

Accession numbers for genetic sequences of virus-enriched field samples and *P. carterae* viruses from laboratory cultures at Bigelow Laboratory for Ocean Sciences, Maine from 2015-2016

Website: <https://www.bco-dmo.org/dataset/670912>

Data Type: experimental

Version:

Version Date: 2016-12-21

Project

» [Persistent Virus Infections in Marine Phytoplankton](#) (Marine Chronic Viruses)

Contributors	Affiliation	Role
Martínez Martínez, Joaquín	Bigelow Laboratory for Ocean Sciences	Principal Investigator, Contact
York, Amber D.	Woods Hole Oceanographic Institution (WHOI BCO-DMO)	BCO-DMO Data Manager

Table of Contents

- [Dataset Description](#)
 - [Methods & Sampling](#)
 - [Data Processing Description](#)
- [Data Files](#)
- [Parameters](#)
- [Instruments](#)
- [Deployments](#)
- [Project Information](#)
- [Funding](#)

Dataset Description

This dataset contains assembled metagenomic contig information from viruses co-infecting a *P. carterae* CCMP 645 laboratory cultures and virus-enriched metagenomics contigs from < 0.45 um-filtered seawater samples collected at the Bigelow Laboratory's dock, at the Damariscotta River Estuary, Maine. Included in this dataset are accession identifiers for the National Center for Biotechnology Information (NCBI) where the sequence data is stored.

When data is unrestricted an update to the data will be made to provide direct links to accession information at NCBI.

Related datasets:

- * [Pleurochrysis carterae growth](#)
- * [Pleurochrysis carterae virus production](#)
- * [Virus dPCR assay primers](#)
- * [TEM Pleurochrysis carterae thin section images](#)
- * [TEM Pleurochrysis carterae virion images](#)

Methods & Sampling

Laboratory cultured *P. carterae* CCMP 645 viruses

A 2 L culture of *P. carterae* CCMP 645 in exponential growth phase, grown in F/2 medium was filtered through a 0.2 um PES membrane under sterile conditions. The filtrate was then initially concentrated down to approximately 25 ml by tangential flow filtration using a Vivaflow 50 cartridge (Sartorius) and further concentrated down to approximately 1 ml using a 30 kDa MWCO AmiconUltra-15 column (Millipore).

DNA from the viral concentrate was extracted using the QiaAMP MinElute Virus spin kit. DNA was tested with universal 16S and 18S primer sets and determined to be free of host, or prokaryotic contamination prior to sequencing.

Two independent sequencing libraries were prepared and sequenced by Illumina MiSeq platforms through Bigelow Laboratory for Ocean Sciences' Single Cell Genomic Center (150 bp paired-end reads) and through the Sequencing Facility at the University of Wisconsin-Madison (300 bp paired-end reads). In addition, PCR products generated to join contig ends were pooled and sequenced in 3 batches (MiSeq), also generating 300 bp paired ends through the University of Wisconsin-Madison.

Virus-enriched Bigelow Dock sample metagenome:

Six large, land-based, 2460 dm³ volume mesocosm tanks were filled on September 24, 2015, from a seawater supply at Bigelow Laboratory's deep-water dock site at the Damariscotta River Estuary, Maine. The samples were screened through 3 mm mesh, with care taken to ensure that each tank was filled simultaneously and contained the same starting phytoplankton composition. The mesocosms were aimed for testing the effects of climate change by simulating predicted increasing temperature and pCO₂ environments in the Gulf of Maine over the next several centuries. However, for our study 1 L samples were collected from each of the mesocosms on Day 0 prior to any further experimental manipulation. All six liters were combined and virus-enriched by filtering through a 0.45 µm PES filter to remove most cellular organisms, and concentrated down to 45 ml by tangential flow filtration. Total DNA was extracted from the sample using the MasterPure(TM) Complete DNA and RNA Purification kit (Epicentre) following manufacturer's recommendations. Total DNA was eluted in 20 µl of nuclease-free water. A DNA library was prepared and sequenced by Illumina MiSeq.

Data Processing Description

Each dataset was assembled independently then, co-assembled using Geneious (v. 8.1.6). De novo assembly pipeline was as follows:

- 1) Adapter removal and quality trimming using trimmomatic (1.0.32)
- 2) Custom python script to merge paired reads into one file (fastx.py)
- 3) Normalization using kmernorm (v. 1.0.5)
- 4) Assembly with SPAdes (3.9.0) using the careful parameter and phred offset 33.
- 5) Contigs were imported into Geneious and assembled using the Geneious assembler to help to stitch together contigs from separate sequencing efforts. Raw reads were aligned to the final contigs in geneious using the bwa aligner to validate the accuracy of the assembly.

Laboratory cultured *P. carterae* CCMP 645 viruses:

Contigs were analyzed through the MetaVir web server.

Virus-enriched Bigelow Dock sample metagenome:

Contigs were annotated with GeneMarkS, then checked manually for any missed potential open reading frames using Artemis. ORFs were blasted against nr, with an evalue threshold of E-5. All ORFs were also queried against hhpred for structural similarity and any hits were annotated accordingly.

BCO-DMO Data Manager Processing Notes:

- * Combined field and culture accession datasets preserving relevant data about accession source
- * added a conventional header with dataset name, PI name, version date
- * modified parameter names to conform with BCO-DMO naming conventions
- * blank values replaced with no data value 'nd'

[[table of contents](#) | [back to top](#)]

Data Files

File
accessions.csv (Comma Separated Values (.csv), 3.17 KB) MD5:a8775094174a7a679849cb61914a3e70
Primary data file for dataset ID 670912

[[table of contents](#) | [back to top](#)]

Parameters

Parameter	Description	Units
accession_id	NCBI identifier	unitless
organism_name	description of virus or metagenome sequenced	unitless
location_collected	location of sample collection	unitless
date_collected	date of sample collection in format yyyy-mm-dd	unitless
sequencing_method	Method of sequence generation (e.g. illumina MiSeq)	unitless
analysis_method	description of analysis	unitless
sequence_description	description of sequence indicating type (ssDNA/dsDNA) and completeness of sequence (partial/complete)	unitless

[[table of contents](#) | [back to top](#)]

Instruments

Dataset-specific Instrument Name	Illumina MiSeq
Generic Instrument Name	Automated DNA Sequencer
Generic Instrument Description	A DNA sequencer is an instrument that determines the order of deoxynucleotides in deoxyribonucleic acid sequences.

Dataset-specific Instrument Name	Bio Rad icycler
Generic Instrument Name	Thermal Cycler
Generic Instrument Description	A thermal cycler or "thermocycler" is a general term for a type of laboratory apparatus, commonly used for performing polymerase chain reaction (PCR), that is capable of repeatedly altering and maintaining specific temperatures for defined periods of time. The device has a thermal block with holes where tubes with the PCR reaction mixtures can be inserted. The cycler then raises and lowers the temperature of the block in discrete, pre-programmed steps. They can also be used to facilitate other temperature-sensitive reactions, including restriction enzyme digestion or rapid diagnostics. (adapted from http://serc.carleton.edu/microbelife/research_methods/genomics/pcr.html)

[[table of contents](#) | [back to top](#)]

Deployments

Bigelow_Martinez_2015-2016

Website	https://www.bco-dmo.org/deployment/670437
Platform	lab Bigelow
Start Date	2015-01-01
End Date	2016-12-30
Description	Bigelow Laboratory for Ocean Sciences Methods & Sampling laboratory experiment

[[table of contents](#) | [back to top](#)]

Project Information

Persistent Virus Infections in Marine Phytoplankton (Marine Chronic Viruses)

Description from NSF award abstract:

Viruses are prevalent in every part of the environment of our living planet, and yet our understanding of type, distribution, and function is the least well-known aspect of biodiversity. In recent years we have developed an increased appreciation for the role viruses play in driving host evolution in the environment, but fundamental knowledge about the mechanisms involved remain lacking. Additionally, viruses may influence diversity indirectly through "kill the winner" scenarios, as well as through cell lysis and subsequent release of dissolved nutrients, which facilitate restructuring of microbial communities. The majority of research on marine viruses to date has focused on combinations of acutely susceptible host strains with highly virulent virus isolates. However, it is likely that marine viruses also employ a persistent infection life strategy, arguably preferring it to the more widely recognized lytic cycle. The objective of this project is to demonstrate that persistent virus infections occur in marine phytoplankton, and that these are a crucial component of ocean ecosystem function and a key evolutionary driver in primary producers. Using a range of persistent virus:host systems, this project will investigate:

- 1) how pervasive persistent virus infections are in marine systems; and
- 2) the role of non-coding RNAs in maintaining host:virus symbiosis.

This is a high risk-high pay research as it involves a radically different approach to the analysis of viruses in

marine systems. The investigators plan to apply a suite of molecular (transcriptomics, genomics and development of novel diagnostic markers) techniques to include the analysis of microRNAs to determine the functional importance of persistent viruses in the ocean. The results of this project will be potentially transformative for our understanding of virus-driven phytoplankton evolution and its potential impact on biodiversity in marine phytoplankton, a vital component of the global carbon cycle.

Note: William Wilson (Bigelow Laboratory) was the Former Principal Investigator on this project award.

[[table of contents](#) | [back to top](#)]

Funding

Funding Source	Award
NSF Division of Ocean Sciences (NSF OCE)	OCE-1346272

[[table of contents](#) | [back to top](#)]