

## 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

\*Corresponding author. Email: kirchman@udel.edu

*Aquatic Microbial Ecology* 66: 107–116 (2012)

**Supplement.** Additional information about the fosmid sequence data, including the similarity of 16S rRNA genes to previously identified sequences (Table S1), BLAST hits for sequences identified by Prophage Finder (Table S2), the frequency of sequences identified by Prophage Finder (Table S3), and metabolic pathways identified in the Arctic fosmid end sequences (Table S4)

Table S1. Percent similarity of 16S rRNA genes in the Arctic fosmids to their closest matching sequences in the Greengenes database

Fosmid	Classification	Closest match	Similarity (%)
EU795101	<i>Actinobacteria</i>	<i>Microsphaeraceae</i>	92.4
EU795092	<i>Alphaproteobacteria</i>	Clone pLW-38	89.5
EU795087	<i>Alphaproteobacteria</i>	<i>Rhodobacter</i>	96.7
EU795106	<i>Alphaproteobacteria</i>	<i>Rhodobacter</i>	96.7
EU795084	<i>Alphaproteobacteria</i>	<i>Rhodobacter</i>	97.8
EU795102	<i>Alphaproteobacteria</i>	<i>Rhodobacter</i>	97.8
EU795100	<i>Alphaproteobacteria</i>	<i>Rhodobacter</i>	97.9
EU795094	<i>Bacteroidetes</i>	Unclassified	87.3
EU795252	<i>Bacteroidetes</i>	<i>Polaribacter</i>	91.9
EU795086	<i>Bacteroidetes</i>	<i>Psychroserpens burtonensis</i>	92.8
EU795244	<i>Bacteroidetes</i>	<i>Psychroserpens burtonensis</i>	93.3
EU795237	<i>Bacteroidetes</i>	<i>Cytophaga</i>	93.7
EU795097	<i>Bacteroidetes</i>	<i>Psychroserpens burtonensis</i>	94.1
EU795239	<i>Bacteroidetes</i>	<i>Cytophaga</i>	94.6
EU795245	<i>Bacteroidetes</i>	<i>Cytophaga</i>	94.6
EU795246	<i>Bacteroidetes</i>	<i>Algoriphagus</i>	95.5
EU795242	<i>Bacteroidetes</i>	<i>Polaribacter</i>	98.2
EU795085	<i>Bacteroidetes</i>	<i>Polaribacter</i>	98.2
EU795091	<i>Bacteroidetes</i>	<i>Polaribacter</i>	98.7
EU795104	<i>Bacteroidetes</i>	<i>Polaribacter</i>	98.9
EU795249	<i>Betaproteobacteria</i>	<i>Methylotenera</i>	94.8
EU795088	<i>Betaproteobacteria</i>	Clone OM43	95.6

EU795251	<i>Gammaproteobacteria</i>	<i>Cobetia</i>	86.5
EU795250	<i>Gammaproteobacteria</i>	<i>Marinomonas</i>	88.1
EU795253	<i>Gammaproteobacteria</i>	<i>Marinomonas</i>	88.1
EU795103	<i>Gammaproteobacteria</i>	<i>Marinobacterium</i>	90.0
EU795090	<i>Gammaproteobacteria</i>	Clone SUP05	95.6
EU795096	<i>Gammaproteobacteria</i>	Clone SUP05	95.9
EU795247	<i>Gammaproteobacteria</i>	<i>Colwelliaceae</i>	97.7
EU795093	<i>Gammaproteobacteria</i>	Strain HTCC2207	98.1
EU795095	<i>Gammaproteobacteria</i>	Strain HTCC2207	98.6
EU795083	Eukaryote	<i>Mantoniella</i>	98.0
EU795098	Eukaryote	<i>Mantoniella</i>	98.0
EU795254	Eukaryote	Clone DYF25	98.8

Table S2. Frequency of prophages estimated using Prophage Finder in bacterial genomic DNA from Antarctic waters (Grzyski et al. 2006), Monterey Bay (McCarren & DeLong 2007) and the Arctic Ocean (the present study). n: number of fosmids or BACs

Prophage frequency (per Mbp)				Best BLAST hits		
<u>Location</u>	<u>Mean</u>	<u>SD</u>	<u>n</u>	<u>Description</u>	<u>Virus</u>	<u>e-value</u>
Antarctic	27	9.9	6	Hypothetical protein MdSGHV066	<i>Musca domestica</i> salivary gland hypertrophy virus	4.00E-11
				Hypothetical protein AR158_C785L	<i>Paramecium bursaria Chlorella</i> virus AR158	2.00E-18
				Tmp	<i>Listeria</i> phage B054	4.00E-16
				Hypothetical protein AR158_C788L	<i>Paramecium bursaria Chlorella</i> virus AR158	4.00E-17
				Putative GroEL-like chaperonine protein	<i>Pseudomonas</i> phage EL	9.00E-18
				Hypothetical protein OTV1_036	<i>Ostreococcus tauri</i> virus 1	5.00E-10
				Hypothetical protein	<i>Paramecium bursaria Chlorella</i> virus 1	6.00E-13
				Putative dUTPase	<i>Xanthomonas</i> phage OP2	3.00E-47
Monterey Bay	22	13	18	Hypothetical protein TNAV2c_gp071	<i>Trichoplusia ni</i> ascovirus 2c	0.006
				Prenyltransferase	African swine fever virus	3.00E-05
				Putative serine palmitoyltransferase	<i>Emiliana huxleyi</i> virus 86	2.00E-32
				HSP70	<i>Acanthamoeba polyphaga</i> mimivirus	1.00E-154
				Putative helicase	<i>Escherichia</i> phage rv5	4.00E-42
				MazG	<i>Prochlorococcus</i> phage P-SSM2	1.00E-28
				Putative DNA-dependent RNA polymerase II largest subunit	<i>Emiliana huxleyi</i> virus 86	3.00E-33

Tmp	<i>Listeria</i> phage B054	1.00E-18
Ban	<i>Enterobacteria</i> phage P1	1.00E-109
Hypothetical protein ATCV1_Z679L	<i>Acanthocystis turfacea</i> <i>Chlorella</i> virus 1	9.00E-09
Hypothetical protein TNAV2c_gp071	<i>Trichoplusia ni</i> ascovirus 2c	5.00E-24
Hypothetical protein TNAV2c_gp071	<i>Trichoplusia ni</i> ascovirus 2c	4.00E-07
Hypothetical protein MIMI_L532	<i>Acanthamoeba polyphaga</i> mimivirus	5.00E-17
IMPB/MUCB/SAMB family protein	<i>Bacillus</i> phage SPBc2	1.00E-07
Hypothetical protein MIMI_L532	<i>Acanthamoeba polyphaga</i> mimivirus	1.00E-14
Hypothetical protein RB43ORF157c	<i>Enterobacteria</i> phage RB43	1.00E-44
HSP70	<i>Acanthamoeba polyphaga</i> mimivirus	1.00E-138
Putative cysteine synthase	<i>Lactococcus</i> phage P087	4.00E-38
Putative DNA-dependent RNA polymerase II largest subunit	<i>Emiliana huxleyi</i> virus 86	3.00E-33
Hypothetical protein AR158_C785L	<i>Paramecium bursaria</i> <i>Chlorella</i> virus AR158	5.00E-32
Putative DNA methylase	<i>Streptococcus</i> phage 5093	2.00E-19
Hypothetical protein FR483_N670L	<i>Paramecium bursaria</i> <i>Chlorella</i> virus FR483	4.00E-18
Envelope glycoprotein J	Equid herpesvirus 9	0.94
S-adenosylmethionine decarboxylase	<i>Synechococcus</i> phage S-PM2	1.00E-15
HSP70	<i>Acanthamoeba polyphaga</i> mimivirus	1.00E-138

Arctic Ocean	29	17	61	HSP70	<i>Acanthamoeba polyphaga</i> mimivirus	1.00E-141
				Hypothetical protein MIMI_R290	<i>Acanthamoeba polyphaga</i> mimivirus	3.00E-77
				N-acetylglucosamine-1-phosphate uridyltransferase	<i>Acanthamoeba polyphaga</i> mimivirus	3.00E-26
				Putative N-acetylmuramoyl-L-alanine amidase	<i>Clostridium</i> phage c-st	6.00E-20
				Hypothetical protein OTV1_038	<i>Ostreococcus tauri</i> virus 1	1.00E-27
				Hypothetical protein TNAV2c_gp071	<i>Trichoplusia ni</i> ascovirus 2c	4.00E-15
				Lon domain protease	<i>Acanthamoeba polyphaga</i> mimivirus	4.00E-20
				Integrase	<i>Salmonella</i> phage c341	2.00E-29
				Fiber protein	Cyanophage PSS2	1.00E-33
				DNA polymerase III alpha subunit	<i>Bacillus</i> phage 0305phi8-36	4.00E-95
				gp58	<i>Burkholderia</i> phage phi1026b	2.00E-05
				Hypothetical protein AR158_C785L	<i>Paramecium bursaria</i> <i>Chlorella</i> virus AR158	1.00E-17
				Putative ATP-binding cassette transporter	<i>Amsacta moorei</i> entomopoxvirus 'L'	2.00E-06
				Putative transglycosylase	<i>Bacillus</i> phage SPBc2	2.00E-16
				Putative acetyl transferase	<i>Pseudomonas</i> phage PAJU2	3.00E-73
				p081	<i>Rhizobium</i> phage 16-3	3.00E-72
				DNA primase	Cyanophage PSS2	1.00E-30
				Putative thymidylate synthase	<i>Enterobacteria</i> phage T5	1.00E-72
				Putative HD phosphohydrolase	<i>Acanthamoeba polyphaga</i> mimivirus	8.00E-19

NudE nudix hydrolase	<i>Enterobacteria</i> phage RB69	0.007
Hypothetical protein	<i>Paramecium bursaria Chlorella</i> virus 1	3.00E-14
Hypothetical protein	<i>Paramecium bursaria Chlorella</i> virus 1	1.00E-04
Putative phosphoribosyl formylglycinamide (FGAM) synthase II	<i>Microbacterium</i> phage Min1	1.00E-142
Glutamine synthetase (glutamate-aminon ligase)	<i>Acanthamoeba polyphaga</i> mimivirus	1.00E-71
Putative triacylglycerol lipase	<i>Acanthamoeba polyphaga</i> mimivirus	5.00E-23
Hypothetical protein OsV5_037f	<i>Ostreococcus</i> virus OsV5	1.00E-40
ORF MSV152 putative core protein P4a homolog (vaccinia A10L), similar to SW:P33817	<i>Melanoplus sanguinipes</i> entomopoxvirus	1.9
Putative fatty acid desaturase	<i>Emiliana huxleyi</i> virus 86	0.002
DNA polymerase I	<i>Thermus</i> phage P74-26	8.00E-49
tRNA (Uracil-5-)-methyltransferase	<i>Acanthamoeba polyphaga</i> mimivirus	3.00E-25
Putative HD phosphohydrolase	<i>Acanthamoeba polyphaga</i> mimivirus	4.00E-18
Acetylcholinesterase	<i>Acanthamoeba polyphaga</i> mimivirus	1.00E-30
Csp	<i>Lactococcus</i> phage bIL312	5.00E-10
Putative membrane protein	<i>Emiliana huxleyi</i> virus 86	0.024
Hypothetical protein NY2A_B018L	<i>Paramecium bursaria Chlorella</i> virus NY2A	1.00E-29
Putative ribonuclease	<i>Emiliana huxleyi</i> virus 86	1.00E-22

RNA polymerase beta subunit	Lymphocystis disease virus - isolate China	3.00E-28
Translation elongation factor EF-1alpha	Marseillevirus	1.00E-12
Putative HD phosphohydrolase	<i>Acanthamoeba polyphaga</i> mimivirus	1.00E-20
Dihydrofolate reductase	<i>Pseudomonas</i> phage 201phi2-1	7.00E-20
Nuclear protein UL24	Papiine herpesvirus 2	1.2
T87	Tupaiid herpesvirus 1	0.15
HSP70	<i>Acanthamoeba polyphaga</i> mimivirus	1.00E-134
Hypothetical protein ATCV1_Z077L	<i>Acanthocystis turfacea</i> <i>Chlorella</i> virus 1	3.00E-07
gp58	<i>Burkholderia</i> phage phi1026b	8.00E-06
gp183	<i>Mycobacterium</i> phage Myrna	5.00E-07
Hypothetical protein ATCV1_Z544R	<i>Acanthocystis turfacea</i> <i>Chlorella</i> virus 1	4.00E-53
DNA directed RNA polymerase subunit 1	<i>Acanthamoeba polyphaga</i> mimivirus	3.00E-22
Putative DNA gyrase A subunit	<i>Clostridium</i> phage c-st	2.00E-71
Putative ATP-binding cassette transporter	<i>Amsacta moorei</i> entomopoxvirus 'L'	4.00E-10
Hypothetical protein ATCV1_Z804L	<i>Acanthocystis turfacea</i> <i>Chlorella</i> virus 1	1.00E-122
Putative cysteine synthase	<i>Lactococcus</i> phage P087	3.00E-62
Large tegument protein	Cercopithecine herpesvirus 2	1.1
gp118	<i>Mycobacterium</i> phage Kostya	3.00E-90

tRNA (Uracil-5-)-methyltransferase	<i>Acanthamoeba polyphaga</i> mimivirus	2.00E-24
GtrB; bactoprenol glucosyl transferase	<i>Salmonella</i> phage epsilon34	2.00E-21
Hypothetical protein ATCV1_Z169R	<i>Acanthocystis turfacea</i> <i>Chlorella</i> virus 1	5.00E-09
Putative plasmid partitioning protein Soj	<i>Natrialba</i> phage PhiCh1	7.00E-18
RNAse III (dsRNA binding motif)	<i>Acanthamoeba polyphaga</i> mimivirus	1.00E-15
Putative 2-component sensor protein YehU	<i>Enterobacteria</i> phage YYZ-2008	4.00E-19
Glutamine synthetase (glutamate-ammonia ligase)	<i>Acanthamoeba polyphaga</i> mimivirus	1.00E-71
Hypothetical protein FR483_N733R	<i>Paramecium bursaria</i> <i>Chlorella</i> virus FR483	1.00E-16
PurH	<i>Prochlorococcus</i> phage P-SSM2	5.00E-64

Table S3. Frequency of prophages estimated using Prophage Finder in bacterial genomic DNA from Antarctic waters (Grzymalski et al. 2006), Monterey Bay (McCarren & DeLong 2007) and the Arctic Ocean (the present study). SD: standard deviation

<u>Location</u>	Prophage frequency (per Mbp)		BLAST e-value		<u>n</u>
	<u>Mean</u>	<u>SD</u>	<u>Mean</u>	<u>SD</u>	
Antarctic	27	0.99	$2.2 \times 10^{-16}$	$1.8 \times 10^{-10}$	6
Monterey Bay	22	1.3	$5.0 \times 10^{-24}$	$1.9 \times 10^{-2}$	18
Arctic Ocean	29	1.7	$2.0 \times 10^{-21}$	$3.1 \times 10^{-1}$	61

Table S4. Metabolic pathways (as defined in the COG database) identified in the Arctic fosmid end sequences. There was no statistically significant difference in the metabolic pathways present in the Arctic versus Cape May and Arctic versus HOT metagenomes based on the gene categories test of the ShotgunFunctionalizeR package. A total of 33% of the fosmid end sequences could be assigned to a metabolic pathway

<u>Metabolic pathway</u>	<u>% of end sequences</u>
Amino acid transport and metabolism	11.5
Energy production and conversion	11.3
Carbohydrate transport and metabolism	8.9
General function prediction only	8.5
Translation, ribosomal structure and biogenesis	8.5
Replication, recombination and repair	7.7
Cell wall, membrane and envelope biogenesis	5.5
Lipid transport and metabolism	5.5
Inorganic ion transport and metabolism	5.1
Transcription	4.9
Posttranslational modification and protein turnover	4.7
Nucleotide transport and metabolism	4.5
Coenzyme transport and metabolism	4.2
Secondary metabolites biosynthesis and transport	2.2
Defense mechanisms	2.2
Signal transduction mechanisms	1.4
Intracellular trafficking and secretion	1.4
Cell cycle control and cell division	0.6
Cell motility	0.6
Function unknown	0.6
Cytoskeleton	0.2
Total end sequences =	1265

### **LITERATURE CITED**

- Grzymiski 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
- McCarren J, DeLong EF (2007) Proteorhodopsin photosystem gene clusters exhibit co-evolutionary trends and shared ancestry among diverse marine microbial phyla. *Environ Microbiol* 9:846–858