruses harbour some nucleotide differences in the consensus sequence in the lake water from 2007 to 2010 and many more between cyanobacterial mats and lake water samples. Importantly, the quasispecies of APLV2 and APLV3 in cyanobacterial mat displayed reduced complexity indicators that resemble those of APLV1 in the water column. These results led us to propose a hypothetical scenario where Antarctic RNA viruses, each of them infecting one or several micro-organisms, are well adapted to particular environments and harbour quasispecies with low-complexity indexes, as APLV1 in the water column or APLV2 and APLV3 in cyanobacterial mats. The seasonal transfer of APLV2 and APLV3 to the lake water by run-off radically increases their quasispecies complexity (Fig. 4).

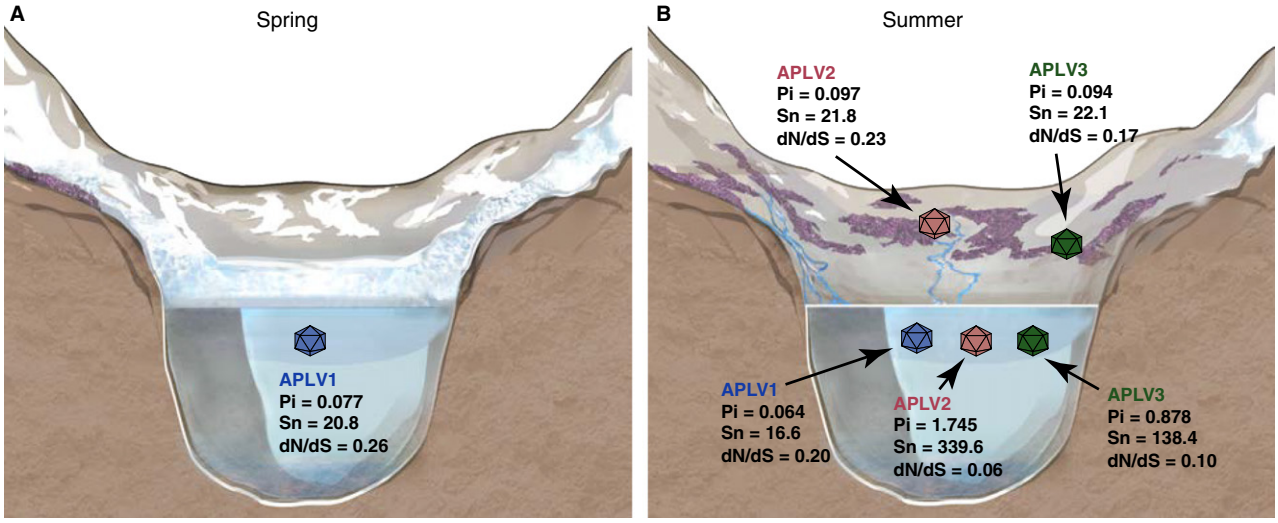

Fig. 4 Effects of ecological connectivity on the quasispecies structure of Antarctic RNA viruses. Normalized Shannon index (Sn), nucleotide diversity (π) and global *dN/dS* obtained applying the best-fit nucleotide substitution model are indicated for Antarctic picorna-like viruses (APLVs) in cyanobacterial mats (purple coloured) and the water column of Lake Limnopolar in spring 2006 (A) and summer 2010 (B). APLV1 is coloured in blue, APLV2 in red and APLV3 in green.

This intriguing shift may respond to several possibilities. First, the lake may function as a temporal collector of the diversity generated in multiple cyanobacterial mats subjected to specialization. The mixture of quasispecies from various local environments in the lake gives rise to an overall increase of complexity compared to that of a single cyanobacterial mat patch. This hypothetical scenario would not require either viral adaptation or active replication in the lake and is in agreement with the low *dN/dS* estimates in both environments. However, unlike APLV2 and APLV3, other highly abundant *Picornavirales* in the cyanobacterial mats were almost not detected in Lake Limnopolar (Table S8, Supporting information). Since all of them are subjected to the same run-off process, the chances of replication in hosts populating the lake water might explain the presence or absence of these allochthonous viruses in the lake. Consequently, a second possibility would be that viruses encounter in the lake a wider plethora of new hosts and the adaptation to these different hosts expands quasispecies complexity. However, success in the adaptation to new hosts in the lake water should lead to permanence of this virus in the lake along the year and neither APLV2 nor APLV3 was detected in the lake water after the winter. Moreover, this second scenario requires a certain degree of adaptation, and therefore, the quasispecies should be subjected to positive selection strengths, at least during the initial viral replication cycles in lake water. However, the overall *dN/dS* estimates from APLV2 and APLV3 ORFs decreased in the lake water. This reduction might be explained in the first scenario by the convergence of quasispecies that differs more in synonymous mutations than in nonsynonymous mutations as a result of being subjected to similar selection forces derived from the same environment. Finally, an intermediate scenario would be the replication of APLV2 and APLV3 in a diversity of allochthonous hosts also transferred by run-off to the lake from cyanobacterial mats. This alternative is supported by the observed seasonal and run-off-mediated increase of diatoms in the water column of the lake during summer (Toro *et al.* 2007; Rochera *et al.* 2013) and the similar diatom types detected in cyanobacterial mats by metagenomics (D. Velázquez, A. López-Bueno, D. Aguirre de Cárcer, A. de los Ríos, A. Alcamí, A. Quesada. Manuscript submitted for publication) and in the Lake Limnopolar water by electron microscopy (Kopalová & van de Vijver 2013).

This report explores, over a 4-year period, the diversity of RNA viruses in a polar lake dominated by micro-organisms. The four most abundant viruses are related to those from the order *Picornavirales* and exhibit different patterns of seasonality and geographical distribution in the catchment area. To our knowledge, this is the first assessment of viral quasispecies in aquatic ecosystems and we provide examples of the impact of ecological connectivity on basic complexity indicators of quasispecies. Further research analysing the geographical and temporal distribution of host-RNA virus pairs in the watershed will shed light into the processes governing the quasispecies structure of RNA viruses in aquatic ecosystems and their influence on microbial ecosystems.

Acknowledgements

This project was funded by the Spanish Polar Programme and the Spanish Ministry of Economy and Competitiveness (CTM2008-05134-E/ANT and CTM2009-08644-E). We thank Roche for their support to the Antarctic expedition. A.L-B. and M.A. were recipients of a 'Ramón y Cajal' and a 'Juan de la Cierva' fellowships from the Spanish Ministry of Economy and Competitiveness, respectively. We thank Esteban Domingo, Francisco Sobrino and Antonio Quesada for helpful comments and discussion. We thank the logistic support from the Maritime Technology Unit (CSIC) and Las Palmas crew (Spanish Navy). We also thank David Velázquez for sampling the lake in 2007 and Pepe Belio for digital art support.

References

Anantharaman K, Duhaime MB, Breier JA *et al.* (2014) Sulfur oxidation genes in diverse deep-sea viruses. *Science*, **344**, 757–760.

Angly FE, Felts B, Breitbart M *et al.* (2006) The marine viromes of four oceanic regions. *PLoS Biology*, **4**, e368.

Bolduc B, Shaughnessy DP, Wolf YI *et al.* (2012) Identification of novel positive-strand RNA viruses by metagenomic analysis of archaea-dominated Yellowstone hot springs. *Journal of Virology*, **86**, 5562–5573.

Bouquet J, Cheval J, Rogee S, Pavio N, Eloit M (2012) Identical consensus sequence and conserved genomic polymorphism of hepatitis E virus during controlled interspecies transmission. *Journal of Virology*, **86**, 6238–6245.

Brum JR, Schenck RO, Sullivan MB (2013) Global morphological analysis of marine viruses shows minimal regional variation and dominance of non-tailed viruses. *ISME Journal*, **7**, 1738–1751.

Brussaard CP, Noordeloos AA, Sandaa RA, Heldal M, Bratbak G (2004) Discovery of a dsRNA virus infecting the marine photosynthetic protist *Micromonas pusilla*. *Virology*, **319**, 280–291.

Bull RA, Luciani F, McElroy K *et al.* (2011) Sequential bottlenecks drive viral evolution in early acute hepatitis C virus infection. *PLoS Pathogens*, **7**, e1002243.

Camacho A (2006) Planktonic microbial assemblages and the potential effects of metazooplankton predation on the food web of lakes from the maritime Antarctica and sub-Antarctic islands. *Re/Views in Environmental Science and Bio/Technology (Online)*, **5**, 167–185.

Cantalupo PG, Calgua B, Zhao G *et al.* (2011) Raw sewage harbors diverse viral populations. *MBio*, **2**, e00180-11.

Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, **25**, 1972–1973.

Caron DA, Dam HG, Kremer P *et al.* (1995) The contribution of microorganisms to particulate carbon and nitrogen in surface waters of the Sargasso Sea near Bermuda. *Deep Sea Research Part I: Oceanographic Research Papers*, **42**, 943–972.

Culley AI, Lang AS, Suttle CA (2003) High diversity of unknown picorna-like viruses in the sea. *Nature*, **424**, 1054–1057.

Culley AI, Lang AS, Suttle CA (2006) Metagenomic analysis of coastal RNA virus communities. *Science*, **312**, 1795–1798.

Culley AI, Lang AS, Suttle CA (2007) The complete genomes of three viruses assembled from shotgun libraries of marine RNA virus communities. *Virology Journal*, **4**, 69.

Culley AI, Suttle CA, Steward GF (2010) Chapter 19: Characterization of the diversity of marine RNA viruses. In: *Manual of Aquatic Viral Ecology* (eds Wilhelm SW, Weinbauer MG, Suttle CA), pp. 193–201. ASLO.

Culley AI, Mueller JA, Belcaid M *et al.* (2014) The characterization of RNA viruses in tropical seawater using targeted PCR and metagenomics. *MBio*, **5**, e01210-14.

Cuvelier ML, Allen AE, Monier A *et al.* (2010) Targeted metagenomics and ecology of globally important uncultured eukaryotic phytoplankton. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 14679–14684.

Djikeng A, Kuzmickas R, Anderson NG, Spiro DJ (2009) Metagenomic analysis of RNA viruses in a fresh water lake. *PLoS One*, **4**, e7264.

Domingo E, Martin V, Perales C *et al.* (2006) Viruses as quasispecies: biological implications. *Current Topics in Microbiology and Immunology*, **299**, 51–82.

Domingo E, Sheldon J, Perales C (2012) Viral quasispecies evolution. *Microbiology and Molecular Biology Reviews*, **76**, 159–216.

Gregori J, Esteban JI, Cubero M *et al.* (2013) Ultra-deep pyrosequencing (UDPS) data treatment to study amplicon HCV minor variants. *PLoS One*, **8**, e83361.

Hoffmann C, Minkah N, Leipzig J *et al.* (2007) DNA bar coding and pyrosequencing to identify rare HIV drug resistance mutations. *Nucleic Acids Research*, **35**, e91.

Horiuchi K (1975) Genetic studies of RNA phages. In: *RNA Phages* (ed. Zinder ND), pp. 29–50. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Huson DH, Scornavacca C (2012) Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Systematic Biology*, **61**, 1061–1067.

Jerzak G, Bernard KA, Kramer LD, Ebel GD (2005) Genetic variation in West Nile virus from naturally infected mosquitoes and birds suggests quasispecies structure and strong purifying selection. *Journal of General Virology*, **86**, 2175–2183.

Kapoor A, Simmonds P, Lipkin WI, Zaidi S, Delwart E (2010) Use of nucleotide composition analysis to infer hosts for three novel picorna-like viruses. *Journal of Virology*, **84**, 10322–10328.

Karlsson OE, Belak S, Granberg F (2013) The effect of preprocessing by sequence-independent, single-primer amplification (SISPA) on metagenomic detection of viruses. *Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science*, **11** (Suppl 1), S227–S234.

Koonin EV (1991) The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses. *Journal of General Virology*, **72**(Pt 9), 2197–2206.

Kopalová K, van de Vijver B (2013) Structure and ecology of freshwater benthic diatom communities from Byers Peninsula, Livingston Island, South Shetland Islands. *Antarctic Science*, **25**, 239–253.

Kosakovsky Pond SL, Frost SD (2005) Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Molecular Biology and Evolution*, **22**, 1208–1222.

Labonte JM, Suttle CA (2013) Previously unknown and highly divergent ssDNA viruses populate the oceans. *ISME Journal*, **7**, 2169–2177.

Lang AS, Culley AI, Suttle CA (2004) Genome sequence and characterization of a virus (HaRNAV) related to picorna-like viruses that infects the marine toxic bloom-forming alga *Heterosigma akashiwo*. *Virology*, **320**, 206–217.

Lang AS, Rise ML, Culley AI, Steward GF (2009) RNA viruses in the sea. *FEMS Microbiology Reviews*, **33**, 295–323.

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods*, **9**, 357–359.

Laybourn-Parry J (2009) Microbiology. No place too cold. *Science*, **324**, 1521–1522.

Laybourn-Parry J, Pearce DA (2007) The biodiversity and ecology of Antarctic lakes: models for evolution. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, **362**, 2273–2289.

Li H, Handsaker B, Wysoker A *et al.* (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

López-Bueno A, Villarreal LP, Almendral JM (2006) Parvovirus variation for disease: a difference with RNA viruses? *Current Topics in Microbiology and Immunology*, **299**, 349–370.

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

Massana R (2011) Eukaryotic picoplankton in surface oceans. *Annual Review of Microbiology*, **65**, 91–110.

McElroy K, Zagordi O, Bull R, Luciani F, Beerenwinkel N (2013) Accurate single nucleotide variant detection in viral populations by combining probabilistic clustering with a statistical test of strand bias. *BMC Genomics*, **14**, 501.

Nagasaki K, Tomaru Y, Katanozaka N *et al.* (2004) Isolation and characterization of a novel single-stranded RNA virus infecting the bloom-forming diatom *Rhizosolenia setigera*. *Applied and Environmental Microbiology*, **70**, 704–711.

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution*, **3**, 418–426.

Ng TF, Marine R, Wang C *et al.* (2012) High variety of known and new RNA and DNA viruses of diverse origins in untreated sewage. *Journal of Virology*, **86**, 12161–12175.

Phan TG, Kapusinszky B, Wang C *et al.* (2011) The fecal viral flora of wild rodents. *PLoS Pathogens*, **7**, e1002218.

Rochera C, Toro M, Rico E *et al.* (2013) Structure of planktonic microbial communities along a trophic gradient in lakes of Byers Peninsula, South Shetland Islands. *Antarctic Science*, **25**, 277–287.

Rodriguez-Valera F, Martin-Cuadrado AB, Rodriguez-Brito B *et al.* (2009) Explaining microbial population genomics through phage predation. *Nature Reviews Microbiology*, **7**, 828–836.

Roossinck MJ, Schneider WL (2006) Mutant clouds and occupation of sequence space in plant RNA viruses. *Current Topics in Microbiology and Immunology*, **299**, 337–348.

Rosario K, Duffy S, Breitbart M (2009a) Diverse circovirus-like genome architectures revealed by environmental metagenomics. *Journal of General Virology*, **90**, 2418–2424.

Rosario K, Nilsson C, Lim YW, Ruan Y, Breitbart M (2009b) Metagenomic analysis of viruses in reclaimed water. *Environmental Microbiology*, **11**, 2806–2820.

Rosario K, Duffy S, Breitbart M (2012) A field guide to eukaryotic circular single-stranded DNA viruses: insights gained from metagenomics. *Archives of Virology*, **157**, 1851–1871.

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

Sakai A, Kaneko S, Honda M, Matsushita E, Kobayashi K (1999) Quasispecies of hepatitis C virus in serum and in three different parts of the liver of patients with chronic hepatitis. *Hepatology*, **30**, 556–561.

Sanz-Ramos M, Diaz-San Segundo F, Escarmis C, Domingo E, Sevilla N (2008) Hidden virulence determinants in a viral quasispecies in vivo. *Journal of Virology*, **82**, 10465–10476.

Sarker S, Patterson EI, Peters A *et al.* (2014) Mutability dynamics of an emergent single stranded DNA virus in a naive host. *PLoS One*, **9**, e85370.

Schmieder R, Edwards R (2011) Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, **27**, 863–864.

Sharon I, Alperovitch A, Rohwer F *et al.* (2009) Photosystem I gene cassettes are present in marine virus genomes. *Nature*, **461**, 258–262.

Shirai Y, Tomaru Y, Takao Y *et al.* (2008) Isolation and characterization of a single-stranded RNA virus infecting the marine planktonic diatom *Chaetoceros tenuissimus* Meunier. *Applied and Environmental Microbiology*, **74**, 4022–4027.

Steward GF, Culley AI, Mueller JA *et al.* (2013) Are we missing half of the viruses in the ocean? *ISME Journal*, **7**, 672–679.

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

Takacs CD, Priscu JC (1998) Bacterioplankton dynamics in the McMurdo Dry Valley Lakes, Antarctica: production and biomass loss over four seasons. *Microbial Ecology*, **36**, 239–250.

Takao Y, Nagasaki K, Mise K, Okuno T, Honda D (2005) Isolation and characterization of a novel single-stranded RNA virus infectious to a marine fungoid protist, *Schizochytrium* sp. (Thraustochytriaceae, Labyrinthulea). *Applied and Environmental Microbiology*, **71**, 4516–4522.

Tomaru Y, Takao Y, Suzuki H, Nagumo T, Nagasaki K (2009) Isolation and characterization of a single-stranded RNA virus infecting the bloom-forming diatom *Chaetoceros socialis*. *Applied and Environmental Microbiology*, **75**, 2375–2381.

Tomaru Y, Takao Y, Suzuki H *et al.* (2011) Isolation and characterization of a single-stranded DNA virus infecting *Chaetoceros lorenzianus* Grunow. *Applied and Environmental Microbiology*, **77**, 5285–5293.

Tomaru Y, Toyoda K, Kimura K *et al.* (2012) First evidence for the existence of pennate diatom viruses. *ISME Journal*, **6**, 1445–1448.

Toro M, Camacho A, Rochera C *et al.* (2007) Limnological characteristics of the freshwater ecosystems of Byers Peninsula, Livingston Island, in maritime Antarctica. *Polar Biology*, **30**, 635–649.

Toro M, Granados I, Pla S *et al.* (2013) Chronostratigraphy of the sedimentary record of Limnopolar Lake, Byers Peninsula, Livingston Island, Antarctica. *Antarctic Science*, **25**, 198–212.

Velázquez D, Ángeles Lezcano M, Frias A, Quesada A (2013) Ecological relationships and stoichiometry within a Maritime Antarctic watershed. *Antarctic Science*, **25**, 191–197.

Victoria JG, Kapoor A, Dupuis K, Schnurr DP, Delwart EL (2008) Rapid identification of known and new RNA viruses from animal tissues. *PLoS Pathogens*, **4**, e1000163.

Vignuzzi M, Stone JK, Arnold JJ, Cameron CE, Andino R (2006) Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature*, **439**, 344–348.

Villaescusa JA, Casamayor EO, Rochera C *et al.* (2010) A close link between bacterial community composition and environmental heterogeneity in maritime Antarctic lakes. *International Microbiology*, **13**, 67–77.

Wen K, Ortmann AC, Suttle CA (2004) Accurate estimation of viral abundance by epifluorescence microscopy. *Applied and Environmental Microbiology*, **70**, 3862–3867.

Willner D, Furlan M, Schmieder R *et al.* (2011) Metagenomic detection of phage-encoded platelet-binding factors in the human oral cavity. *Proceedings of the National Academy of Sciences of the United States of America*, **108**(Suppl 1), 4547–4553.

A.L.-B. and A.A. planned and designed the research. The experiments were performed by A.L.-B. Bioinformatic analyses of the data were performed by A.L.-B., A.R., R.P. and M.A. Article was written by A.L.-B. and A.A. All authors discussed the results and commented on the manuscript.

Data accessibility

The sequence read archive accession number for the RNA viromes reported in this study is SRP044919. Primer-clipped and quality-filtered metagenomic reads of L2006, L2007 and L2010 are also publicly accessible in MG-RAST (http://metagenomics.anl.gov/) under id numbers 4572579.3, 4572623.3 and 4572643.3 and in *Metavir* (http://metavir-meb.univ-bpclermont.fr/) under the project id numbers 4488, 4489 and 4490. APLV 1–4 genomes have been deposited at GenBank under Accession nos. KM259869–KM259872.

Contigs obtained with NEWBLER and CLC GENOMIC WORKBENCH *de novo* assemblers, alignment of the RpRd domain in fasta format, phylogenetic tree in newick format, a table with validated SNVs (CLC GENOMIC WORKBENCH approach) for all quasispecies and the dinucleotide bias input for PCA analysis are available at Dryad Digital Repository (doi: 10.5061/dryad.1dp6s).

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1 Taxonomy overview of reads contained in contigs from Lake Limnopolar viromes.

Fig. S2 Genomic organization of four APLVs.

Fig. S3 Principal component analysis (PCA) of the dinucleotide bias of APLV1-4 and a comprehensive set of viruses from the order *Picornavirales*.

Fig. S4 Genome stability of APLVs.

Fig. S5 Statistical analysis of complexity indexes.

Fig. S6 SNV calling calculated by four alternative methods.

Table S1 Details of sample collection.

Table S2 Oligonucleotides used for scaffolding and 3′ end sequencing.

Table S3 Metagenomic reads statistics and accession numbers.

Table S4 Taxonomic profile of contigs.

Table S5 BLASTP against GenBank nr of the APLVs-ORFs.

Table S6 Percentage of amino acid identity among conserved domains of APLVs.

Table S7 Complexity indexes of APLV1 and APLV2 viral quasispecies.

Table S8 RNA viral contigs from cyanobacterial mats surrounding Lake Limnopolar.

Table S9 Alternative methods for global *dN*/*dS* rates estimation.