Vol. 78: 135–145, 2017
doi: 10.3354/ame01808

AQUATIC MICROBIAL ECOLOGY
Aquat Microb Ecol

Published online February 9

Viral assemblage variation in an Arctic shelf seafloor

Tan Thi Nguyen[1,2], Espen Mikal Robertsen[3], Bjarne Landfald[2,*]

[1]Centre for Research-based Innovation on Marine Bioactives and Drug Discovery (MabCent-SFI), UiT The Arctic University of Norway, 9037 Tromsø, Norway

[2]Norwegian College of Fishery Science, Faculty of Biosciences, Fisheries and Economics, UiT The Arctic University of Norway, Breivika, 9037 Tromsø, Norway

[3]NorStruct, Department of Chemistry, Faculty of Science and Technology, UiT The Arctic University of Norway, Breivika, 9037 Tromsø, Norway

ABSTRACT: Spatial differences in microbial communities are observable even in habitats with moderate environmental variation, such as within the pelagic zone or seafloor of geographically finite regions of the oceans. Here we explore whether biogeographical variations are also manifested at this level in the structure of viral assemblages by comparing DNA viromes from the Barents Sea upper seafloor, collected at 5 geographically separated locations. Of the open reading frames, 27 to 44 % showed significant similarity to genes of viral genomes in the Refseq database. The majority of the identified open reading frames, i.e. 86 to 95 %, were affiliated with sequences of single-stranded DNA (ssDNA) viruses, but the ssDNA virus genetic material was likely strongly overrepresented due to the use of phi29 DNA polymerase for amplifying viral DNA. The majority of ssDNA virus sequences originated from the *Microviridae* family of phages and the eukaryotic Circular Rep-encoding ssDNA (CRESS-DNA) viruses. The sediment virus assemblages showed higher overall similarity to counterparts from deep-sea sediment of the Pacific Ocean than to, e.g., Arctic Ocean pelagic viromes, supporting the presence of common compositional features in sediment viral assemblages across continental-scale geographical separations. The Barents Sea viromes grouped biogeographically in accordance with the south–north environmental division of this Arctic sea by the oceanic polar front, thereby mirroring a corresponding 16S rRNA gene-based biogeographical division of the bacterial communities. However, compositional differences in the eukaryotic virus assemblages rather than the bacteriophages appeared to be the primary basis for this spatial separation.

KEY WORDS: Viruses · ssDNA virus · Barents Sea · Marine sediments · Biogeography

INTRODUCTION

Viruses and virus-like particles are the most abundant biological entities in marine ecosystems and they impact processes such as horizontal gene transfer, microbial community structuring and biogeochemical cycling (Bergh et al. 1989, Fuhrman 1999, Weinbauer 2004, Weinbauer & Rassoulzadegan 2004, Suttle 2005, Bouvier & del Giorgio 2007, Yang et al. 2010). Their genomes may have double-stranded (ds) or single-stranded (ss) DNA or RNA. Traditionally, the dsDNA group has been considered the most abundant in marine environments (Wommack & Colwell 2000, Breitbart et al. 2002, Breitbart et al. 2004, Steward & Preston 2011, Fancello et al. 2013). In recent years, high abundances of ssDNA viruses have also been reported in various aquatic environments, including marine waters (Angly et al. 2006, Tucker et al. 2011, Labonté & Suttle 2013a,b), freshwater (López-Bueno et al. 2009, Roux et al. 2012), deep seafloor sediments (Yoshida et al. 2013) and methane seep sediments (Bryson et al. 2015). However, it has become increasingly clear that the high prevalence of ssDNA viruses is, to a large extent, a consequence of amplification bias caused by the use of phi29 DNA polymerase-based multiple-

*Corresponding author: bjarne.landfald@uit.no

displacement DNA amplification (MDA) to obtain sufficient metagenomic material for high-throughput sequencing (Kim & Bae 2011). While the majority of marine dsDNA viral genomes range from 25 to 70 kilobases (kb) in size (Steward et al. 2000), circular ssDNA viruses are small with respect to both physical and genomic size, as the latter largely vary in the range from 1 to 9 kb (Day & Hendrix 2005). Subsequently, the ssDNA viruses are not easily enumerated by epifluorescent microscopy or flow cytometry viral counts because of weak fluorescence signal (Tomaru & Nagasaki 2007, Holmfeldt et al. 2012).

Temporal and geographical variations in aquatic viral assemblages are well documented (Bergh et al. 1989, Wommack & Colwell 2000, Short & Suttle 2003, Clasen et al. 2008, Danovaro et al. 2008, Yang et al. 2010, Helton et al. 2012, Parsons et al. 2012, Hurwitz et al. 2014) and these variations are, at some level, presupposed to be related to concomitant variations in the host communities, although these relationships may be complex. Uncertainty factors include, e.g., lytic versus temperate strategies of the viruses and r versus K type strategies of the hosts (Suttle 2007). Conceptual models such as 'kill the winner', implying that the fastest growing host populations in a community are the ones most susceptible to viral infections (Thingstad 2000), have been fruitful in explaining the viral contribution to maintenance of high host diversities in natural environments.

While there has been a vast increase in the available genomic information from marine microbiotas in the last decade owing to the implementation of next-generation sequencing technologies, the number of isolated and characterized virus–host systems is still limited. Due to their ecological importance, marine primary producers, including the ubiquitous *Prochlorococcus* and *Synechococcus* among the prokaryotes and diatoms and dinoflagellates among the eukaryotes, have been principal targets of such studies (Sullivan et al. 2003, Nagasaki 2008, Avrani et al. 2011, Kimura & Tomaru 2015). Among heterotrophic bacteria, phages infecting the *Bacteroidetes* species *Cellulophaga baltica* have been thoroughly characterized by Holmfeldt et al. (2007) and later infection studies of the ubiquitous, but hard-to-cultivate pelagic groups SAR11 and SAR116 have been added to the list (Kang et al. 2013, Zhao et al. 2013). To evade the limitations caused by the consistent lack of cultivability of the large majority of marine prokaryotes, single-cell genomics-based approaches have been introduced to identify and characterize individual bacterioplankton–virus systems (Roux et al. 2014b, Labonté et al. 2015). To our knowledge, no isolated virus–host system from marine sediments has been characterized.

Although a large fraction of the individual or contiguous gene sequences resulting from metagenomic sequencing projects on aquatic viromes has been classified as having unknown function and taxonomic affiliation (Breitbart et al. 2004, Angly et al. 2006, López-Bueno et al. 2009, Fancello et al. 2013, Yoshida et al. 2013), such projects have still given valuable insights into compositional and biogeographical variations. The pioneering shotgun cloning-based study by Breitbart et al. (2004) indicated a higher viral diversity in Californian coastal sediment than in the adjacent pelagic water masses. The 2 assemblage types were also phylogenetically distinct. Yoshida et al. (2013) have compared pyrosequence data from 3 deep-sea sediment viromes and concluded that they show marked similarities that distinguish them from, e.g., seawater viromes, despite spatial separations by more than 1000 km and marked host community and environmental differences.

The objective of the present study was to complement the hitherto fragmentary knowledge on diversity variations in seafloor viromes by adding data from an oceanic area that deviates both geographically and environmentally from those addressed in preceding studies in the field. Sampling at different sites, separated by up to 640 km, in the Barents Sea shelf seafloor allowed for the elucidation of viral biogeographical variations within a coherent and environmentally moderately variable shelf sediment area. In a recent study based on the same sediment material, 16S rRNA gene-based prokaryotic community variation was linked to the environmental and geographic separation of the Barents Sea by the oceanic polar front (Nguyen & Landfald 2015). Hence, the present analysis allowed us to determine whether coincident biogeographical patterns could be observed for bacterial and bacteriophage assemblages originating from the same material.

MATERIALS AND METHODS

Sampling

Sediment samples were collected by Van Veen grab during 61 h (20–23 May 2009) from 5 locations separated by up to 640 km (Fig. 1) in the western part of the Barents Sea. The 3 southern sampling stations (Stns 1–3) were south of the polar front, and hence dominated by relatively warm and saline Atlantic

water (temperature 2.5–2.7°C at seafloor at time of sampling), while the northern stations (Stns 4 and 5) were dominated by colder Arctic water (temperature 0.9–1.2°C). Furthermore, ocean depth was greatest at the southern stations (442–474 m versus 230–290 m at the northern ones), while other recorded physicochemical variables (salinity, grain size distribution, organic carbon content) showed minimal or inconsistent differences between sampling sites. Additional details of sampling, locations and sediment characteristics are given in Nguyen & Landfald (2015). Stns 1 to 5 of the present study correspond to Stns 1, 3, 4, 7 and 10 in that study.

Purification and isolation of viral DNA

Viral purification was performed with modifications of the procedure of Thurber et al. (2009). Three milliliters of sediment was suspended in 3 ml of 2% (v/w) tetrasodium pyrophosphate and incubated for 15 min in the dark at 4°C. To ensure release of the viral particles from the sediment samples, the suspension was sonicated on ice for 3 × 45 s with a Branson 3210 sonicator (Triad Scientific), and centrifuged at 1000 × *g* for 5 min. Thereafter, the sediment pellets were twice re-suspended in 3 ml of 2% (v/w) NaCl and centrifuged at 2500 × *g* for 5 min. The supernatants were pooled and successively passed slowly through 0.45 µm and 0.2 µm Whatman filters to remove residual cells. The filtrate was concentrated using a 30 kDa NMWCO centrifugal ultrafiltration device (Ultracel PL-30 membrane, Millipore). The concentrate was treated with DNase I (2.5 U ml^{-1}) and incubated for 1 h at 37°C to remove free DNA. To check for purity of the viral preparation, each sample was stained with SYBRGold (Invitrogen) and inspected under an epifluorescence microscope (Leica DM6000 B). If no contamination by prokaryotic cells was revealed, the viral nucleic acids were extracted using the formamide and phenol/chloroform/isoamyl alcohol procedure (Thurber et al. 2009). After the extraction, the absence of detectable quantities of bacterial DNA was confirmed by subjecting the DNA preparation to PCR amplification with universal bacterial 16S rRNA gene primers 27F (5′-AGA GTT TGA TCM TGG CTC AG-3′) and 1492R (5′-TAC CTT GTT ACG ACT T-3′). The PCR master mix contained 2.5 µl of 10X PCR buffer (Invitrogen), 2.0 µl of 0.2 mM dNTPs (Invitrogen), 2.5 µl of 0.5 µM of each primer (Eurofins MWG), 1.25 U of Taq polymerase (Invitrogen), 10 ng of genomic DNA template, and Milli-Q water to a total volume of 25 µl. The thermocycler (Applied Biosystems) conditions were: an initial denaturation step at 95°C for 5 min; 30 cycles at 95°C for 30 s, 55°C for 30 s, and 72°C for 1 min; and a final extension at 72°C for 5 min. Only preparations where no bacterial cells were observed and no bacterial PCR products were generated from the DNA extracts were used for subsequent viral amplification.

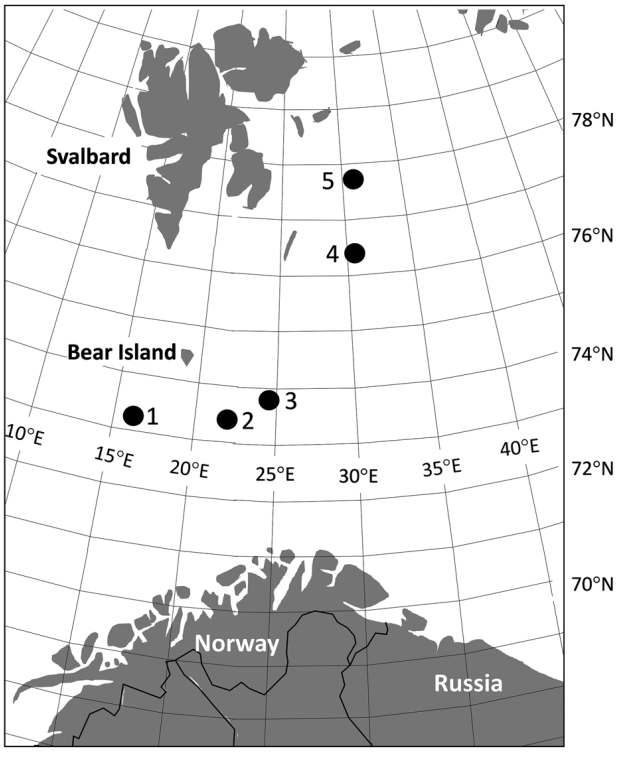

Fig. 1. The western Barents Sea with geographical positions of the 5 sampling stations

Amplification of viral DNA and sequencing

Viral genomic DNA (10 ng) was amplified using the GenomiPhi V2 DNA Amplification Kit (GE Healthcare Life Sciences) according to the manufacturer's instructions. After amplification, amplicon DNA was purified on silica columns (Qiagen Dnase Kit), precipitated with ethanol and resuspended in pure water. The concentration and quality of amplified DNA were determined using a NanoDrop ND-1000 spectrophotometer. The amplifications were done in duplicate for each sample to reduce bias. The samples were subjected to shotgun pyrosequencing (GS FLX Titanium; 454 Life Sciences, installed at The Norwegian High-Throughput Sequencing Centre, University of Oslo, Norway; www.sequencing.uio.no).

Sequence assembly

Flowgram sequence data in sff format from the GS FLX Titanium were assembled using Newbler version 2.5.3 for all 5 stations. Several runs with varying parameters were tested to produce optimal assemblies based on contig length distribution (N50) and total contig yield (bp). Default settings for *de novo* assembly and 'autotrimming' were considered the most successful. A minimum contig size threshold of 500 bp was applied to discard the smallest contigs.

Taxonomic classifications

The metagenomic sequence and contig data were analyzed using the Metavir (Roux et al. 2011) and Metavir2 (Roux et al. 2014a) pipeline (http://metavir-meb.univ-bpclermont.fr/) with taxonomic affiliations based on BLASTx queries in the NCBI Refseq complete viral genomes protein sequence database (release 2016-01-19). BLAST hits with E-values $<10^{-3}$ were considered significant. Sequences affiliated with individual *Circoviridae*, *Geminiviridae* and *Nanoviridae* genomes or metagenomic genomes categorized as Circular Rep-encoding ssDNA (CRESS-DNA) viruses were classified as CRESS-DNA virus sequences in accordance with Rosario et al. (2015). Taxonomic assignments were also generated via MG-RAST (Meta Genome Rapid Annotation using Subsystem Technology, v3.1) (Meyer et al. 2008) based on BLASTx searches against the M5NR non-redundant database. An E-value cutoff of 10^{-3}, a minimum identity cutoff of 60% and minimum alignment length of 15 bp were used as parameters for this analysis.

Phylogenetic analyses of marker genes

The protein sequences of the major capsid protein F (Vp1) of the *Microviridae* family (Desnues et al. 2008) and the replication-associated protein (Rep) of the *Circoviridae*/*Nanoviridae*/*Geminiviridae* families (Rosario et al. 2009a) were used to infer phylogenetic trees. Representative sequences covering the whole or major parts of the complete gene products from the Barents Sea sediment viromes were aligned with homologous sequences obtained from translated genomic and metagenomic sequences of public data and the Metavir server, using the ClustalW algorithm in MEGA 6 program package (Tamura et al. 2013). Phylogenetic trees were constructed using the maximum likelihood method, with the Jones-Taylor-Thornton substitution model, gamma distribution among sites, partial deletion of missing data and bootstrapping with 100 resamplings, as implemented in the MEGA package.

Comparative metagenomics of viral communities

The viral metagenomics data from our study were compared with public virome projects from other environments, available in the Metavir web server (Roux et al. 2011). The imported data only included studies in which multiple displacement amplification (MDA) had been used to increase DNA quantities and comprised data from deep-sea surface sediment (Yoshida et al. 2013), seawater (Angly et al. 2006, Cassman et al. 2012) and freshwater (Roux et al. 2012, Fancello et al. 2013). In the Metavir workflow, the similarity of the virus metagenomes were compared and scored as tBLASTx sequence similarity scores. Only viromes that contained more than 50 000 sequences with length exceeding 100 bp on average were used for the comparison. Similarity scores for 2 by 2 viromes were computed as the sum of the best BLAST hit scores of a sequence component in one virome community against its counterpart in the other virome. The score sums for the virome pairs were stored in a matrix format and used to hierarchically cluster the communities in the Metavir web server.

Separate hierarchical clustering trees for the *Microviridae* and the eukaryotic subgroups of virus sequences of the Barents Sea sediment data were inferred using the *pvclust* package with default parameters (Suzuki & Shimodaira 2006) in R software (www.r-project.org). Bootstrap values (1000 resamplings) were calculated.

Bacterial community comparisons

The total DNA was extracted using the PowerSoil™ DNA Isolation kit (Mo Bio Labs) according to the manufacturer's instructions. The bacterial 16S rRNA genes were amplified use the primers 27F (5′-AGA GTT TGA TCM TGG CTC AG-3′) and 338R (5′-TGC TGC CTC CCG TAG GAG T-3′). Subsequently, the amplicons were sequenced by multiplex pyrosequencing on a 454/Roche GS-FLX Titanium instrument, and the sequences were binned into operational taxonomic units (OTUs) at 97% identity level, as described in Nguyen & Landfald (2015).

Sequence-based hierarchical clustering of the bacterial communities was done by average linkage in the R package *pvclust*.

Virus/bacteria community relationships

Correlations between variations in bacterial communities and viral assemblages (complete or subsections) were quantified as Spearman rank correlation coefficients calculated from Bray-Curtis distance matrices. Bacterial distances were based on OTU tables, while viral distances were based on tables of significant BLAST hits (E-value < 10^{-3}). The statistical significance of the correlations was tested using the Mantel test with 10^4 Monte Carlo permutations in the *vegan* package in R.

Virome accession numbers

Viral metagenomes from the 5 sampling stations are accessible in the Metavir server under the name 'Barents Sea sediment'. In the MG-RAST annotation server, they are available under accession numbers 4564380.3, 4564381.3, 4564382.3, 4564383.3 and 4565015.3. The 14 completed circular genomes are deposited in GenBank under accession numbers KX828610 to KX829623.

RESULTS AND DISCUSSION

The Barents Sea sediment virome

The shotgun pyrosequencing produced a total of 674 115 sequence reads with an average length of 507 bp from the 5 Barents Sea sediment samples (Table 1). Between 81 and 87% of reads in the different samples were assembled into 17 805 contiguous sequences (minimum 500 bp) by the Newbler assembler. The median contig length varied in the range 1131 to 1228 bp, the longest being 5745 bp. A total of 39 756 complete or partial open reading frames (ORFs) were identified in the contigs and 37 to 51% of the predicted ORFs in the separate samples showed significant homology to sequences present in the NCBI Refseq database. Fourteen contigs, ranging from 819 to 4217 bp in size, were categorized as circular genomes by employing the Metavir2 pipeline, but only 4 of them produced significant hits (E-value < 10^{-3}) to established viral genomes. The largest identified genome (4217 bp) was similar to that of a complete marine *Gokushovirinae* from the Gulf of Mexico (KC131021; Labonté & Suttle 2013a), while 2 genomes of size 2163 and 1938 bp, respectively, showed highest similarity to the eukaryotic circular Rep-encoding ssDNA (CRESS-DNA) viruses Avon-Heathcote Estuary associated circular virus 5 and 6 (KM874301 and KM874301) (Dayaram et al. 2015). The former was categorized as circo-like by encoding both the conserved Rep and a putative capsid (Cap) protein (Rosario et al. 2015), while the latter encoded just the Rep protein, as did the fourth, smallest genome (1265 bp), which was similar to *Circoviridae* 3 LDMD-2013 (KF133810; McDaniel et al. 2014).

Table 1. Outcome of the virome sequence analyses at the individual sampling sites

Stn	No. of reads	Average length (bp)	Assembled reads (% of reads)	No. of contigs (>500 bp)	Viral hits[a] (% of reads)
1	118 313	497	81.3	3642	40.5
2	158 119	511	80.1	5110	27.3
3	154 403	516	78.5	4526	40.3
4	111 206	502	79.9	2308	32.2
5	132 074	509	86.6	2219	43.8

[a]Significant similarity (E-value < 10^{-3}) to the NCBI Refseq database

Analyses of the individual sequences identified 27 to 44% as having significant similarity (BLASTx, E-values < 10^{-3}) to genes within the complete viral genomes of the Refseq database (Table 1). Hence, both the contiguous and individual sequence data confirmed the consistent pattern from preceding environmental virome studies that the majority of sequences remain unassigned. However, the proportions of virus-affiliated sequences compared favorably with preceding metagenomic studies employing the Refseq database, as 1–35% viral genes have been reported at <10^{-3} significance level (Breitbart et al. 2004, Edwards & Rohwer 2005, Angly et al. 2006, López-Bueno et al. 2009, Roux et al. 2012, Fancello et al. 2013). Successful enrichment of viral metagenomic material was also indicated by low proportions of prokaryotic (6.7%) and eukaryotic (0.6%) sequences by search in the M5NR database of MG-RAST (Meyer et al. 2008). The rather straightforward method employed in the present study for isolation of virus particles from sediment therefore appeared efficient for enrichment of viral genomic material.

Further phylogenetic classification showed 86.1 to 94.8% of reads affiliated to viruses from the 5 sampling stations to be confidently associated with

ssDNA virus genes of the Refseq database (E-value $< 10^{-3}$), while reads of dsDNA virus origin constituted 3.0 to 11.6% (Table 2). In accordance with the pattern for the ssDNA viral groups (see below), both a bacteriophage taxon (*Caudovirales*) and a eukaryote-infecting taxon (*Phycodnaviridae*) were represented among the most predominant dsDNA groups. Several previous metagenomic studies have indicated a strong presence of ssDNA viruses in marine environments (Desnues et al. 2008, Kim & Bae 2011, Rosario & Breitbart 2011, Tucker et al. 2011, Yoshida et al. 2013). However, a common feature of both the present and the pervious studies is the use of MDA with phi29 polymerase for amplification of viral metagenomic DNA. MDA biases the genome type distribution towards ssDNA viruses and, in particular, towards small circular ssDNA viruses. For seawater samples, this has been demonstrated by comparing viral DNA amplified by the linker amplified shotgun library (LASL) method of Breitbart et al. (2002) with the same samples amplified by MDA (Kim & Bae 2011), and McDaniel et al. (2014) found large differences in abundances of ssDNA virus genes between MDA and unamplified DNA preparations from seawater. Chromatographic separation of viral groups prior to LASL amplification, as devised by Andrews-Pfannkoch et al. (2010), represents an improvement, but it still remains unclarified to what extent the diversities of the resulting metagenomic data are truly representative of the source material. In our study, the proportion of dsDNA viruses may have been lowered also by the filtration step used for removal of cellular and abiotic particles. Electron micrographs show some tailed or filamentous environmental virions to have physical dimensions that limit their passage through the 0.2 µm pore size filters used in the present study (Bergh et al. 1989, Zhao et al. 2013), while circular ssDNA viruses are less likely to be hindered due to their overall smaller size. To secure the inclusion of large-size virions in the preparations, some recent virome studies have preferred the use of 0.45 µm pore size filters or no filtration step at all (López-Bueno et al. 2009, Fancello et al. 2013, Martínez Martínez et al. 2014). In contrast, environmental viral diversity studies have also demonstrated variable proportions of ssDNA virus genes and even dominance of dsDNA virus genes in 0.2 µm filtrates amplified by MDA (Rosario et al. 2009b, Cassman et al. 2012, McDaniel et al. 2014), indicating a relationship between abundances of ssDNA viruses in the source material and ssDNA virus genes in the amplified metagenome. Hence, the apparent predominance by ssDNA viral material in our amplified virome DNA likely reflects a lower but substantial proportion of this virus group in the source material.

The International Committee on Taxonomy of Viruses has classified the ssDNA viruses into 11 families and one genus not assigned to a family (Krupovic & Forterre 2015, Krupovic et al. 2016). The BLASTx search affiliated the ORFs of the Barents Sea sediment material with 6 of these families, i.e.

Table 2. Phylogenetic affiliations of viral sequences from the 5 sampling stations

Group	Taxonomic composition	Stn 1 (%) (n = 47948)	Stn 2 (%) (n = 43193)	Stn 3 (%) (n = 62148)	Stn 4 (%) (n = 36110)	Stn 5 (%) (n = 57784)
ssDNA	*Microviridae*	36.0	15.7	43.9	7.9	12.9
	CRESS-DNA virus	57.8	69.7	45.7	81.2	81.4
	Parvoviridae	0.1	0.1	0.1	0.0	0.0
	Inoviridae	0.0	0.1	0.1	0.0	0.0
	Unclassified ssDNA viruses	0.4	0.5	0.5	0.5	0.5
dsDNA	*Phycodnaviridae*	2.4	6.3	1.9	2.1	1.1
	Caudovirales	0.5	3.4	4.9	4.4	1.2
	Mimiviridae	0.2	0.5	0.2	0.3	0.2
	Poxviridae	0.2	0.5	0.2	0.2	0.1
	Herpesvirales	0.1	0.2	0.2	0.1	0.1
	Iridoviridae	0.1	0.1	0.1	0.0	0.0
	Unclassified dsDNA	0.1	0.6	0.7	0.8	0.3
Satellites		1.7	2.1	1.1	2.1	2.0
ssRNA viruses		0.1	0.0	0.1	0.0	0.0
Unclassified viruses		0.3	0.2	0.3	0.4	0.2
Chloroplast 16S rRNA genes (%)[a]		0.1	0.0	0.2	2.5	1.8

[a]Nguyen & Landfald (2015); CRESS-DNA, Circular Rep-encoding ssDNA

Microviridae, *Circoviridae*, *Nanoviridae*, *Geminiviridae*, *Parvoviridae* and *Inoviridae*, among which the *Microviridae* phage family constituted 8–44% whereas a larger fraction, 46–81%, was assigned to the 3 eukaryotic families of circular Rep-encoding ssDNA viruses (*Circoviridae*, *Nanoviridae*, *Geminiviridae*), collectively named CRESS-DNA viruses (Rosario et al. 2015). More than half of the CRESS-DNA virus sequences were affiliated with the *Circoviridae* family by showing highest similarity to genes in complete genomes carrying both Rep and putative capsid encoding genes.

There are still rather few described examples of ssDNA viruses infecting marine microorganisms, the first one being a novel virus that infects bloom-forming diatoms (Nagasaki et al. 2005). More recently, putative *Microviridae* infecting marine *Bacteroidetes* phylotypes have been reported (Holmfeldt et al. 2012). Identification of the virus–host associations by *in silico* analysis generally requires complete or near-complete genomic information of both virus and host (Roux et al. 2015). This is a challenging task by metagenomic approaches even in rather well-characterized microbiota such as the human gut (Pérez-Brocal et al. 2013), and even more so in highly diverse habitat types such as marine environments, which are dominated by hitherto uncultivated and genomically uncharacterized phylogenetic groups. However, Mizuno et al. (2013) were able to assign putative hosts to more than 500 viral contigs in a seawater fosmid library by combining several *in silico* analytical approaches. Potential bacterial and archaeal hosts of viral assemblages in hydrothermal vent environments have been identified by clustered regularly interspaced short palindromic repeat (CRISPR) analysis (Anderson et al. 2011), while Cassman et al. (2012) were unable to identify virus–host pairs by CRISPR analysis in Pacific seawater material. Recently, single-cell genome amplification has been demonstrated as a viable new approach to identify virus–bacteria associations even among the uncultivated majority of bacteria in a pelagic marine system (Roux et al. 2014b).

Marker gene phylogenies

Phylogenetic relationships of the major capsid protein F (Vp1) as marker gene for the *Microviridae* and the replication-associated protein (Rep) as marker gene for the CRESS-DNA virus were inferred by comparing sequences covering conserved regions with corresponding sequences from relevant metagenome and genome studies (Fig. 2). *Gokushovirinae* originating from intracellular parasitic bacteria (*Chlamydiae, Bdellovibrio* and *Spiroplasma* spp.) (Brentlinger et al. 2002) and *Microvirus* from enterobacterial species (including φX174) were the sources of the most closely related Vp1 genes of established genomic origin. The actual host groups of the Vp1-carrying phages within the Barents Sea sediment communities were not established. A previous study of bacterial community structure in the same material identified *Gammaproteobacteria* and *Deltaproteobacteria* as predominant groups by jointly constituting nearly two-thirds of the 16S rRNA gene pool (Nguyen & Landfald 2015). Outside the marine realm, the *Microviridae* have been shown to infect a diverse range of hosts, including *Proteobacteria*, *Spiroplasma* and *Chlamydiales* (Read et al. 2000, Brentlinger et al. 2002, Garner et al. 2004), and a recent, comprehensive search for viral genomes in prokaryotic genomic data sets (Roux et al. 2015) also ties the *Microviridae* to *Alpha-*, *Gamma-* and *Deltaproteobacteria* as important host classes. It therefore appears likely that the mentioned classes of *Proteobacteria* constitute substantial parts of the *Microviridae* host range.

Different Barents Sea Rep-encoding sequences showed affiliation to both animal-infecting (*Circoviridae*) and plant-infecting (*Nanoviridae*, ssDNA Alphasatellites, *Geminiviridae*) genomic virus sequences, and metagenomic sequences from seawater, freshwater and deep-sea sediments were also distributed throughout the Rep gene diversity. No distinct grouping of the picked Vp1 or Rep metagenomic sequences in accordance with the south–north climatic division of the Barents Sea was observed, indicating that important groups of host organisms were abundant throughout the sampling range.

Community comparisons

A hierarchical clustering tree, including both the 5 Barents Sea sampling sites and additional aquatic and sediment viromes (Angly et al. 2006, Cassman et al. 2012, Roux et al. 2012, Fancello et al. 2013, Yoshida et al. 2013), was inferred from BLAST-based comparisons (Fig. 3) by tools integrated in the Metavir server (Roux et al. 2011). The 3 Pacific Ocean marine sediment viromes (Yoshida et al. 2013) clustered with the Barents Sea samples and the environmentally most comparable location, i.e. pelagic sediment (1181 m water depth) off the Japanese coast, also showed the highest assemblage similarity. Even the

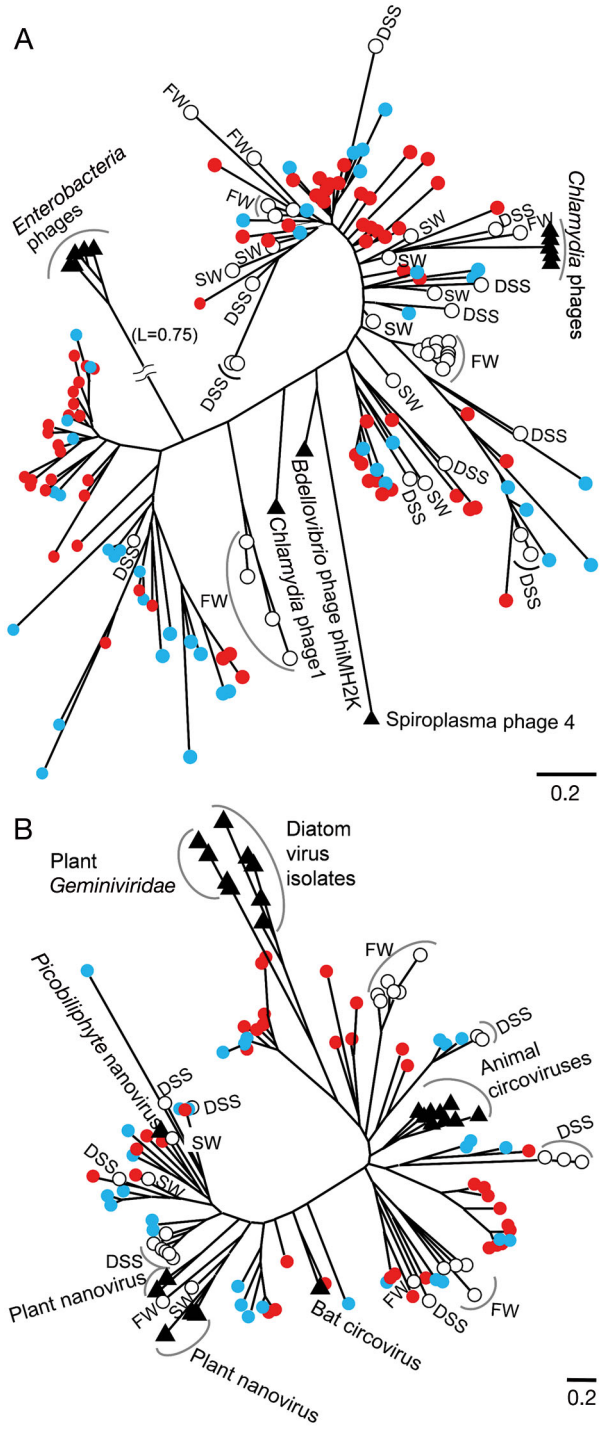

Fig. 2. Unrooted maximum likelihood phylogenetic trees computed from (A) the major capsid protein Vp1 of *Microviridae* and (B) the replication-associated protein (Rep) of Circular Rep-encoding ssDNA (CRESS-DNA) viruses. Marker coding: filled circles, sequences of the Barents Sea virome from southern stations (Stns 1–3, red) and northern stations (Stns 4 and 5, blue); open circles, sequences from other metagenomic projects, including freshwater (FW; Roux et al. 2012), seawater (SW; Labonté & Suttle 2013a, Labonté & Suttle 2013b) and deep-sea sediment (DSS; Yoshida et al. 2013); black triangles, sequences from completed genomes. L = 0.75: true branch length of *Enterobacteria* phages clade

viromes from 2 extremely deep Pacific trenches were more similar to the Barents Sea material than any seawater or freshwater sample, including the ones collected from Arctic regions. These common phylogenetic features in seafloor sediment viromes across large geographical separations and extensive variations in water depth likely reflect corresponding similarities in the prokaryotic and eukaryotic host communities, but the specific basis for the observed commonness remains vague, as community data are lacking for the microorganisms.

In the present study, some level of coincidence in bacterial and viral biogeography patterns among the Barents Sea sediment sampling stations was indicated by the hierarchical clustering patterns. A previous comprehensive study including 10 sampling stations established significant differences among bacterial communities in accordance with the main environmental division of this ocean area into a southern and a northern region by the oceanic polar front (Nguyen & Landfald 2015). Here we reconfirmed this bacterial pattern in a study comprising just the 5 stations included in the virome analysis. The 2 northernmost stations (Stns 4 and 5) constituted a distinct cluster based on 16S rRNA gene sequence data (Fig. 4A). BLAST-based similarity scores indicated the same partition of the viral assemblages if the analysis was based on the complete virome sequence data (Fig. 3). However, if restricting the analysis to the dominant group of bacteria-associated virus in the assemblages, i.e. the *Microviridae* family of phages, the clustering pattern of the viromes did not reflect that exerted by the bacterial communities (Fig. 4B). In contrast, Stns 4 and 5 showed a distinct grouping of the eukaryotic CRESS-DNA virus assemblages (Fig. 4C), suggesting that these taxa were inclined to stronger host variations along the south–north axis than the bacteria-infecting viruses. The absence of a relationship between bacteria and bacteriophage community variations was also confirmed by a Mantel test of correlation between the respective community distance matrices ($p = 0.29$). One factor that may have contributed to this lack of coherence is the aforementioned expected bias towards ssDNA virus in MDA-amplified viromes. This would lead to an overestimation of the importance of the *Microviridae* as infective agents in the bacterial communities. Furthermore, the bacterial community profiling was based on ribosomal RNA gene frequency comparisons. This implies that the phylogenetic distribution of actively proliferating bacterial phylotypes, which expectedly are the more influential ones on phage assemblage

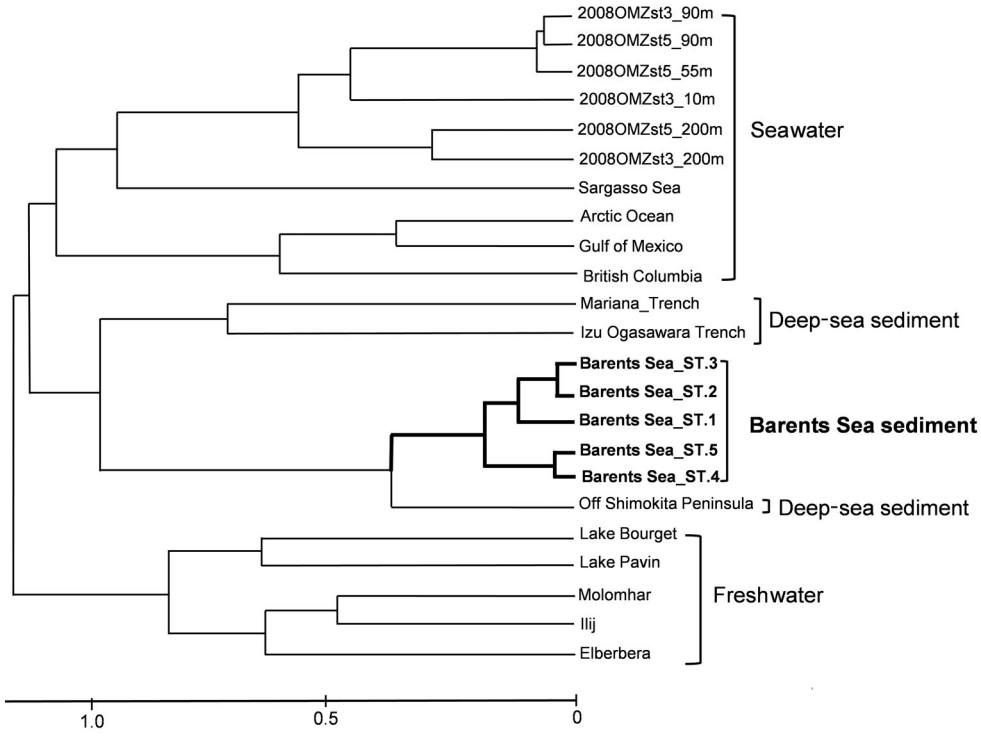

Fig. 3. tBLASTx similarity based comparison of environmental virome assemblages. All included sequence data originate from multiple-displacement amplified viral DNA. Sources: seawater (Angly et al. 2006, Cassman et al. 2012), deep-sea surface sediment (Yoshida et al. 2013) and freshwater (Roux et al. 2012, Fancello et al. 2013)

structuring, may have deviated from the observed total 16S rRNA gene-based biogeographical variation. The elevated relative abundance of eukaryotic virus types in the north (Table 2) coincided with higher levels of chlorophyll *a* and chloroplast 16S rRNA gene frequencies at the same locations, as reported elsewhere (Nguyen & Landfald 2015). Hence, increased seafloor deposition of fresh algal material in the northern region, due to the recent ice margin spring bloom, may have contributed to the observed differences among the eukaryotic virus assemblages.

In conclusion, the present study confirmed that virome biogeographical variation within a coherent and environmentally moderately variable shelf seafloor area was small as compared with differences among more distinctly separated ecosystems, such as geographically distant lakes or oceans. It furthermore indicated some degree of commonness in composition of seafloor viral assemblages at a global scale. The observed Barents Sea virome variation was principally associated with changes in relative abundance and composition of eukaryotic ssDNA virus genes, while the established bacterial community variation within the same ocean range was not mirrored in the bacteriophage metagenomic material. Our work underlines the need to complement this kind of community-wide approach with identification and characterization of representative virus–microorganism associations to obtain comprehensive pictures of the interactions between the groups in their natural environment.

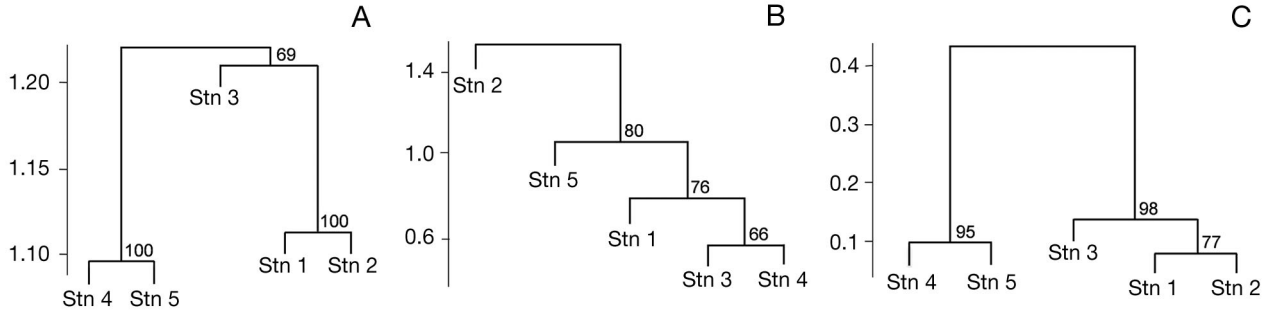

Fig. 4. Hierachical clustering of (A) bacterial communities, (B) *Microviridae* assemblages and (C) Circular Rep-encoding ssDNA (CRESS-DNA) virus assemblages of the 5 Barents Sea sampling sites. Bacterial 16S rRNA gene sequence data were obtained from Nguyen & Landfald (2015). Multiscale bootstrap values (1000 re-samplings) are indicated at the nodes

Acknowledgements. This work was supported by grants from the UiT The Arctic University of Norway, the Centre for Marine Bioactives and Drug Discovery (MABCENT-SFI) and the program Marine Biotechnology in Northern Norway (MABIT). We thank the crew of the FF Helmer Hansen for their technical and logistic assistance during the sampling cruise.

LITERATURE CITED

Anderson RE, Brazelton WJ, Baross JA (2011) Is the genetic landscape of the deep subsurface biosphere affected by viruses? Front Microbiol 2:219

Andrews-Pfannkoch C, Fadrosh DW, Thorpe J, Williamson SJ (2010) Hydroxyapatite-mediated separation of double-stranded DNA, single-stranded DNA, and RNA genomes from natural viral assemblages. Appl Environ Microbiol 76:5039–5045

Angly FE, Felts B, Breitbart M, Salamon P and others (2006) The marine viromes of four oceanic regions. PLoS Biol 4: e368

Avrani S, Wurtzel O, Sharon I, Sorek R, Lindell D (2011) Genomic island variability facilitates *Prochlorococcus*–virus coexistence. Nature 474:604–608

Bergh O, Børsheim KY, Bratbak G, Heldal M (1989) High abundance of viruses found in aquatic environments. Nature 340:467–468

Bouvier T, del Giorgio PA (2007) Key role of selective viral-induced mortality in determining marine bacterial community composition. Environ Microbiol 9:287–297

Breitbart M, Salamon P, Andresen B, Mahaffy JM and others (2002) Genomic analysis of uncultured marine viral communities. Proc Natl Acad Sci USA 99:14250–14255

Breitbart M, Felts B, Kelley S, Mahaffy JM, Nulton J, Salamon P, Rohwer F (2004) Diversity and population structure of a near-shore marine-sediment viral community. Proc R Soc B 271:565–574

Brentlinger KL, Hafenstein S, Novak CR, Fane BA, Borgon R, McKenna R, Agbandje-McKenna M (2002) Microviridae, a family divided: isolation, characterization, and genome sequence of phi MH2K, a bacteriophage of the obligate intracellular parasitic bacterium *Bdellovibrio bacteriovorus.* J Bacteriol 184:1089–1094

Bryson SJ, Thurber AR, Correa AM, Orphan VJ, Vega Thurber R (2015) A novel sister clade to the enterobacteria microviruses (family Microviridae) identified in methane seep sediments. Environ Microbiol 17:3708–3721

Cassman N, Prieto-Davo A, Walsh K, Silva GG and others (2012) Oxygen minimum zones harbour novel viral communities with low diversity. Environ Microbiol 14: 3043–3065

Clasen JL, Bigden SM, Payet JP, Suttle CA (2008) Evidence that viral abundance across oceans and lakes is driven by different biological factors. Freshw Biol 53: 1090–1100

Danovaro R, Corinaldesi C, Filippini M, Fischer UR and others (2008) Viriobenthos in freshwater and marine sediments: a review. Freshw Biol 53:1186–1213

Day LA, Hendrix RW (2005) The single stranded DNA viruses. In: Fauquet C, Mayo M, Maniloff J, Desselberger U, Ball L (eds) Virus taxonomy. Eighth Report of the International Committee on Taxonomy of Viruses. Academic Press, San Diego, CA, p 277–369

Dayaram A, Goldstien S, Argüello-Astorga GR, Zawar-Reza P, Gomez C, Harding JS, Varsani A (2015) Diverse small circular DNA viruses circulating amongst estuarine molluscs. Infect Genet Evol 31:284–295

Desnues C, Rodriguez-Brito B, Rayhawk S, Kelley S and others (2008) Biodiversity and biogeography of phages in modern stromatolites and thrombolites. Nature 452: 340–343

Edwards RA, Rohwer F (2005) Viral metagenomics. Nat Rev Microbiol 3:504–510

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

Fuhrman JA (1999) Marine viruses and their biogeochemical and ecological effects. Nature 399:541–548

Garner SA, Everson JS, Lambden PR, Fane BA, Clarke IN (2004) Isolation, molecular characterisation and genome sequence of a bacteriophage (Chp3) from *Chlamydophila pecorum.* Virus Genes 28:207–214

Helton RR, Wang K, Kan J, Powell DH, Wommack KE (2012) Interannual dynamics of viriobenthos abundance and morphological diversity in Chesapeake Bay sediments. FEMS Microbiol Ecol 79:474–486

Holmfeldt K, Middelboe M, Nybroe O, Riemann L (2007) Large variabilities in host strain susceptibility and phage host range govern interactions between lytic marine phages and their *Flavobacterium* hosts. Appl Environ Microbiol 73:6730–6739

Holmfeldt K, Odic D, Sullivan MB, Middelboe M, Riemann L (2012) Cultivated single-stranded DNA phages that infect marine Bacteroidetes prove difficult to detect with DNA-binding stains. Appl Environ Microbiol 78:892–894

Hurwitz BL, Westveld AH, Brum JR, Sullivan MB (2014) Modeling ecological drivers in marine viral communities using comparative metagenomics and network analyses. Proc Natl Acad Sci USA 111:10714–10719

Kang I, Oh HM, Kang D, Cho JC (2013) Genome of a SAR116 bacteriophage shows the prevalence of this phage type in the oceans. Proc Natl Acad Sci USA 110: 12343–12348

Kim KH, Bae JW (2011) Amplification methods bias metagenomic libraries of uncultured single-stranded and double-stranded DNA viruses. Appl Environ Microbiol 77:7663–7668

Kimura K, Tomaru Y (2015) Discovery of two novel viruses expands the diversity of single-stranded DNA and single-stranded RNA viruses infecting a cosmopolitan marine diatom. Appl Environ Microbiol 81:1120–1131

Krupovic M, Forterre P (2015) Single-stranded DNA viruses employ a variety of mechanisms for integration into host genomes. Ann NY Acad Sci 1341:41–53

Krupovic M, Ghabrial SA, Jiang D, Varsani A (2016) *Genomoviridae*: a new family of widespread single-stranded DNA viruses. Arch Virol 161:2633–2643

Labonté JM, Suttle CA (2013a) Metagenomic and whole-genome analysis reveals new lineages of gokushoviruses and biogeographic separation in the sea. Front Microbiol 4:404

Labonté JM, Suttle CA (2013b) Previously unknown and highly divergent ssDNA viruses populate the oceans. ISME J 7:2169–2177

Labonté JM, Swan BK, Poulos B, Luo H and others (2015) Single-cell genomics-based analysis of virus–host interactions in marine surface bacterioplankton. ISME J 9: 2386–2399

López-Bueno A, Tamames J, Velazquez D, Moya A, Quesada A, Alcami A (2009) High diversity of the viral community from an Antarctic lake. Science 326:858–861
Martínez Martínez J, Swan BK, Wilson WH (2014) Marine viruses, a genetic reservoir revealed by targeted viromics. ISME J 8:1079–1088
McDaniel LD, Rosario K, Breitbart M, Paul JH (2014) Comparative metagenomics: natural populations of induced prophages demonstrate highly unique, lower diversity viral sequences. Environ Microbiol 16:570–585
Meyer F, Paarmann D, D'Souza M, Olson R and others (2008) The metagenomics RAST server — a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics 9:386
Mizuno CM, Rodriguez-Valera F, Kimes NE, Ghai R (2013) Expanding the marine virosphere using metagenomics. PLoS Genet 9: e1003987
Nagasaki K (2008) Dinoflagellates, diatoms, and their viruses. J Microbiol 46:235–243
Nagasaki K, Tomaru Y, Takao Y, Nishida K, Shirai Y, Suzuki H, Nagumo T (2005) Previously unknown virus infects marine diatom. Appl Environ Microbiol 71:3528–3535
Nguyen TT, Landfald B (2015) Polar front associated variation in prokaryotic community structure in Arctic shelf seafloor. Front Microbiol 6:17
Parsons RJ, Breitbart M, Lomas MW, Carlson CA (2012) Ocean time-series reveals recurring seasonal patterns of virioplankton dynamics in the northwestern Sargasso Sea. ISME J 6:273–284
Pérez-Brocal V, Garcia-Lopez R, Vazquez-Castellanos JF, Nos P, Beltran B, Latorre A, Moya A (2013) Study of the viral and microbial communities associated with Crohn's disease: a metagenomic approach. Clin Transl Gastroenterol 4:e36
Read TD, Brunham RC, Shen C, Gill SR and others (2000) Genome sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39. Nucleic Acids Res 28: 1397–1406
Rosario K, Breitbart M (2011) Exploring the viral world through metagenomics. Curr Opin Virol 1:289–297
Rosario K, Duffy S, Breitbart M (2009a) Diverse circovirus-like genome architectures revealed by environmental metagenomics. J Gen Virol 90:2418–2424
Rosario K, Nilsson C, Lim YW, Ruan Y, Breitbart M (2009b) Metagenomic analysis of viruses in reclaimed water. Environ Microbiol 11:2806–2820
Rosario K, Schenck RO, Harbeitner RC, Lawler SN, Breitbart M (2015) Novel circular single-stranded DNA viruses identified in marine invertebrates reveal high sequence diversity and consistent predicted intrinsic disorder patterns within putative structural proteins. Front Microbiol 6:696
Roux S, Faubladier M, Mahul A, Paulhe N, Bernard A, Debroas D, Enault F (2011) Metavir: a web server dedicated to virome analysis. Bioinformatics 27:3074–3075
Roux S, Enault F, Robin A, Ravet V and others (2012) Assessing the diversity and specificity of two freshwater viral communities through metagenomics. PLOS ONE 7: e33641
Roux S, Tournayre J, Mahul A, Debroas D, Enault F (2014a) Metavir 2: new tools for viral metagenome comparison and assembled virome analysis. BMC Bioinformatics 15: 76
Roux S, Hawley AK, Beltran MT, Scofield M and others (2014b) Ecology and evolution of viruses infecting uncultivated SUP05 bacteria as revealed by single-cell and meta-genomics. eLife 3:e03125
Roux S, Hallam SJ, Woyke T, Sullivan MB (2015) Viral dark matter and virus–host interactions resolved from publicly available microbial genomes. eLife 4:e08490
Short SM, Suttle CA (2003) Temporal dynamics of natural communities of marine algal viruses and eukaryotes. Aquat Microb Ecol 32:107–119
Steward GF, Preston CM (2011) Analysis of a viral metagenomic library from 200 m depth in Monterey Bay, California constructed by direct shotgun cloning. Virol J 8:287
Steward GF, Montiel JL, Azam F (2000) Genome size distributions indicate variability and similarities among marine viral assemblages from diverse environments. Limnol Oceanogr 45:1697–1706
Sullivan MB, Waterbury JB, Chisholm SW (2003) Cyanophages infecting the oceanic cyanobacterium *Prochlorococcus.* Nature 424:1047–1051
Suttle CA (2005) Viruses in the sea. Nature 437:356–361
Suttle CA (2007) Marine viruses — major players in the global ecosystem. Nat Rev Microbiol 5:801–812
Suzuki R, Shimodaira H (2006) Pvclust: an R package for assessing the uncertainty in hierarchical clustering. Bioinformatics 22:1540–1542
Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol 30:2725–2729
Thingstad TF (2000) Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. Limnol Oceanogr 45:1320–1328
Thurber RV, Haynes M, Breitbart M, Wegley L, Rohwer F (2009) Laboratory procedures to generate viral metagenomes. Nat Protocol 4:470–483
Tomaru Y, Nagasaki K (2007) Flow cytometric detection and enumeration of DNA and RNA viruses infecting marine eukaryotic microalgae. J Oceanogr 63:215–221
Tucker KP, Parsons R, Symonds EM, Breitbart M (2011) Diversity and distribution of single-stranded DNA phages in the North Atlantic Ocean. ISME J 5:822–830
Weinbauer MG (2004) Ecology of prokaryotic viruses. FEMS Microbiol Rev 28:127–181
Weinbauer MG, Rassoulzadegan F (2004) Are viruses driving microbial diversification and diversity? Environ Microbiol 6:1–11
Wommack KE, Colwell RR (2000) Virioplankton viruses in aquatic ecosystems. Microbiol Mol Biol Rev 64:69–114
Yang Y, Motegi C, Yokokawa T, Nagata T (2010) Large-scale distribution patterns of virioplankton in the upper ocean. Aquat Microb Ecol 60:233–246
Yoshida M, Takaki Y, Eitoku M, Nunoura T, Takai K (2013) Metagenomic analysis of viral communities in (hado)-pelagic sediments. PLOS ONE 8:e57271
Zhao Y, Temperton B, Thrash JC, Schwalbach MS and others (2013) Abundant SAR11 viruses in the ocean. Nature 494:357–360

Editorial responsibility: Mya Breitbart, St. Petersburg, Florida, USA

Submitted: June 15, 2016; Accepted: November 18, 2016
Proofs received from author(s): January 27, 2017