

FEATURE ARTICLE



Virus genes in Arctic marine bacteria identified by metagenomic analysis

Matthew T. Cottrell, David L. Kirchman*

School of Marine Science and Policy, University of Delaware, 700 Pilottown Rd., Lewes, Delaware 19958, USA

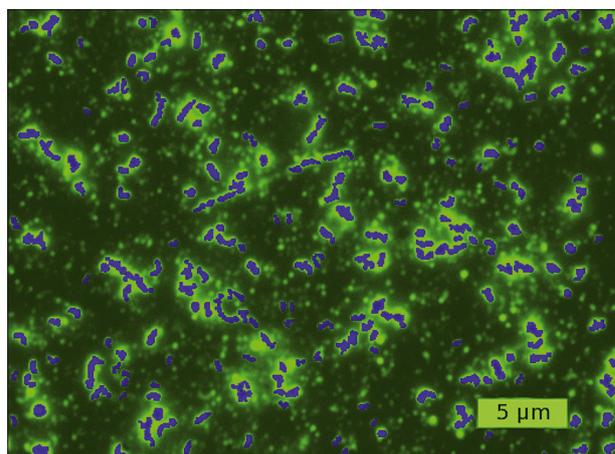
ABSTRACT: The relationship between viruses and prokaryotes in the Arctic appears to differ from that at lower latitudes because of low temperature and low abundances of viruses and bacteria, but the impact of viruses on Arctic bacterial genomes is unclear. In the present study, genome fragments of bacteria from the western Arctic Ocean were sequenced to explore the prevalence of virus genes in the genomes of bacteria inhabiting these perennially cold waters. Arctic bacterial genomes were sampled by cloning bacterial environmental DNA into a fosmid vector, and virus DNA within the bacterial genomes was identified using marine viral metagenomes to query the bacterial genomes using the basic local alignment search tool (BLAST). Virus DNA flanked by bacterial DNA was identified as being part of bacterial genomes. Virus DNA was ~3-fold more abundant in Arctic bacterial genomes than in bacterial genomes from Monterey Bay and ~10-fold more abundant than in bacterial genomes from Antarctic waters. Phage terminase genes involved in packaging DNA into phage capsids were the most abundant gene family identified in the Arctic bacterial genomes. Viruses appear to have a larger impact on prokaryotic communities in the Arctic than what might be inferred from the low bacterial and viral abundances in this high latitude ocean.

KEY WORDS: Bacterial metagenome · Viruses · Phages · Lysogeny · Arctic · Fosmid

Resale or republication not permitted without written consent of the publisher

INTRODUCTION

Heterotrophic prokaryotes actively participate in the Arctic Ocean food web and play key roles in the overall metabolism of the Arctic. The prokaryotic



Composite epifluorescence micrograph showing viruses (green, SYBR Gold-stained) with uncultured marine bacteria (blue, DAPI-stained)

Image by Matthew T. Cottrell

communities in surface waters of the Arctic Ocean are dominated by *Bacteria* even though the abundance of *Archaea* is higher in this perennially cold, high latitude environment than in temperate and tropical waters (Wells et al. 2006, Kirchman et al. 2007, Alonso-Sáez et al. 2008). Although bacteria appear to be adapted to low temperatures and other environmental conditions found at high latitudes (Connelly et al. 2006, Grzymiski et al. 2006), their biogeochemical impact appears to differ from that in lower latitude oceans. The level of bacterial biomass production relative to primary production is lower in the Arctic Ocean and other polar waters than in warmer, lower latitude waters, indicating a unique food web structure and carbon cycling in these perennially cold waters (Kirchman et al. 2009).

The relationship between viruses and prokaryotes in the Arctic also appears to differ from that in lower

*Corresponding author. Email: kirchman@udel.edu

latitudes. Bacterial and viral abundances are low in the Arctic (Steward et al. 2007, Payet & Suttle 2008), implying that contact rates between viral particles and host cells are probably also low (Murray & Jackson 1992). The frequency of visibly infected cells indicates that viral lysis is low in the upper water column of the Arctic (Steward et al. 1996, 2007). In contrast, estimates of virus production rates measured in Arctic bottom waters suggest that viral lysis rates are comparable to bacterial growth rates (Wells & Deming 2006).

Metagenomic analyses have provided clues about virulent viruses actively replicating in uncultivated bacteria in the oceans (DeLong et al. 2006, Williamson et al. 2008). The majority of the virus genes identified in the Global Ocean Survey (GOS) sequence data base appear to come from actively replicating viruses (Williamson et al. 2008) rather than viral DNA integrated into bacterial genomes. The high abundance of *Prochlorococcus* cyanophage DNA that co-purified with bacterial genomic DNA suggests that on average 6% of *Prochlorococcus* cells contain actively replicating viruses (Williamson et al. 2008). Similarly, environmental DNA extracted from the bacterial size fraction of waters in the North Pacific Subtropical Gyre contained a large amount of DNA from actively replicating cyanophages, accounting for 1 to 10% of the total sequences and 60 to 80% of the virus sequences (DeLong et al. 2006). A recent study at the Hawaii Ocean Time-series Station, however, did not find large amounts of phage DNA in the bacterial size fraction (Shi et al. 2011).

Virus genes are likely to be common in genomes of uncultivated bacteria because prophages have been found in 60 to 70% of all sequenced genomes of cultivated bacteria (Paul 2008). In addition, about half of all cultivated marine bacteria contain inducible prophages (Jiang & Paul 1994, 1998). However, no study has systematically surveyed virus DNA in genomes of uncultivated bacteria. One study in Tampa Bay, Florida, waters detected a *Vibrio* phage-like and a *Clostridium* phage-like integrase gene, which enables virus DNA to be integrated into the host genome, in only 3% and 0.5%, respectively, of the bacterial genomes in the GOS data set (McDaniel et al. 2008). Other viral genes were not examined. The GOS project did not sample the Arctic Ocean (www.jcvi.org/cms/research/projects/gos/overview/), and no clear relationship has yet emerged between environment and integrase gene phylotype or frequency (McDaniel et al. 2008).

We examined 58 genome fragments of bacteria from the western Arctic Ocean to explore questions

about the prevalence of virus genes in uncultivated bacteria inhabiting these perennially cold waters. In the present study, genomic fragments of 40 kb were cloned into a fosmid vector and sequenced to explore virus DNA in bacterial genomes. In contrast to previous metagenomic studies, a fosmid-based approach can distinguish between viruses in the bacterial size fraction (previous studies) and viral genes in bacterial genomes (present study). Our analysis uncovered a higher frequency of phage genes in uncultivated bacterial genomes in the Arctic than in bacterial genomes in lower latitude oceans. Viruses apparently have substantial interactions with bacteria in the Arctic, despite the low host and viral abundances in these waters (Steward et al. 2007, Payet & Suttle 2008).

MATERIALS AND METHODS

Fosmid library construction and screening

The Arctic fosmid library, containing 4800 clones, was constructed using DNA isolated from cells in the <0.8 μm to >0.2 μm size fraction of a seawater sample collected at a depth of 10 m in the Chukchi Sea (72° 19.33' N, 151° 59.07' W) in July 2004 and screened as described previously (Cottrell et al. 2005). Pools of 96 fosmid clones were screened for 16S rRNA genes by denaturing gradient gel electrophoresis (DGGE) of PCR amplicons generated with primers GC358F and 517R (Muyzer et al. 1995). Selected bands resolved on an 8% polyacrylamide gel containing a gradient of 25 to 55% denaturant (13.8 to 22% formamide and 10.5 to 23% urea) were re-amplified and sequenced. Phylogenetic classification of 16S rRNA genes was conducted using the Greengenes classifier (DeSantis et al. 2006).

The fosmid pools were also screened for proteorhodopsin (PR) genes by PCR (Sabehi et al. 2005) using buffer provided with the *Taq* polymerase (Promega), 200 nM primers and 0.7 mM MgCl_2 . The thermal cycling consisted of 96°C denaturing, 50°C primer annealing and 72°C DNA polymerizing steps repeated 35 times.

Fosmid sequencing and annotation

The fosmids bearing 16S rRNA genes, PR genes and hydrolase genes were completely sequenced by the Broad Institute and are available from the GenBank data base under accession no. EU795084 to

EU795107, EU795222 to EU795255 and EU795278. Open reading frames were identified using the FGENESB (Softberry) web site (www.softberry.com) and then annotated using AutoFACT software (Koski et al. 2005), which generates an annotation based on a consensus of the basic local alignment search tool (BLAST) (Altschul et al. 1990) results obtained with searches against the UniRef90, COG, KEGG, Pfam, Smart, LSU ribosomal RNA and SSU ribosomal RNA data bases. The annotation was obtained using the default setting in the software. The BLAST results obtained from the different data bases were also used for the different analyses presented below, including taxonomic identification based on a consensus of the top BLAST hits of all of the fosmid protein-encoding genes.

The 16S rRNA genes carried by fosmid clones were compared to a data base of Arctic 16S rRNA gene tags generated by 454 pyrosequencing (Kirchman et al. 2010) using BLAST with the *blastn* option. Hits with e-values $<10^{-8}$ and similarities $>85\%$ were considered significant.

The Arctic metagenome and those from the Hawaii Ocean Time-series (HOT) (DeLong et al. 2006) and the coastal waters near Cape May, NJ (Rusch et al. 2007), were compared using end sequences of 1920 randomly selected Arctic fosmid clones. The fosmid end sequences are available from the GenBank data base under accession no. JS800118 to JS803928. End sequences were used because complete fosmid sequences are not available from the other sites. Functional genes in the Arctic fosmid end sequences were compared to those from HOT and Cape May using the ShotgunFunctionalizeR software (Kristiansson et al. 2009) with the default settings. Functional genes in the Arctic fosmids were first grouped into gene families of functionally similar genes identified in the Cluster of Orthologous Groups (COG) data base (www.ncbi.nlm.nih.gov/COG). The COG assignments were used directly in gene-centric analyses and were grouped according to metabolic pathways (Markowitz et al. 2008) for pathway-centric analyses (Kunin et al. 2008).

Proteorhodopsin genes

PCR screening identified 10 fosmids containing PR genes. The Arctic PR genes were 72 to 75% similar to the PR genes in the cultivated *Alphaproteobacteria* strain HTCC2255, the *Bacteroidetes* strain *Polaribacter* sp. MED152 and the *Gammaproteobacteria* *Photobacterium* sp. SKA345. One Arctic PR gene was

most similar (78%) to the PR gene in the HOT2C01 *Alphaproteobacteria* fosmid from the HOT site. Five other Arctic PR genes were most similar to PR genes from the Mediterranean Sea.

Identifying virus DNA in genomes of uncultivated bacteria

The 57 Arctic fosmid sequences, 19 Monterey Bay bacterial artificial chromosome (BAC) sequences (Béjà et al. 2000) and 6 Antarctic BAC sequences (Grzymski et al. 2006) were examined for phage genes using BLAST analyses with viral metagenome sequences obtained from the Arctic Ocean, coastal British Columbia waters and the Gulf of Mexico (Angly et al. 2006). The viral metagenome sequences were first filtered to remove bacterial sequences that were identified as having matches with e-values $<10^{-5}$ to bacterial sequences in the SEED data base. The fosmid and BAC clones were analyzed using the *tblastx* option, and top hits with e-values of $\leq 10^{-5}$ and similarities of $\geq 90\%$ were mapped onto the bacterial DNA. The Arctic, Monterey Bay (Béjà et al. 2000) and Antarctic (Grzymski et al. 2006) environmental bacterial sequences were also screened for phage sequences with the Prophage Finder tool (Bose & Barber 2006) using the default software settings and the viral genomic DNA sequences available from NCBI (ftp.ncbi.nih.gov/refseq/release/viral) in May 2010.

Fluorescence *in situ* hybridization

Bacterial community structure was assessed using fluorescence *in situ* hybridization (FISH) following the procedure outlined by Kirchman et al. (2007). Seawater samples were preserved in paraformaldehyde and filtered through 0.2 μm polycarbonate filters in preparation for the analysis using Cy3-labeled probes (MWG Biotech) to enumerate *Alphaproteobacteria* (Alf968) (Amann et al. 1990), *Betaproteobacteria* (Bet42a) (Glöckner et al. 1999), *Gammaproteobacteria* (Gam42a) (Glöckner et al. 1999), the *Bacteroidetes* group (CF319a) (Manz et al. 1996), *Actinobacteria* (HGC69a) (Roller et al. 1994) and all bacteria (EUB338) (Amann et al. 1990). A negative control probe (Karner & Fuhrman 1997) was used to assess non-specific probe binding. Microscopic analysis was done with a semi-automated image analysis system coupled to an Olympus epifluorescence microscope (Cottrell & Kirchman 2003).

RESULTS

The phylogenetic structure of the bacterial community sampled by the fosmid library was determined using 16S rRNA genes carried by 34 fosmid clones. The majority of the Arctic fosmids contained DNA from *Bacteroidetes* and *Gammaproteobacteria* (Table 1); 13 and nine 16S rRNA gene-containing fosmids contained DNA from *Bacteroidetes* and *Gammaproteobacteria*, respectively. The rest of the Arctic fosmids were from *Alphaproteobacteria*, *Betaproteobacteria*, *Actinobacteria* and eukaryotic algae (Table 1).

The phylogenetic composition of the 16S rRNA gene-bearing fosmids differed somewhat from that of the original community, as determined by FISH (Table 1). *Alphaproteobacteria* made up 33% of the total prokaryotic community, which was ~2-fold higher than the representation of *Alphaproteobacteria* in the fosmid library. In contrast, *Betaproteobacteria* accounted for 6.5% of the 16S rRNA-bearing clones and 4% of the bacterial community as determined by FISH. The abundance of *Gammaproteobacteria* in the original community as determined by FISH was ~50% lower than in the fosmid library. The percentage of *Bacteroidetes* in the original community was 37%, which is slightly lower than the abundance of clones belonging to this group in the fosmid library. The 16S rRNA genes in the Arctic fosmids closely matched genes from previously identified bacterial taxa (see Table S1 in the supplement at www.int-res.com/articles/suppl/a066p107_supp.pdf).

Frequency of phage genes in Arctic bacterial DNA

The number of phage genes in genomes of uncultivated Arctic bacteria was significantly higher (ANOVA, $p < 0.001$) than the number found in bacterial genomes from Monterey Bay and Antarctic waters (Fig. 1). BLAST analyses using viral metagenome queries revealed ~2-fold more viral genes in Arctic bacterial DNA than in the Monterey Bay bacterial DNA and ~10-fold more than in the Antarctic bacterial DNA (Fig. 1). The abundance of virus DNA detected in bacterial DNA was higher in the Arctic than in bacterial DNA from Monterey Bay or the Antarctic ($p < 0.001$), regardless of which virus metagenome was used in the BLAST analysis.

The Prophage Finder tool, which relies on a data base of complete viral genomes, yielded a different accounting of phage-like gene sequences. According to this tool, the phage genes found in the Arctic fosmids matched the genes from a variety of aquatic and

non-aquatic viruses (see Table S2 in the supplement at www.int-res.com/articles/suppl/a066p107_supp.pdf). However, overall, Prophage Finder revealed no significant difference in the frequencies of prophage genes in the bacterial genomes from the 3 environments (ANOVA, $p > 0.05$). The Prophage Finder tool estimated an average of 26 prophage genes per Mbp of bacterial genomic DNA in the Arctic Ocean, Monterey Bay and Antarctic waters (see Table S3 in the

Table 1. Classification of Arctic fosmids containing 16S rRNA genes. The phylogenetic classification is based on the rRNA genes analyzed using the Greengenes classifier. The phylogenetic structure of the original community ('% of total prokaryotes', mean \pm SE) was determined by fluorescence *in situ* hybridization. % of total: percent of fosmids bearing a bacterial 16S rRNA gene

Classification	No. of fosmids	% of total	% of total prokaryotes
Total	34	100	
<i>Alphaproteobacteria</i>	6	19.4	33 \pm 14
<i>Betaproteobacteria</i>	2	6.5	4 \pm 7
<i>Gammaproteobacteria</i>	9	29.0	13 \pm 3
<i>Bacteroidetes</i>	13	41.9	37 \pm 7
<i>Actinobacteria</i>	1	3.2	6 \pm 3
Eukarya	3		

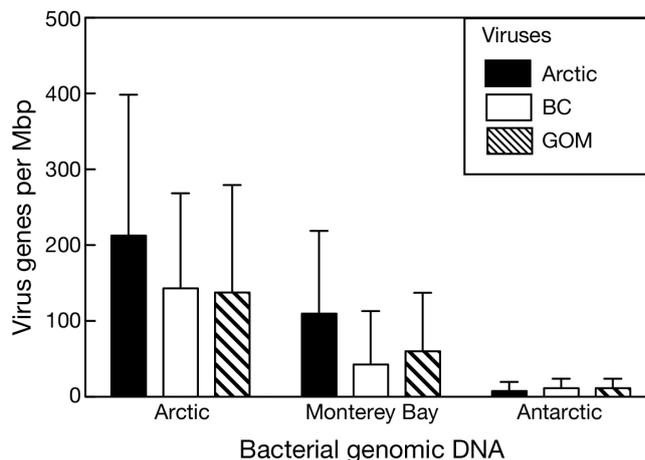


Fig. 1. Frequency of virus genes (per million base pairs, Mbp) in bacterial genomic DNA from Arctic Ocean, Monterey Bay and Antarctic waters. The virus genes were identified by querying the bacterial DNA with metagenomic viral DNA sequences obtained from the Arctic Ocean, British Columbia (BC) coastal water waters and the Gulf of Mexico (GOM) (Angly et al. 2006). The virus genes were enumerated after mapping the viral metagenomic sequences to the bacterial genomic DNA. BLAST hits with e-values < 0.0001 and percent similarities $> 90\%$ were considered significant in the tblastx analysis. The error bars correspond to 1 SD calculated for sample sizes of 58, 14 and 6 clones for the Arctic fosmids, Monterey Bay bacterial artificial chromosome (BAC) and the Antarctic BAC libraries, respectively

supplement). In contrast, the average based on BLAST analyses using the viral metagenome sequences was 82 virus genes per Mbp (Fig. 1), ~4-fold higher than the Prophage Finder estimate.

Most (75%) of the prophage genes identified by Prophage Finder in the Arctic bacterial genomes were similar to viruses from other environments and viruses that infect insects and vertebrates but not to viruses from marine or even aquatic environments. The few aquatic viral sequences identified in the bacterial genomic DNA included *Ostreococcus tauri* virus 1, *Cyanophage* PSS2, *Paramecium bursaria* *Chlorella* virus AR158, *Emiliana huxleyi* virus 86 and *Prochlorococcus* phage P-SSM2.

Genomic context of phage genes in Arctic bacterial genomes

We performed a more detailed analysis of the fosmid sequences by searching the annotation for the

keywords 'phage' and 'virus'. The annotations of 4 such fosmids are depicted in Fig. 2. The phage genes in 3 of the 4 fosmids are flanked by entirely bacterial genes (Fig. 2), confirming that these phage genes occur in bacterial genomes and are not from free phages contaminating the bacterial size fraction. Unclassified and unassigned genes were significantly more abundant in the phage-bearing fosmids than in fosmids without phage DNA (Student's *t*-test, $p < 0.001$). On average, the fosmids with phage DNA contained 6.6 ± 10 unclassified or unassigned genes compared to 2.5 ± 2.4 genes in fosmids without phage DNA.

Fosmid EU795240 was the only fosmid apparently containing a fragment of viral genomic DNA. None of the genes in Fosmid EU795240 could be identified, except for 3 phage-like genes coding for a tyrosine recombinase, a phage integrase and a putative bacteriophage protein. No bacterial genes could be identified flanking these viral genes, suggesting that this fosmid contains a fragment of viral genomic

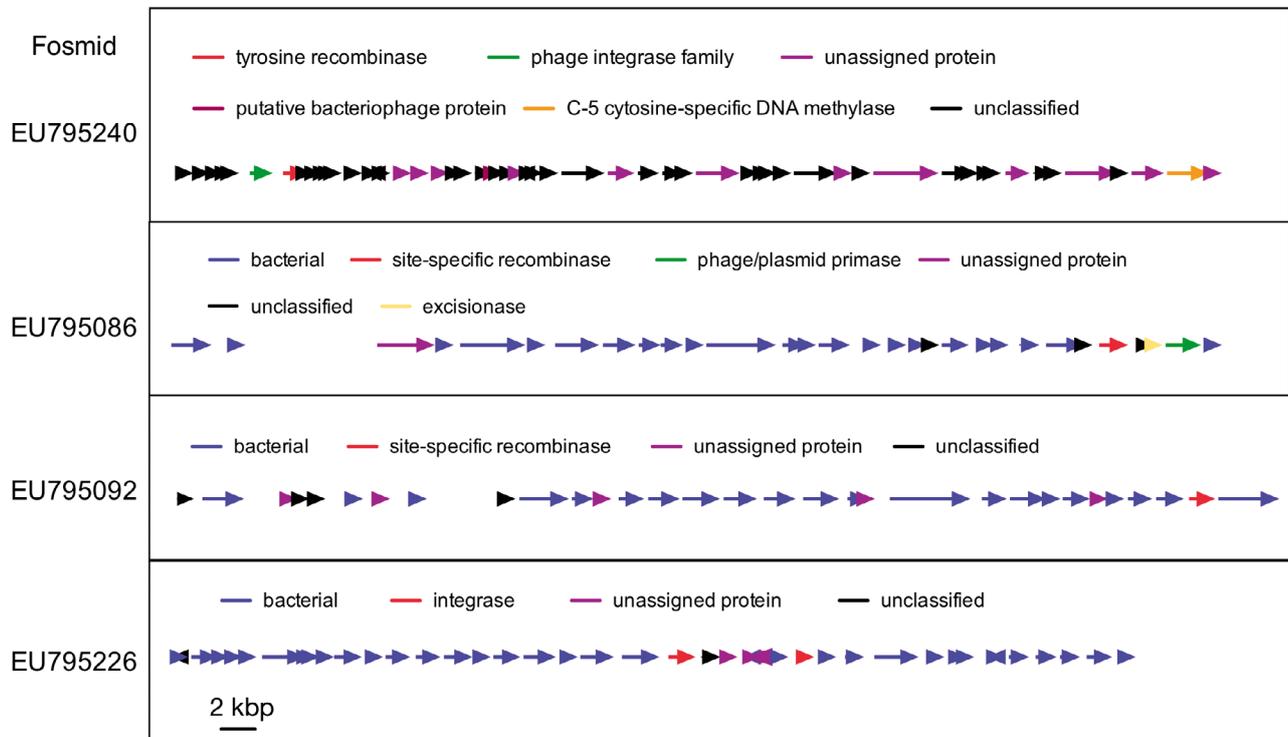


Fig. 2. Gene maps of 4 fosmids (GenBank accession numbers indicated) showing the relationship of virus-related genes to neighboring genes. Open reading frames are depicted as arrows, and the colored legends indicate various gene annotations. Genes not indicated as being 'bacterial' are potentially from viruses. The empty space between genes indicates that the annotator did not report anything for that stretch of DNA. The bulk of the virus-related genes were most similar to integrases or recombinases (see 'Results: Genomic context of phage genes ...' for details). 'Unclassified' indicates no significant BLAST hit to any of the rRNA, UniRef, nr, KEGG, COG, Pfam or Smart data bases. 'Unassigned protein' indicates that none of the significant BLAST hits to the UniRef, KEGG, nr or COG data bases included an informative term in the description, and there was no significant RPS-BLAST hit to the Pfam or Smart data bases (Koski et al. 2005)

DNA (Fig. 2). Aside from the phage genes, 11 genes were classified as encoding unassigned proteins based on BLAST search results yielding no significant hits or a hit to only one of the uniref90, nr, COG and KEGG data bases. In about half of these cases with no consensus among the BLAST results from the different data bases, the BLAST results included hits in a subset of the data bases to genes from phages, such as the bacteriophage ST64T, cyanophage P60, and cyanophage Pf-WMP4. Because it may contain genes from an intact virus, fosmid EU795240 was not included in the analysis of virus DNA in bacterial genomes.

Fosmid EU795226 contained phage genes annotated as integrases, which were adjacent to genes that could not be classified as belonging to bacteria or phages (Fig. 2). However, in 2 of the other fosmids with phage-like genes, the non-bacterial genes were located side by side with bacterial genes (Fig. 2). Fosmid EU795086 contained 2 phage-related genes, which were classified as a phage or plasmid primase and an excisionase. These phage-like genes were immediately flanked by genes that were identified as being bacterial, along with the rest of the 20 other genes on this fosmid. The other phage-containing fosmid (EU795092) contained only a single phage-related gene, which was classified as a site-specific recombinase. All of the other classified genes on this fosmid appeared to be from bacteria and were related to genes found in the alphaproteobacterial strain HTCC2255. In addition to the phage and bacterial genes, fosmid EU795092 included 5 unassigned proteins and 4 genes that were completely unclassified.

Metagenomic comparison of bacteria from the Arctic Ocean versus lower latitude environments

We used fosmid end sequence data to compare the Arctic Ocean metagenome with metagenomes from lower latitude waters where similar experiments using fosmids and random shotgun sequencing have been conducted. The most relevant previous studies were from the coastal Atlantic Ocean sampled near Cape May, NJ (Rusch et al. 2007) and the central North Pacific sampled at the HOT site near Hawaii (DeLong et al. 2006).

The most abundant gene family in the Arctic metagenome was of the large subunit of a phage terminase, which is involved in packaging DNA into phage capsids (Table 2). No phage terminase genes were found in the Cape May or HOT sites. The top 5 most abundant genes in one metagenome were typically present in the other metagenomes, but only 2 gene families, those for permeases and dehydrogenases of short-chain alcohols, were among the 5 most abundant gene families in all 3 environments. Only one other gene family exhibited a distribution as skewed as the phage genes. The family of universal stress proteins, which was highly abundant in the Arctic and Cape May samples, was absent from the HOT sample (Table 2).

In spite of a few highly contrasting gene families, there was little statistical support for any overall difference in the abundance of genes families among the 3 metagenomes. In fact, only the abundance of the phage terminase large subunit gene family was significantly different among the metagenomes from the Arctic Ocean, Cape May and the HOT environ-

Table 2. Abundant gene families in the Arctic (present study), Cape May (Rusch et al. 2007) and the Hawaii Ocean Time-series (HOT) (DeLong et al. 2006) metagenomes. The list includes the 5 most abundant gene families in each metagenome. Only COG0477 and COG1028 were among the 5 most abundant gene families in all 3 metagenomes. The Arctic, Cape May and HOT samples were randomly subsampled to include 3840 sequences each

Gene family	Number of sequences			Description
	Arctic	Cape May	HOT	
COG1783	25	0	0	Phage terminase large subunit
COG0477	18	26	16	Permeases of the major facilitator superfamily
COG1028	18	21	30	Dehydrogenases with different specificities (short-chain alcohol dehydrogenases)
COG0697	17	12	10	Permeases of the drug/metabolite transporter (DMT) superfamily
COG0589	15	2	0	Universal stress protein UspA and related nucleotide-binding proteins
COG0596	13	9	16	Predicted hydrolases or acyltransferases (α/β hydrolase superfamily)
COG1012	12	24	8	NAD-dependent aldehyde dehydrogenases
COG0665	10	17	10	Glycine/b-amino acid oxidases (deaminating)
COG0642	9	18	4	Signal transduction histidine kinase
COG3119	6	8	22	Arylsulfatase A and related enzymes
COG0154	2	4	14	Asp-tRNA ^{Asn} /Glu-tRNA ^{Gln} amidotransferase A subunit and related amidases

ments (Table 3). Similar to the large subunit, the small subunit of the phage terminase was detected in the Arctic Ocean but not in the Cape May and HOT regions, but the difference was not statistically significant ($p > 0.05$, ShotgunFunctionalizeR package). As expected for universal single-copy genes found in all bacteria (Santos & Ochman 2004), the number of *rpoB* RNA polymerase and *gyrB* DNA gyrase genes in these metagenomes (normalized to equal number of sequences) did not differ between the Arctic Ocean and lower latitude environments (Table 3).

We hypothesized that the metabolic pathways would be different in the microbial communities of the cold Arctic Ocean, the oligotrophic Pacific Ocean and the nutrient-rich coastal Atlantic Ocean because the environmental conditions differ so greatly. However, a test of gene categories using the ShotgunFunctionalizeR package revealed no significant differences in the metabolic pathways among the metagenomes from these waters (Table S4 in the supplement at www.int-res.com/articles/suppl/a066p107_supp.pdf).

DISCUSSION

Persistently cold water temperature and low bacterial and viral abundances are expected to limit the impact of viruses on bacteria in Arctic waters. Contact rates between viruses and bacteria are probably low in the Arctic because bacterial and viral abundances are low (Steward et al. 2007, Payet & Suttle 2008). In addition, the elevated viscosity of cold seawater is expected to reduce the contact frequency between hosts and viruses. The viscosity of seawater is 75% higher at polar seawater temperatures compared to that at 20°C, which reduces the diffusivity of viruses and thus the contact rate by ~60% (Murray & Jackson 1992). These

physical and biological conditions are consistent with the estimates of mortality due to viruses in the Arctic Ocean. Rates of viral mortality range from 6 to 15% in various regions of the Chukchi Sea (Steward et al. 1996, 2007), which are among the lowest determined for any aquatic system (Weinbauer 2004). These results along with those from our study present a paradox for the Arctic. The frequency of virus DNA in the Arctic bacterial metagenome was high, suggesting a substantial viral impact, yet conditions for viral infection in the Arctic are poor, and current estimates of viral mortality are low.

A high abundance of temperate phages in Arctic waters could be one resolution to the paradox. Temperate phages leave a longer-lasting imprint on bacteria than virulent phages because temperate virus DNA is integrated into the host, whereas virulent viral DNA typically is not (Ackermann & DuBow 1987). In support of this hypothesis, viral metagenomic analysis revealed a 10-fold higher frequency of integrase genes in the Arctic viral metagenome than in British Columbia and Gulf of Mexico viral metagenomes (McDaniel et al. 2008); integrase genes are indicative of temperate phages (Novick et al. 2010). However, DNA of virulent phages may also become integrated into the host DNA, for example in association with clustered, regularly interspaced short palindromic repeats (Sorek et al. 2008). Other mechanisms for integrating virulent phage DNA might involve incomplete or faulty viral genomes (Villafane 2009). When a defective virus infects a new host, the virus DNA may become integrated into the host chromosome by generalized transduction without forming a prophage (Canchaya et al. 2003). Gene transfer agents (GTAs) have been implicated in lateral gene transfer in aquatic systems (Biers et al. 2008), but no GTA sequences were identified in the Arctic fosmids (data not shown).

Table 3. Number of sequences in select gene families in the Arctic, Cape May and the Hawaii Ocean Time-series (HOT) metagenomes. Statistical significance was determined using the binomial model of the testGeneFamilies function in the ShotgunFunctionalizeR package. The Arctic, Cape May and HOT samples were randomly subsampled to include 3840 sequences each

Gene family	No. of sequences			Statistical significance		
	Arctic	Cape May	HOT	Arctic vs. Cape May	Arctic vs. HOT	Cape May vs. HOT
Phage terminase large subunit	25	0	0	$p < 0.05$	$p < 0.05$	ns
Phage terminase small subunit	5	0	0	ns	ns	ns
<i>rpoB</i> - DNA-directed RNA polymerase	5	9	8	ns	ns	ns
<i>gyrB</i> - DNA gyrase	9	8	13	ns	ns	ns
No hits found ^a	1767	819	1526	$p < 0.05$	ns	$p < 0.05$

^aThe number of sequences that were not identified by BLAST analysis using the COG data base

Distinguishing among the different sources of virus DNA in marine bacterial genomes will be important for obtaining the most complete understanding of the impact of viruses on bacterial processes in the ocean. If virus DNA in bacterial genomes is dominated by prophages representing temperate viruses, then much of the viral impact will be dictated by factors determining whether contacts between phages and hosts lead to phage integration or host lysis as well as the factors triggering prophage induction leading to lysis. In contrast, if most of the virus DNA occurring in bacterial genomes is that of virulent phages that failed to lyse the host, then models that predict rates of viral lysis from contact rates may need to be revised (Murray & Jackson 1992). The Prophage Finder tool has the potential to give much information about the prevalence of temperate virus DNA in the bacterial DNA from the marine environments. However, the utility of that tool is now limited because there appears to be little representation of marine viruses in the Prophage Finder data base. Cultivation and characterization of representative virulent and temperate phages would be one approach to generating the data needed to expand the utility of that approach.

Despite the similarities of perennially cold temperatures and seasonal ice cover at northern and southern polar latitudes, the ecology of viruses in Antarctic marine waters appears to differ from that of viruses in Arctic Ocean waters. Rates of bacterial mortality due to virus infection in Antarctic waters are among the highest measured in any aquatic system, equivalent to half to nearly all of the bacterial biomass production (Guixa-Boixereu et al. 2002). In addition, viruses respond rapidly to iron-induced phytoplankton blooms in the Southern Ocean, accounting for essentially all of the bacterial mortality associated with the bloom (Weinbauer et al. 2009). Additional data will be needed to fully describe differences in the relationship between viruses and their bacterial hosts at northern and southern high latitudes, but the evidence points to there being more differences than similarities.

Much like the bacterial communities in other oceanic regions (Pedrós-Alió 2006), the bacterial community in the Arctic is dominated by a few abundant types and many rare ones; the 100 most abundant ribotypes account for ~80% of the total community consisting of 3000 ribotypes (Kirchman et al. 2010). The level of sampling in the present study likely accessed a large fraction of the most abundant types of bacteria in the Arctic Ocean. Half of the 34 fosmid bearing 16S rRNA genes matched the 16S

rRNA gene sequences of the 100 most abundant ribotypes. Unfortunately, one of the most abundant taxa, the SAR11 clade, was probably underrepresented in our data because these bacteria appear to be inefficiently sampled by fosmid cloning for unknown reasons (Feingersch & Béjà 2009). SAR11 made up 12% of the community in the Arctic (Kirchman et al. 2010). However, our fosmid-based approach seems to have sampled the other major ribotypes in the community identified by FISH and many of the abundant Arctic ribotypes, including *Polaribacter*, based on sequence matches between fosmids bearing 16S rRNA genes and the 16S rRNA genes in these waters identified by tag pyrosequencing (Kirchman et al. 2010). Novel approaches using single-cell PCR are likely to provide more insights into viruses infecting abundant host cells (Tadmor et al. 2011).

Examining viral genes in rare taxa would require a much larger sample size and a different approach, such as the high throughput and long read sequencing capabilities expected to become available soon (Eid et al. 2009). Because the frequency of contacts between viruses and their bacterial hosts is density dependent, we expect less of an impact of viruses on rare bacteria and a greater impact on abundant bacteria, similar to that proposed by the 'kill the winner' hypothesis (Thingstad 2000, Sandaa et al. 2009).

Overall, the results of the present study support the hypothesis that viruses have a substantial impact on prokaryotic communities in the Arctic Ocean, even though potential contacts between viruses and bacteria are low in these waters. The interactions that do occur appear to leave many virus genes in the genomes of their bacterial hosts. Taken together, the large amounts of virus DNA found in the Arctic bacterial metagenome and the high abundance of temperate phage genes in the Arctic viral metagenome (Angly et al. 2006) suggest that relationships between viruses and hosts are different in this high latitude environment than elsewhere. The Arctic is a prime example of an environment where viruses play dual roles as agents of bacterial mortality and instruments of genetic exchange. Closer examination of the evolutionary and metabolic consequences of genetic exchange mediated by viruses may help identify the full impact of viral genes carried by bacteria in the Arctic Ocean.

Acknowledgements. We thank Liying Yu for technical assistance and colleagues in the Shelf Basin Interaction Program for help in sampling. Grants from the US National Science Foundation (OPP 0124733 and 0806295) and the Marine Microbial Genome Sequencing Project of the Gordon and Betty Moore Foundation supported this work.

LITERATURE CITED

- Ackermann HW, DuBow MS (1987) Viruses of prokaryotes, Vol. I. General properties of bacteriophages. CRC Press, Boca Raton, FL
- Alonso-Sáez L, Sanchez O, Gasol JM, Balague V, Pedrós-Alió C (2008) Winter-to-summer changes in the composition and single-cell activity of near-surface arctic prokaryotes. *Environ Microbiol* 10:2444–2454
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410
- Amann RI, Binder BJ, Olson RJ, Chisholm SW, Devereux R, Stahl DA (1990) Combination of 16S ribosomal-RNA-targeted oligonucleotide probes with flow-cytometry for analyzing mixed microbial populations. *Appl Environ Microbiol* 56:1919–1925
- Angly FE, Felts B, Breitbart M, Salamon P and others (2006) The marine viromes of four oceanic regions. *PLoS Biol* 4:e368
- Béjà O, Suzuki MT, Koonin EV, Aravind L and others (2000) Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. *Environ Microbiol* 2:516–529
- Biers EJ, Wang K, Pennington C, Belas R, Chen F, Moran MA (2008) Occurrence and expression of gene transfer agent genes in marine bacterioplankton. *Appl Environ Microbiol* 74:2933–2939
- Bose M, Barber RD (2006) Prophage finder: a prophage loci prediction tool for prokaryotic genome sequences. *In Silico Biol* 6:223–227
- Canchaya C, Fournous G, Chibani-Chennoufi S, Dillmann ML, Brussa H (2003) Phage as agents of lateral gene transfer. *Curr Opin Microbiol* 6:417–424
- Connelly TL, Tilburg CM, Yager PL (2006) Evidence for psychrophiles outnumbering psychrotolerant marine bacteria in the springtime coastal Arctic. *Limnol Oceanogr* 51:1205–1210
- Cottrell MT, Kirchman DL (2003) Contribution of major bacterial groups to bacterial biomass production (thymidine and leucine incorporation) in the Delaware estuary. *Limnol Oceanogr* 48:168–178
- Cottrell MT, Yu LY, Kirchman DL (2005) Sequence and expression analyses of *Cytophaga*-like hydrolases in a western Arctic metagenomic library and the Sargasso Sea. *Appl Environ Microbiol* 71:8506–8513
- DeLong EF, Preston CM, Mincer T, Rich V and others (2006) Community genomics among stratified microbial assemblages in the ocean's interior. *Science* 311:496–503
- DeSantis TZ, Hugenholtz P, Larsen N, Rojas M and others (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* 72:5069–5072
- Eid J, Fehr A, Gray J, Luong K and others (2009) Real-time DNA sequencing from single polymerase molecules. *Science* 323:133–138
- Feingersch R, Béjà O (2009) Bias in assessments of marine SAR11 biodiversity in environmental fosmid and BAC libraries? *ISME J* 3:1117–1119
- Glöckner FO, Fuchs BM, Amann R (1999) Bacterioplankton compositions of lakes and oceans: a first comparison based on fluorescence *in situ* hybridization. *Appl Environ Microbiol* 65:3721–3726
- Grzymalski JJ, Carter BJ, DeLong EF, Feldman RA, Ghadiri A, Murray AE (2006) Comparative genomics of DNA fragments from six Antarctic marine planktonic bacteria. *Appl Environ Microbiol* 72:1532–1541
- Guixa-Boixereu N, Vaque D, Gasol JM, Sanchez-Camara J, Pedrós-Alió C (2002) Viral distribution and activity in Antarctic waters. *Deep-Sea Res II* 49:827–845
- Jiang SC, Paul JH (1994) Seasonal and diel abundance of viruses and occurrence of lysogeny/bacteriocinogeny in the marine environment. *Mar Ecol Prog Ser* 104:163–172
- Jiang SC, Paul JH (1998) Gene transfer by transduction in the marine environment. *Appl Environ Microbiol* 64:2780–2787
- Karner M, Fuhrman JA (1997) Determination of active marine bacterioplankton: a comparison of universal 16S rRNA probes, autoradiography, and nucleoid staining. *Appl Environ Microbiol* 63:1208–1213
- Kirchman DL, Elifantz H, Dittel AI, Malmstrom RR, Cottrell MT (2007) Standing stocks and activity of Archaea and Bacteria in the western Arctic Ocean. *Limnol Oceanogr* 52:495–507
- Kirchman DL, Moran XAG, Ducklow H (2009) Microbial growth in the polar oceans—role of temperature and potential impact of climate change. *Nat Rev Microbiol* 7:451–459
- Kirchman DL, Cottrell MT, Lovejoy C (2010) The structure of bacterial communities in the western Arctic Ocean as revealed by pyrosequencing of 16S rRNA genes. *Environ Microbiol* 12:1132–1143
- Koski LB, Gray MW, Lang BF, Burger G (2005) AutoFACT: an automatic functional annotation and classification tool. *BMC Bioinformatics* 6:151
- Kristiansson E, Hugenholtz P, Dalevi D (2009) Shotgunfunctionalizer: an R-package for functional comparison of metagenomes. *Bioinformatics* 25:2737–2738
- Kunin V, Copeland A, Lapidus A, Mavromatis K, Hugenholtz P (2008) A bioinformatician's guide to metagenomics. *Microbiol Mol Biol Rev* 72:557–578
- Manz W, Amann R, Ludwig W, Vancanneyt M, Schleifer KH (1996) Application of a suite of 16S rRNA-specific oligonucleotide probes designed to investigate bacteria of the phylum *Cytophaga-Flavobacter-Bacteroides* in the natural environment. *Microbiology* 142:1097–1106
- Markowitz VM, Ivanova NN, Szeto E, Palaniappan K and others (2008) IMG/M: a data management and analysis system for metagenomes. *Nucleic Acids Res* 36:D534–D538
- McDaniel L, Breitbart M, Mobberley J, Long A, Haynes M, Rohwer F, Paul JH (2008) Metagenomic analysis of lysogeny in Tampa Bay: implications for prophage gene expression. *PLoS ONE* 3:e3263
- Murray AG, Jackson GA (1992) Viral dynamics: a model of the effects of size, shape, motion and abundance of single-celled planktonic organisms and other particles. *Mar Ecol Prog Ser* 89:103–116
- Muyzer G, Teske A, Wirsén CO, Jannasch HW (1995) Phylogenetic relationships of *Thiomicrospira* species and their identification in deep-sea hydrothermal vent samples by denaturing gradient gel-electrophoresis of 16S rDNA fragments. *Arch Microbiol* 164:165–172
- Novick RP, Christie GE, Penades JR (2010) The phage-related chromosomal islands of Gram-positive bacteria. *Nat Rev Microbiol* 8:541–551
- Paul JH (2008) Prophages in marine bacteria: dangerous molecular time bombs or the key to survival in the seas? *ISME J* 2:579–589
- Payet JP, Suttle CA (2008) Physical and biological correlates of virus dynamics in the southern Beaufort Sea and Amundsen Gulf. *J Mar Syst* 74:933–945

- Pedrós-Alió C (2006) Marine microbial diversity: Can it be determined? *Trends Microbiol* 14:257–263
- Roller C, Wagner M, Amann R, Ludwig W, Schleifer KH (1994) *In situ* probing of Gram-positive bacteria with high DNA G+C content using 23S rRNA-targeted oligonucleotides. *Microbiology* 140:2849–2858
- Rusch DB, Halpern AL, Sutton G, Heidelberg KB and others (2007) The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol* 5:e77
- Sabehi G, Loy A, Jung KH, Partha R and others (2005) New insights into metabolic properties of marine bacteria encoding proteorhodopsins. *PLoS Biol* 3:e273
- Sandaa RA, Gomez-Consarnau L, Pinhassi J, Riemann L and others (2009) Viral control of bacterial biodiversity — evidence from a nutrient-enriched marine mesocosm experiment. *Environ Microbiol* 11:2585–2597
- Santos SR, Ochman H (2004) Identification and phylogenetic sorting of bacterial lineages with universally conserved genes and proteins. *Environ Microbiol* 6: 754–759
- Shi Y, Tyson GW, Eppley JM, DeLong EF (2011) Integrated metatranscriptomic and metagenomic analyses of stratified microbial assemblages in the open ocean. *ISME J* 5: 999–1013
- Sorek R, Kunin V, Hugenholtz P (2008) Crispr — a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat Rev Microbiol* 6: 181–186
- Steward GF, Smith DC, Azam F (1996) Abundance and production of bacteria and viruses in the Bering and Chukchi Seas. *Mar Ecol Prog Ser* 131:287–300
- Steward GF, Fandino LB, Hollibaugh JT, Whitley TE, Azam F (2007) Microbial biomass and viral infections of heterotrophic prokaryotes in the sub-surface layer of the central Arctic Ocean. *Deep-Sea Res I* 54:1744–1757
- Tadmor AD, Ottesen EA, Leadbetter JR, Phillips R (2011) Probing individual environmental bacteria for viruses by using microfluidic digital PCR. *Science* 333:58–62
- 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
- Villafane R (2009) Construction of phage mutants. In: Clokie MRJ, Kropinski AM (eds) *Bacteriophages: methods and protocols, Vol 1: isolation, characterization, and interactions*. Humana Press, New York, NY, p 105–127
- Weinbauer MG (2004) Ecology of prokaryotic viruses. *FEMS Microbiol Rev* 28:127–181
- Weinbauer MG, Arrieta JM, Griebler C, Herndl GJ (2009) Enhanced viral production and infection of bacterioplankton during an iron-induced phytoplankton bloom in the southern ocean. *Limnol Oceanogr* 54:774–784
- Wells LE, Deming JW (2006) Significance of bacterivory and viral lysis in bottom waters of Franklin Bay, Canadian Arctic, during winter. *Aquat Microb Ecol* 43:209–221
- Wells LE, Cordray M, Bowerman S, Miller LA, Vincent WF, Deming JW (2006) Archaea in particle-rich waters of the Beaufort Shelf and Franklin Bay, Canadian Arctic: Clues to an allochthonous origin? *Limnol Oceanogr* 51:47–59
- Williamson SJ, Rusch DB, Yooseph S, Halpern AL and others (2008) The Sorcerer II Global Ocean Sampling expedition: metagenomic characterization of viruses within aquatic microbial samples. *PLoS ONE* 3:e1456

Editorial responsibility: Curtis Suttle, Vancouver, Canada

*Submitted: October 27, 2011; Accepted: April 5, 2012
Proofs received from author(s): May 10, 2012*