

ORIGINAL ARTICLE

Diversity and distribution of single-stranded DNA phages in the North Atlantic Ocean

Kimberly P Tucker^{1,3}, Rachel Parsons², Erin M Symonds¹ and Mya Breitbart¹

¹College of Marine Science, University of South Florida, St Petersburg, FL, USA and ²Bermuda Institute of Ocean Sciences, St George's GE, Bermuda

Knowledge of marine phages is highly biased toward double-stranded DNA (dsDNA) phages; however, recent metagenomic surveys have also identified single-stranded DNA (ssDNA) phages in the oceans. Here, we describe two complete ssDNA phage genomes that were reconstructed from a viral metagenome from 80 m depth at the Bermuda Atlantic Time-series Study (BATS) site in the northwestern Sargasso Sea and examine their spatial and temporal distributions. Both genomes (SARss ϕ 1 and SARss ϕ 2) exhibited similarity to known phages of the *Microviridae* family in terms of size, GC content, genome organization and protein sequence. PCR amplification of the replication initiation protein (*Rep*) gene revealed narrow and distinct depth distributions for the newly described ssDNA phages within the upper 200 m of the water column at the BATS site. Comparison of *Rep* gene sequences obtained from the BATS site over time revealed changes in the diversity of ssDNA phages over monthly time scales, although some nearly identical sequences were recovered from samples collected 4 years apart. Examination of ssDNA phage diversity along transects through the North Atlantic Ocean revealed a positive correlation between genetic distance and geographic distance between sampling sites. Together, the data suggest fundamental differences between the distribution of these ssDNA phages and the distribution of known marine dsDNA phages, possibly because of differences in host range, host distribution, virion stability, or viral evolution mechanisms and rates. Future work needs to elucidate the host ranges for oceanic ssDNA phages and determine their ecological roles in the marine ecosystem.

The ISME Journal (2011) 5, 822–830; doi:10.1038/ismej.2010.188; published online 2 December 2010

Subject Category: microbial population and community ecology

Keywords: marine; *Microviridae*; phage; single-stranded DNA; virus

Introduction

Viruses, the majority of which are phages (viruses that infect bacteria), are abundant and important components of marine ecosystems. Phages control microbial abundance, influence community composition through lysis of specific host organisms (Fuhrman and Schwalbach, 2003; Weinbauer and Rassoulzadegan, 2004), and are a critical link in global biogeochemical cycles (Wommack and Colwell, 2000; Weinbauer, 2004). In addition, phages contribute to bacterial diversity through transduction and lysogenic conversion (Jiang and Paul, 1998; Faruque *et al.*, 1999; Dinsdale *et al.*, 2008; Paul, 2008; Rohwer and Thurber, 2009). Direct counts with epifluorescence microscopy and flow cytometry indicate that there are approximately 10^7 virus-like particles in each milliliter of surface seawater (Marie *et al.*, 1999; Wommack

and Colwell, 2000). The majority of these viruses are believed to be double-stranded DNA (dsDNA) viruses (Wommack and Colwell, 2000; Weinbauer and Rassoulzadegan, 2004) containing genomes ranging from 25 to 70 kilobases (kb) in length (Steward *et al.*, 2000; Sandaa, 2008).

Current knowledge of marine viruses is highly biased toward dsDNA viruses. The small genome sizes of single-stranded DNA (ssDNA) viruses (1–9 kb) (LeClerc, 2002; Fauquet *et al.*, 2005) have been a major obstacle for the examination of marine ssDNA viruses, potentially leading to an underestimation of viral abundance and diversity in the oceans. Typical methods for direct viral counts, such as epifluorescence microscopy and flow cytometry, are unable to enumerate these viruses because of the weak fluorescence signal produced (Tomaru and Nagasaki, 2007). The small, circular ssDNA genomes of these viruses also exclude them from pulsed-field gel electrophoresis studies of viral community diversity (Steward, 2001). Moreover, early metagenomic studies utilized methods that only captured the diversity of the dsDNA viral community (Breitbart *et al.*, 2002, 2004), excluding the ssDNA viruses.

In spite of these methodological limitations, several recent discoveries have suggested that

Correspondence: M Breitbart, College of Marine Science, University of South Florida, 140 Seventh Avenue South, St Petersburg, FL 33701, USA.

E-mail: mya@marine.usf.edu

³Current address: Department of Natural Sciences, College of Coastal Georgia, 3700 Altama Avenue, Brunswick, GA, USA.

Received 26 August 2010; revised 26 October 2010; accepted 30 October 2010; published online 2 December 2010

ssDNA viruses are more prevalent in the marine environment than previously recognized. In 2005, the *Chaetoceros salsgineum* nuclear inclusion virus, a ssDNA virus that infects a bloom-forming diatom, was the first ssDNA virus to be discovered in the marine environment through infection studies of algal cultures with natural viral communities (Nagasaki *et al.*, 2005). Subsequently, another ssDNA virus was identified that infected a related diatom species (Tomaru *et al.*, 2008). In addition, a ssDNA prophage was obtained in culture through induction of a *Synechococcus* isolate from the Gulf of Mexico (McDaniel *et al.*, 2006). However, at this time, the number of marine ssDNA viruses that have been cultured in the laboratory is extremely limited. Metagenomic studies including a multiple displacement amplification step, which is known to enrich for small, circular, ssDNA genomes (Haible *et al.*, 2006; Kim *et al.*, 2008), have identified numerous sequences from ssDNA viruses (including both phages and eukaryotic viruses) in marine environments (Angly *et al.*, 2006; Wegley *et al.*, 2007; Desnues *et al.*, 2008; Rosario *et al.*, 2009). However, the ecology of the ssDNA viruses identified through metagenomic sequencing is completely unknown.

Here, we describe two complete marine ssDNA phage genomes belonging to the *Microviridae*, which were reconstructed from a viral metagenome from 80 m depth at the Bermuda Atlantic Time-series Study (BATS) site in the Sargasso Sea. The depth distribution of each of these ssDNA phages at the BATS site was determined at several time points over a 2-year period through PCR amplification of their replication initiation protein (*Rep*) gene. Finally, the diversity of one of the phage groups was examined in temporal samples collected from the BATS site, as well as spatial samples from transects through the North Atlantic Ocean.

Materials and methods

ssDNA phage genome assembly and analysis

Previously sequenced viral metagenomic data from 80 m depth in June 2005 from the BATS site (Hydrostation S: 32°10'N, 64°30'W) (Angly *et al.*, 2006) in the northwestern Sargasso Sea were assembled in SeqMan (Lasergene DNASTAR, Madison, WI, USA) using criteria of 95% identity over 35 nt. All contiguous sequences (contigs) longer than 1000 nt were examined by BLAST to identify contigs with similarity to known *Microviridae*. Several contigs with similarities to the chlamydiophage-like *Microviridae* were identified; however, no contigs exhibited similarities to the ϕ X174-like phages. Two complete circular genomes related to the chlamydiophage-like *Microviridae* were identified (SARss ϕ 1 and SARss ϕ 2; GenBank accession numbers HQ157198 and HQ157199), as well as two contigs representing partial genome sequences. For the two complete genomes, PCR was performed

using several sets of primers designed throughout the genome to verify the genome sequence assembly. Putative Open Reading Frames (ORFs) of at least 300 nt with both start and stop codons were identified in SeqBuilder (Lasergene DNASTAR) and BLASTp was performed to determine similarity to sequences in GenBank (Altschul *et al.*, 1997).

To determine the relationship of the novel phage genomes within the *Microviridae*, a phylogenetic tree was constructed from the family's three conserved ORFs (the major capsid, the minor capsid pilot and the Rep). Each set of protein sequences was aligned separately using the MUSCLE web server (<http://www.ebi.ac.uk/Tools/muscle/index.html>) hosted by the European Bioinformatics Institute (Edgar, 2004a,b) and the aligned sequences were manually curated in BioEdit (Hall, 1999). The best-fitting amino acid substitution models were identified for the individual gene and concatenated alignments using ProtTest (Drummond and Strimmer, 2001; Guindon and Gascuel, 2003; Abascal *et al.*, 2005). For the concatenated alignment, the LG+I+G+F model was implemented in PhyML to construct the maximum likelihood trees with the approximate likelihood test as support for the branches (Guindon and Gascuel, 2003; Anisimova and Gascuel, 2006; Guindon *et al.*, 2010). Phylogenetic trees were also constructed separately for each ORF (data not shown), which agreed with the concatenated alignment.

Sample collection and processing

To examine the depth distributions and temporal variability of the ssDNA phages, water samples were collected from the BATS station (31°40'N, 64°10'W) in August–October 2007, March 2008, June–December 2008 and July 2009 using Niskin bottles.

Polyethylene glycol precipitation (Sambrook *et al.*, 1989) was used to concentrate viruses from the August–October 2007 samples. Briefly, 100 ml of whole seawater were 0.22 μ m filtered, frozen and stored at –20 °C until processing. Samples were thawed at room temperature, solid polyethylene glycol 8000 was added at a 10% w/v ratio and samples were incubated overnight at 4 °C in the dark. Samples were centrifuged at 11 000 \times g for 45 min at 4 °C, and the DNA was extracted from the pellet using a formamide extraction (Sambrook *et al.*, 1989).

Samples of the viral community from June–August 2008, October 2008, December 2008 and July 2009 were processed using the small-scale filtering protocol described by Culley and Steward (2007). Whole seawater samples (50 ml) were filtered through a 0.22 μ m Sterivex filter (Millipore Corp., Billerica, MA, USA) and then onto a 0.02 μ m Anotop filter (Whatman, Maidstone, UK), which was frozen at –80 °C until extraction. DNA was extracted from the Anotop filters using a Masterpure complete DNA and RNA purification kit (Epicenter, Madison, WI,

USA) with a protocol slightly modified from the manufacturer's instructions as previously described (Culley and Steward, 2007).

Large-scale viral concentrates (150–250 l) were collected from the surface and 100 m at the BATS site in March 2008 and September 2008. Each water sample was concentrated to a volume of <500 ml using a 100 kD tangential flow filtration unit (Hollow fiber filtration cartridge; GE Healthcare, Westborough, MA, USA). The viral concentrates were then filtered through a 0.22 µm Sterivex filter (Millipore) to remove bacteria, and stored at 4 °C until processing. Viruses were further concentrated through polyethylene glycol precipitation (as described above) and resuspended in 8 ml of 0.02 µm filtered seawater. The polyethylene glycol-precipitated viral concentrates were then loaded onto cesium chloride density gradients, ultracentrifuged at 61 000 × g for 3 h at 4 °C, and the 1.2–1.5 g ml⁻¹ fraction was collected (Thurber *et al.*, 2009). DNA was extracted from the purified viral fraction using a formamide extraction (Sambrook *et al.*, 1989).

Samples from two separate transects through the Northern Atlantic Ocean were taken to examine the biogeography of the ssDNA phages. Viral concentrates were processed from 20 l of seawater from station BVMS-8 (41°0.059'N, 64°59.869'W) in April 2008 using the large-scale viral concentrate protocol. The small-scale filtering protocol was used to extract viral DNA from samples taken in October 2008 at sites BV42-3 (30°40'N, 64°19'W), BV42-5 (28°40'N, 64°37'W) and BV42-9 (24°40'N, 65°13'W).

PCR amplification, cloning and sequencing

DNA from each sample was amplified using isothermal strand displacement with random hexamer primers and Phi29 DNA polymerase (GenomiPhi V2 DNA Amplification Kit; GE Healthcare, Chalfont St Giles, UK) following the manufacturer's protocol. It is known that this method preferentially amplifies small, ssDNA circular genomes in mixed communities by two to three orders of magnitude (Haible *et al.*, 2006; Kim *et al.*, 2008); therefore, no attempts were made to quantify the abundance of the ssDNA phage genomes.

To confirm that PCR-quality DNA was successfully extracted from each viral sample, PCR for the portal protein (g20) of cyanomyophages was performed on all samples (Sullivan *et al.*, 2008). Primers CPS1.1 and CPS8.1 were used in 25 µl reactions consisting of 1 µl GenomiPhi-amplified template, 1X Sigma REDTaq buffer (10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.1 mM MgCl₂, 0.01% gelatin), 1 µM of each primer, 0.2 mM dNTPs and 1 unit of Sigma REDTaq DNA polymerase (Sigma-Aldrich, St Louis, MO, USA). The cyanomyophage reaction conditions consisted of an initial denaturation step of 94 °C for 5 min, followed by 35 cycles of 94 °C for 1 min, 35 °C for 1 min and 72 °C for 1 min; then a final elongation step at 72 °C for 10 min.

Negative controls were performed without target DNA for each set of reactions to ensure that there was no spurious amplification.

Primer3 (Rozen and Skaletsky, 2000) was used to design primers to the putative *Rep* of the newly described ssDNA phage genomes. The primers for the SARssφ1 genome were based on a consensus of one complete genome (SARssφ1) and two additional large contigs, whereas the primers for the SARssφ2 genome were designed to be specific to that genome. The primer sequences for SARssφ1 were: SARss-consF (5'-TCATAYACTGCGTAATAWACTTTCTKC-3') SARss-consR (5'-CGAATTATATATATCMCCGAATT RSA-3'). PCR was carried out in 50 µl reactions, using 1X Sigma REDTaq buffer, 1 µM of each primer, 0.2 mM dNTPs, and 1 unit of Sigma REDTaq polymerase (Sigma-Aldrich). The touchdown PCR reaction consisted of the following steps: 95 °C for 5 min, 40 cycles of (94 °C for 1 min, an annealing step of 60 °C that decreased by 0.5 °C each cycle, then 72 °C for 3 min), followed by a final extension step at 72 °C for 10 min. The SARssφ1 PCR product was 695 bp in length. The primers for SARssφ2 were SARssφ2_3371F (5'-AACACAAGCG GAAGACCACT-3') and SARssφ2_4473R (5'-TGT TTAGCTGGCGGTTTCTT-3'). Reactions were carried out in 25 µl volumes, using 1X Apex Red Taq buffer (750 mM Tris-HCl, pH 8.5, (NH₄)₂SO₄, 1% Tween20), 1 µM each primer, 0.2 mM dNTPs, 0.5 U Apex Red Taq DNA polymerase. Cycling conditions for these reactions were as follows: 95 °C for 5 min, 40 cycles of 95 °C for 1 min, 63 °C for 1 min and 72 °C for 1 min; followed by a final elongation step of 72 °C for 10 min. Negative controls without template DNA were included for each set of reactions. The SARssφ2 PCR product was 1102 bp in length. PCR products were cloned into the pCR2.1-TOPO vector (Invitrogen, Carlsbad, CA, USA), and the vector was transformed into chemically competent DH5α-T1 cells. Clones were screened by PCR to ensure inserts of the proper size were present, then sequenced with the M13F primer (5'-GTAAAAC GACGGCCAG-3') by Beckman Coulter Genomics (Danvers, MA, USA).

SARssφ1 *Rep* sequence analysis

Vector sequences were removed and the low-quality regions were trimmed using SEQUENCHER 4.5 (Gene Codes, Ann Arbor, MI, USA). Sequences were deposited to GenBank with accession numbers HQ142125–HQ142384. All sequences were aligned using the MUSCLE web server hosted by the European Bioinformatics Institute (Edgar, 2004a, b) and manually curated in BioEdit (Hall, 1999). The best-fitting model of nucleotide substitution was identified for the alignments using jModelTest (Guindon and Gascuel, 2003; Posada, 2008). Phylogenetic trees were then constructed from the nucleotide sequences using the GTR+I+G model in PhyML, and the approximate likelihood ratio test

values were examined for branch support (Guindon and Gascuel, 2003; Anisimova and Gascuel, 2006; Guindon *et al.*, 2010).

To examine temporal variability, the alignments from all sequences obtained from the BATS site were visualized in Geneious Pro v5.0.4 (Drummond *et al.*, 2010) to identify patterns of nucleotide substitutions. To examine spatial variability, sequences obtained from each location in 2008 were de-replicated at 99% sequence identity with gaps using FastGroupII (Yu *et al.*, 2006) before alignments were created. The tree for the spatial samples was integrated with the geospatial data using GenGIS v1.06 and MapMaker v1.0 to create the two-dimensional georeferenced phylogenetic tree (Parks *et al.*, 2009).

Mantel tests were performed using matrices of pairwise differences to determine if genetic divergence was correlated to the time of sampling or geographic location. The pairwise genetic distances were calculated using the GTR model, which was previously determined to be the best-fitting model of nucleotide substitution, implemented in DAMBE (Xia and Xie, 2001) and then manually formatted as a full matrix using Microsoft Excel. Samples collected from the BATS site in June 2005, August–October 2007, March 2008, August 2008, September 2008 and July 2009 were utilized for the temporal comparisons. For the spatial samples, the pairwise geographic distances were calculated for the sequences obtained from the BVMS-8 (March 2008), Hydrostation S (September 2008), BV42-3, BV42-5 and BV42-9 (October 2008) sites, as well as from the BATS samples from 2008. The BATS data set was limited to samples obtained in 2008 to reduce bias associated with any temporal signal. The latitude and longitude for each sample was input into the Geographic Distance Matrix Calculator (http://biodiversityinformatics.amnh.org/open_source/gdmg/) to generate a full pairwise distance matrix in kilometers. The temporal distance matrix and geographic distribution matrix were each compared with genetic distance using a Mantel test as implemented in XLSTAT, with 10 000 permutations of the data to determine significance.

Results and Discussion

Marine ssDNA phage genome sequences

Two complete ssDNA genomes were assembled from the previously sequenced viral metagenome from 80 m at the BATS site (Angly *et al.*, 2006) and verified by PCR. These are the first completely sequenced ssDNA phages from the marine environment. The novel genomes represent a community genome, as they are a mixture of very closely related viral genomes. Genomic characteristics suggest that these viruses are novel members of the family *Microviridae*, which are icosahedral viruses with circular ssDNA genomes between 4.4 and 6.1 kb (Fane, 2005). Members of the *Microviridae* infecting a diverse range of hosts (including proteobacteria, *Spiroplasma*, *Chlamydia*) have been isolated and completely sequenced (Sanger *et al.*, 1977; Renaudin *et al.*, 1987; Storey *et al.*, 1989; Lui *et al.*, 2000; Read *et al.*, 2000; Brentlinger *et al.*, 2002; Garner *et al.*, 2004; Rokyta *et al.*, 2006). Genome comparisons have identified two distinct groups of *Microviridae* (those similar to the chlamydia phages and those similar to *Escherichia coli* ϕ X174), leading to the suggestion that ssDNA phages evolve through different mechanisms than dsDNA phages (Brentlinger *et al.*, 2002; Rokyta *et al.*, 2006). Although the *Microviridae* have been extensively studied, the sequenced representatives infect a small number of bacterial hosts. Expansion of the *Microviridae* family with environmental sequences will enable a better understanding of their evolutionary mechanisms and ecological impacts.

Both phage genomes (SARss ϕ 1 and SARss ϕ 2) exhibited similarities to known phages of the *Microviridae* in terms of size, GC content, genome organization and protein sequences. The SARss ϕ 1 genome was 4487 nt with 43.9% GC and the SARss ϕ 2 genome was 4478 nt with 46.9% GC. The genome organizations of these phages were syntenous with the other members of the *Microviridae* (Figure 1). The phage genomes were most closely related to partial ssDNA phage sequences identified in microbialites through viral metagenomics (Desnues *et al.*, 2008), but also displayed similarities to cultured members of the *Microviridae*. Maximum

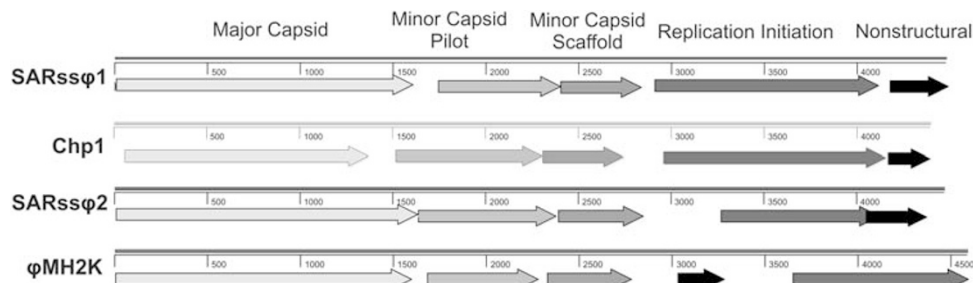


Figure 1 Organization of five major genes in the SARss ϕ 1 and SARss ϕ 2 genomes compared with two other members of the *Microviridae* (*Chlamydia* phage 1 (Chp1), and *Bdellovibrio* phage ϕ MH2K).

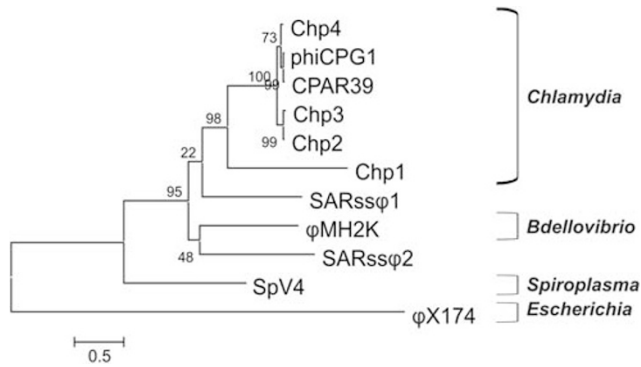


Figure 2 Phylogenetic relationship of the marine ssDNA phages compared with known members of the *Microviridae* based on the three conserved ORFs (the major capsid, the minor capsid pilot and the Rep). The phages included in the tree are *Escherichia* phage ϕ X174 (NC_001422.1), *Spiroplasma* phage SpV4 (NC_003438.1), *Bdellovibrio* phage ϕ MH2K (NC_002643.1), the *Chlamydia* phages 1–4 (NC001741, NC_002194, NC_008355, NC_007461, respectively), CPAR39 (NC_002180) and phiCPG1 (NC_001998).

likelihood phylogenetic trees of three concatenated ORFs (the major capsid, the minor capsid pilot and the Rep) demonstrated that these two genomes represent novel members of the *Microviridae*. The SARss ϕ 1 phage genome clustered just outside of the known *Chlamydia* phages, whereas the SARss ϕ 2 phage genome clustered with *Bdellovibrio* phage ϕ MH2K (Figure 2). As these two novel ssDNA phages originate from the marine environment and infect unknown hosts, it is interesting that they clearly cluster with the chlamydiaophage subfamily of the *Microviridae*, and are only distantly related to ϕ X174. This phylogenetic placement supports the divided nature of the *Microviridae* family, suggesting that intermediates between these two subfamilies may not exist because of the nature of evolution of ssDNA phages (Brentlinger *et al.*, 2002).

Distribution of ssDNA phages throughout the water column at the BATS site

The depth distribution for each of the newly identified ssDNA phage genomes throughout the upper 200 m of the water column at the BATS site was determined using PCR for the Rep gene. The presence of high-quality phage DNA in each sample was verified through PCR for the *g20* gene of cyanomyophages (dsDNA phages that infect cyanobacteria), which was found at all depths tested. In contrast to the cyanomyophages, the two ssDNA phages displayed narrower depth distributions, and these distributions were distinct from each other (Table 1). SARss ϕ 1 was most frequently identified at 80 m depth (75% of the samples; $n = 8$), but also recovered in at least 50% of the samples taken from 100 to 120 m. SARss ϕ 1 was identified at depths ranging from 40 to 160 m, but was never recovered from surface waters (0 or 20 m) or the deepest samples (200 m). In contrast, SARss ϕ 2 was most

Table 1 Presence of SARss ϕ 1 and SARss ϕ 2 at the BATS site for a variety of depths sampled between 2007 and 2009. The presence of high-quality phage DNA in each of the samples was verified through PCR amplification of the cyanomyophage *g20* gene

Depth (No.)	SARss ϕ 1	SARss ϕ 2
0 m ($n = 5$)	0	0
20 m ($n = 7$)	0	0
40 m ($n = 7$)	1	0
60 m ($n = 8$)	2	2
80 m ($n = 8$)	6	0
100 m ($n = 9$)	6	4
120 m ($n = 8$)	4	5
140 m ($n = 7$)	2	6
160 m ($n = 2$)	1	2
200 m ($n = 5$)	0	4

Abbreviation: BATS, Bermuda Atlantic Time-series Study. The number of samples examined is shown in the first column, and the number of samples positive for each phage is shown in the second and third columns. Dark gray indicates that the phage was present in >75% of the samples from that depth, light gray indicates that the phage was present in fewer than 75% of the samples from that depth and white indicates that the phage was not detected at that depth.

frequently detected at depths of 100 m or greater. SARss ϕ 2 was never detected in samples from 40 m or shallower, and was only sporadically detected at 60 m. Interestingly, at 80 m, where SARss ϕ 1 was most frequently detected, SARss ϕ 2 was never found. SARss ϕ 2 was found in >75% of the samples from 140 m or deeper, including 200 m, where this phage was detected in 80% of the samples ($n = 5$). As 200 m was the deepest depth tested, it is unknown if SARss ϕ 2 was also found deeper in the water column. The absence of both ssDNA phages in surface waters at the BATS site was verified through examination of large-scale (200 l) viral concentrates collected from the surface and 100 m depths at three unique time points. Although both ssDNA phages were consistently identified in the 100 m viral concentrates, neither phage was recovered from the surface water viral concentrates.

The distinct depth distributions of these two ssDNA phages suggest that they infect different hosts. Although some potential host microorganisms can be found throughout the upper 200 m, highly structured depth distributions, which vary seasonally in accordance with the degree of water column stratification, have been demonstrated for bacterial communities at the BATS site (Morris *et al.*, 2005; Carlson *et al.*, 2009; Treusch *et al.*, 2009). No seasonal differences in the presence or depth distribution of the ssDNA phages were apparent from our data and the hosts for these phages still remain unknown.

Temporal variation in ssDNA phage sequence diversity at the BATS site

A portion of the Rep gene from phages related to SARss ϕ 1 was amplified and sequenced from samples collected from the BATS site in June 2005 and at various time points between August 2007 and July

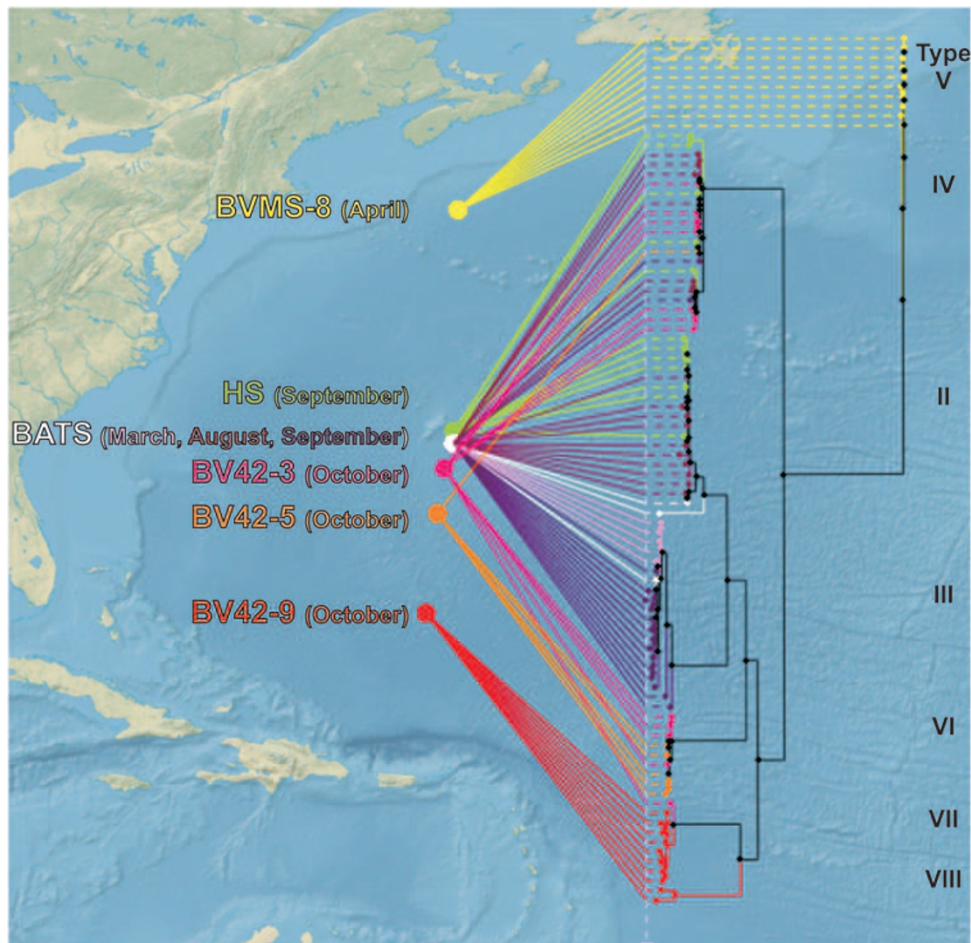


Figure 4 Spatial variation in the diversity of the *Rep* gene from SARss ϕ 1 phages from different sites in the North Atlantic Ocean in 2008. Sequences were de-replicated at 99% sequence identity with gaps, then a maximum likelihood phylogeny was constructed and layered on top of a map. Samples are color-coded based on location. The SARss ϕ 1 phage genome is indicated with a white star, and the other two contigs used to design the primers are also shown in white.

was amplified from different sites between Massachusetts and Puerto Rico in 2008. Throughout the transects, SARss ϕ 1 was identified at depths of 80–100 m, but never found in surface water samples, consistent with the depth distribution of this phage at the BATS site. The georeferenced tree showed that the *Rep* sequences generally clustered by location (Figure 4). Sequences obtained from Hydrostation S clustered with sequences from the BATS site, which is likely due to the fact that these sites are only ~56 km apart. Divergent sequences appeared with increasing distance from the BATS site, and a Mantel test indicated a positive correlation between genetic distance and geographic distance between sampling locations ($n = 177$; $r(\text{AB}) = 0.573$; $P < 0.0001$). Although some sequences from sites BV42-3 and BV42-5 clustered with sequences from the BATS site (types II–IV), many sequences from these southern sites belonged to novel clades (types VI and VII). Each viral type contained a variety of nucleotide polymorphisms, but many of the changes were synonymous (Supplementary Figure 1). Although a diversity of ssDNA phage sequences was recovered from the

southernmost (BV42-9; types VII and VIII) and northernmost (BVMS-8; type V) sites, these sequences displayed no genetic overlap with sequences identified from BATS, demonstrating that genetic distance of sequences increased with geographic distance.

Differences between ssDNA and dsDNA phages

This is the first in-depth analysis of the prevalence and diversity of environmental ssDNA phages over depth, time and space. Several previous studies have examined the biogeographical distribution of specific dsDNA viruses and found that these viruses are widely distributed in nature (Kellogg *et al.*, 1995; Short and Suttle, 2002, 2005; Breitbart and Rohwer, 2004; Labonte *et al.*, 2009; Huang *et al.*, 2010). The recovery of nearly identical sequences from disparate environments throughout the world has led to the suggestion that there is a shared global gene pool for viruses (Breitbart and Rohwer, 2005; Angly *et al.*, 2006). In contrast to the cosmopolitan distribution of some dsDNA phage sequences (Breitbart and Rohwer, 2004; Short and Suttle, 2005; Huang *et al.*, 2010), the

ssDNA phages described here displayed narrower (and distinct) depth distributions and their sequence divergence correlated with geographic distance. These findings are consistent with previous observations for ssDNA phages in microbialite systems, which were found to have extremely narrow geographic distributions (Desnues *et al.*, 2008). At present, the reason behind this difference in biogeography between the marine ssDNA phages described here and previously studied dsDNA phages is unknown; however, it may be related to the distribution of their hosts or fundamental differences in viral lifestyle, genome evolution rates or virion stability. Future work needs to elucidate the host ranges of these and other ssDNA phages, and determine the roles of ssDNA phages in marine microbial ecology.

Acknowledgements

This research was funded by a grant to MB from the National Science Foundation (MCB-0701984). Thanks to Steven Bell, Craig Carlson, Dawn Goldsmith, Kevin Lew, Ben Van Mooy, Karyna Rosario, and the crew of the RV Atlantic Explorer for help collecting and processing samples. Thanks to Bhakti Dwivedi for bioinformatic assistance.

References

- Abascal F, Zardoya R, Posada D. (2005). ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* **21**: 2104–2105.
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W *et al.* (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402.
- Angly F, Felts B, Breitbart M, Salamon P, Edwards R, Carlson C *et al.* (2006). The marine viromes of four oceanic regions. *PLoS Biol* **4**: e368.
- Anisimova M, Gascuel O. (2006). Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. *Syst Biol* **55**: 539–552.
- Breitbart M, Felts B, Kelley S, Mahaffy JM, Nulton J, Salamon P *et al.* (2004). Diversity and population structure of a near-shore marine-sediment viral community. *Proc R Soc Lond B Biol Sci* **271**: 565–574.
- Breitbart M, Rohwer F. (2004). Global distribution of nearly identical phage-encoded DNA sequences. *FEMS Microbiol Lett* **236**: 245–252.
- Breitbart M, Rohwer F. (2005). Here a virus, there a virus, everywhere the same virus? *Trends Microbiol* **13**: 278–284.
- Breitbart M, Salamon P, Andresen B, Mahaffy JM, Segall AM, Mead D *et al.* (2002). Genomic analysis of uncultured marine viral communities. *Proc Natl Acad Sci USA* **99**: 14250–14255.
- Brentlinger K, Hafenstein S, Novak C, Fane B, Borgon R, McKenna R *et al.* (2002). *Microviridae*, a family divided: isolation, characterization, and genome sequence of ϕ MH2K, a bacteriophage of the obligate intracellular parasitic bacterium *Bdellovibrio bacteriovorus*. *J Bacteriol* **184**: 1089–1094.
- Carlson C, Morris R, Parsons R, Treusch A, Giovannoni S, Vergin K. (2009). Seasonal dynamics of SAR11 populations in the euphotic and mesopelagic zones of the northwestern Sargasso Sea. *ISME J* **3**: 283–295.
- Cuevas J, Duffy S, Sanjuan R. (2009). Point mutation rate of bacteriophage ϕ X174. *Genetics* **183**: 747–749.
- Culley A, Steward G. (2007). New genera of RNA viruses in subtropical seawater, inferred from polymerase gene sequences. *Appl Environ Microbiol* **73**: 5937–5944.
- Desnues C, Rodriguez-Brito B, Rayhawk S, Kelley S, Tran T, Haynes M *et al.* (2008). Biodiversity and biogeography of phages in modern stromatolites and thrombolites. *Nature* **452**: 340–343.
- Dinsdale EA, Edwards RA, Hall D, Angly F, Breitbart M, Brulic JM *et al.* (2008). Functional metagenomic profiling of nine biomes. *Nature* **452**: 629–632.
- Drummond A, Ashton B, Buxton S, Cheung M, Cooper A, Heled J *et al.* (2010). Geneious v5.0.4 <http://www.geneious.com>.
- Drummond A, Strimmer K. (2001). PAL: an object-oriented programming library for molecular evolution and phylogenetics. *Bioinformatics* **17**: 662–663.
- Duffy S, Holmes E. (2009). Validation of high rates of nucleotide substitution in geminiviruses: phylogenetic evidence from *East African cassava mosaic viruses*. *J Gen Virol* **90**: 1539–1547.
- Duffy S, Shackleton L, Holmes E. (2008). Rates of evolutionary change in viruses: patterns and determinants. *Nat Rev Genet* **9**: 267–276.
- Edgar R. (2004a). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792–1797.
- Edgar R. (2004b). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinform* **5**: 113.
- Fane B (2005). Family microviridae. In: Fauquet C, Mayo M, Maniloff J, Desselberger U, Ball L (eds). *Virus Taxonomy: Classification and Nomenclature of Viruses*. Elsevier Academic Press: San Diego, pp 289–299.
- Faruque SM, Rahman MM, Asadulghani, Islam KMN, Mekalanos JJ. (1999). Lysogenic conversion of environmental *Vibrio mimicus* strains by CTX ϕ . *Infect Immun* **67**: 5723–5729.
- Fauquet C, Mayo M, Maniloff J, Desselberger U, Ball L (eds) (2005). *Virus Taxonomy: Classification and Nomenclature of Viruses*. Elsevier Academic Press: San Diego.
- Fuhrman J, Schwalbach M. (2003). Viral influence of aquatic bacterial communities. *Biol Bull* **204**: 192–195.
- Garner S, Everson J, Lambden P, Fane B, Clarke I. (2004). Isolation, molecular characterisation and genome sequence of a bacteriophage (Chp3) from *Chlamydomonas percorum*. *Virus Genes* **28**: 207–214.
- Guindon S, Daufayard J, Lefort V, Anisimova M, Hordijk W, Gascuel O. (2010). New algorithms and methods to estimate maximum likelihood phylogenies: assessing the performance of PhyML3.0. *Syst Biol* **59**: 307–321.
- Guindon S, Gascuel O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**: 696–704.
- Haible D, Kober S, Jeske H. (2006). Rolling circle amplification revolutionizes diagnosis and genomics of geminiviruses. *J Virol Methods* **135**: 9–16.
- Hall T. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* **41**: 95–98.
- Huang SJ, Wilhelm SW, Jiao NAZ, Chen F. (2010). Ubiquitous cyanobacterial podoviruses in the global oceans unveiled through viral DNA polymerase gene sequences. *ISME J* **4**: 1243–1251.

- Jiang SC, Paul JH. (1998). Gene transfer by transduction in the marine environment. *Appl Environ Microbiol* **64**: 2780–2787.
- Kellogg CA, Rose JB, Jiang SC, Turmond JM, Paul JH. (1995). Genetic diversity of related vibriophages isolated from marine environments around Florida and Hawaii, USA. *Mar Ecol Prog Ser* **120**: 89–98.
- Kim K, Chang H, Nam Y, Roh S, Kim M, Sung Y et al. (2008). Amplification of uncultured single-stranded DNA viruses from rice paddy soil. *Appl Environ Microbiol* **74**: 5975–5985.
- Labonte JM, Reid KE, Suttle CA. (2009). Phylogenetic analysis indicates evolutionary diversity and environmental segregation of marine podovirus DNA polymerase gene sequences. *Appl Environ Microbiol* **75**: 3634–3640.
- LeClerc J. (2002). Single-stranded DNA phages. In: Streips U, Yasbin R (eds). *Modern Microbial Genetics* 2nd edn. Wiley-Liss, Inc.: New York, pp 145–176.
- Lui B, Everson J, Fane B, Giannikopoulou P, Vertou E, Lambden P et al. (2000). Molecular characterization of a bacteriophage (Chp2) from *Chlamydia psittaci*. *J Virol* **74**: 3464–3469.
- Marie D, Brussaard C, Thyrhaug R, Bratbak G, Vaulot D. (1999). Enumeration of marine viruses in culture and natural samples by flow cytometry. *Appl Environ Microbiol* **65**: 45–52.
- McDaniel L, Paul J, de la Rosa M. (2006). Temperate and lytic cyanophages from the Gulf of Mexico. *J Mar Biol Assoc UK* **86**: 517–527.
- Morris RM, Vergin KL, Cho JC, Rappe MS, Carlson CA, Giovannoni SJ. (2005). Temporal and spatial response of bacterioplankton lineages to annual convective overturn at the Bermuda Atlantic Time-series Study site. *Limnol Oceanogr* **50**: 1687–1696.
- Nagasaki K, Tomaru Y, Takao Y, Nishida K, Shirai Y, Suzuki H et al. (2005). Previously unknown virus infects marine diatom. *Appl Environ Microbiol* **71**: 3528–3535.
- Parks D, Porter M, Churcher S, Wang S, Blouin C, Whalley J et al. (2009). GenGIS: a geospatial information system for genomic data. *Genome Res* **19**: 1896–1904.
- Paul JH. (2008). Prophages in marine bacteria: dangerous molecular time bombs or the key to survival in the seas? *ISME J* **2**: 579–589.
- Posada D. (2008). jModelTest: phylogenetic model averaging. *Mol Biol Evol* **25**: 1253–1256.
- Read T, Brunham R, Shen C, Gill S, Heidelberg J, White O et al. (2000). Genome sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39. *Nucleic Acids Res* **28**: 1397–1406.
- Renaudin J, Paracel M, Bove J. (1987). Spiroplasma virus 4: nucleotide sequence of the viral DNA, regulatory signals, and the proposed genome organization. *J Bacteriol* **169**: 4950–4961.
- Rohwer F, Thurber RV. (2009). Viruses manipulate the marine environment. *Nature* **459**: 207–212.
- Rokyta D, Burch C, Caudle S, Wichman H. (2006). Horizontal gene transfer and the evolution of microvirid coliphage genomes. *J Bacteriol* **188**: 1134–1142.
- Rosario K, Duffy S, Breitbart M. (2009). Diverse circovirus-like genome architectures revealed by environmental metagenomics. *J Gen Virol* **90**: 2418–2424.
- Rozen S, Skaletsky H. (2000). Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds). *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Humana Press: Totowa, NJ, pp 365–386.
- Sambrook J, Fritsch E, Maniatis T. (1989). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press: New York.
- Sandaa R. (2008). Burden or benefit? Virus–host interactions in the marine environment. *Res Microbiol* **159**: 374–381.
- Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes CA et al. (1977). Nucleotide sequence of bacteriophage ϕ X174 DNA. *Nature* **265**: 687–695.
- Short C, Suttle C. (2005). Nearly identical bacteriophage structural gene sequences are widely distributed in marine and freshwater environments. *Appl Environ Microbiol* **71**: 480–486.
- Short SM, Suttle CA. (2002). Sequence analysis of marine virus communities reveals that groups of related algal viruses are widely distributed in nature. *Appl Environ Microbiol* **68**: 1290–1296.
- Steward G. (2001). Fingerprinting viral assemblages by pulsed field gel electrophoresis. In: Paul J (ed). *Methods in Microbiology*. Academic Press: San Diego, pp 85–103.
- Steward G, Montiel J, Azam F. (2000). Genome size distributions indicate variability and similarities among marine viral assemblages from diverse environments. *Limnol Oceanogr* **45**: 1697–1706.
- Storey CC, Lusher M, Richmond SJ. (1989). Analysis of the complete nucleotide sequence of Chp1, a phage which infects avian *Chlamydia psittaci*. *J Gen Virol* **70**: 3381–3390.
- Sullivan M, Coleman M, Quinlivan V, Rosenkrantz J, DeFrancesco A, Tan G et al. (2008). Portal protein diversity and phage ecology. *Environ Microbiol* **10**: 2810–2823.
- Thurber R, Haynes M, Breitbart M, Wegley L, Rohwer F. (2009). Laboratory procedures to generate viral metagenomes. *Nat Protocols* **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.
- Tomaru Y, Shirai Y, Suzuki H, Nagumo T, Nagasaki K. (2008). Isolation and characterization of a new single-stranded DNA virus infecting the cosmopolitan marine diatom *Chaetoceros debilis*. *Aqua Microbiol Ecol* **50**: 103–112.
- Treusch A, Vergin K, Finlay L, Donatz M, Burton R, Carlson C et al. (2009). Seasonality and vertical structure of microbial communities in an ocean gyre. *ISME J* **3**: 1148–1163.
- Wegley L, Edwards R, Rodriguez-Brito B, Liu H, Rohwer F. (2007). Metagenomic analysis of the microbial community associated with the coral *Porites astreoides*. *Environ Microbiol* **9**: 2707–2719.
- Weinbauer M. (2004). Ecology of prokaryotic viruses. *FEMS Microbiol Rev* **28**: 127–181.
- Weinbauer M, Rassoulzadegan F. (2004). Are viruses driving microbial diversification and diversity? *Environ Microbiol* **6**: 1–11.
- Wommack K, Colwell R. (2000). Virioplankton: viruses in aquatic ecosystems. *Microbiol Mol Biol Rev* **64**: 69–114.
- Xia X, Xie Z. (2001). DAMBE: software package for data analysis in molecular biology and evolution. *J Hered* **92**: 371–373.
- Yu Y, Breitbart M, McNairnie P, Rohwer F. (2006). FastGroupII: a web-based bioinformatics platform for analyses of large 16S rDNA libraries. *BMC Bioinform* **7**: 57.

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