ssDNA-dsDNA viromes from the Great Lakes, off Bermuda and the South Pacific (GOV project)

Website: https://www.bco-dmo.org/dataset/715475
Data Type: Cruise Results
Version:
Version Date: 2017-09-21

Project
» Ecological impacts and drivers of double-stranded DNA viral communities in the global oceans (GOV)

Contributors	Affiliation	Role
Sullivan, Matthew	Ohio State University	Principal Investigator
Coleman, Maureen L.	University of Chicago	Co-Principal Investigator
Copley, Nancy	Woods Hole Oceanographic Institution (WHOI BCO-DMO)	BCO-DMO Data Manager

Table of Contents

Coverage

Spatial Extent: **N**:48 **E**:-64 **S**:-5.2529 **W**:-87

Dataset Description

This dataset includes metadata and links to scripts and data for the analyses of double- and single-stranded DNA from viruses collected in the Great Lakes, off Bermuda (N. Atlantic) and off Peru (S. Pacific) used to provide first estimates of the natural abundance of ssDNA viruses.

These data were published in
Roux S, Solonenko NE, Dang VT, Poulos BT, Schwenck SM, Goldsmith DB, Coleman ML, Breitbart M, Sullivan MB. (2016) Towards quantitative viromics for both double-stranded and single-stranded DNA viruses. PeerJ 4:e2777 DOI: 10.7717/peerj.2777

Methods & Sampling

For full methodology, see Roux et al (2016).

Great Lakes methodology:
For each lake, 33 to 45L of water were 0.22 micron m-filtered and viruses were concentrated from the filtrate using iron chloride flocculation followed by storage at 4 degrees C. After resuspension in ascorbic-EDTA buffer (0.1 M EDTA, 0.2 M Mg, 0.2 M ascorbic acid, pH 6.0), viral particles were concentrated using Amicon Ultra 100 kDa centrifugal devices (Millipore), treated with DNase I (100U/mL) followed by the addition of 0.1 M EDTA and 0.1 M EGTA to halt enzyme activity, and extracted with the QIAamp DNA Mini Kit (Qiagen 51304).

Tara Expedition methodology:
20 L of seawater were 0.22 µ m-filtered, and viruses were concentrated from the filtrate using iron chloride flocculation followed by storage at 4°C. After resuspension in ascorbic-EDTA buffer (0.1 M EDTA, 0.2 M Mg, 0.2 M ascorbic acid, pH 6.0), viral particles were concentrated using Amicon Ultra 100 kDa centrifugal devices (Millipore), treated with DNase I (100U/mL) followed by the addition of 0.1 M EDTA and 0.1 M EGTA to halt enzyme activity, and extracted with the QIAamp DNA Mini Kit (Qiagen 51304).

Bermuda BATS methodology:
Approximately 180L of seawater were concentrated using a 100kDa tangential flow filter, 0.22 µ m-filtered, PEG precipitated, cesium chloride purified, and DNA was extracted using formamide.

Data Processing Description

Trimmed and filtered reads were assembled with Spades 3.6.2 with the "sc" and "careful" options. All contigs >500 bp and with at least one complete gene were run through VirSorter in the "virome decontamination" mode as well as a custom detection pipeline for small genomes (see Roux et al., 2016). Contig affiliation was based on best BLAST hit against RefSeqVirus (thresholds: bit score > 50 and e-value < 10 −3).

BCO-DMO Processing:
- created a table with dataset id, description, latitude, longitude, depth and link to the sequence and associated data at iVirus
- added conventional header with dataset name, PI name, version date

[table of contents | back to top]

Data Files

File
ssDNA_dsDNA.csv (Comma Separated Values (.csv), 1.88 KB) MD5:f421d71eefba8f85d326881cfe3aeb63 Primary data file for dataset ID 715475

[table of contents | back to top]

Parameters

Parameter	Description	Units
sample_type	type of analysis performed on the samples	unitless
iVirus_link	hypertext link to the CyVerse iVirus data archive	unitless
iVirus_url	url for link to the CyVerse iVirus data archive	unitless
dataset_id	identifier for dataset	unitless
location	description of sample collection location	unitless
lat	latitude; north is positive	decimal degrees
lon	longitude; east is positive	decimal degrees
depth	sample depth	meters

[table of contents | back to top]

Instruments

Dataset-specific Instrument Name	
Generic Instrument Name	Automated DNA Sequencer
Generic Instrument Description	General term for a laboratory instrument used for deciphering the order of bases in a strand of DNA. Sanger sequencers detect fluorescence from different dyes that are used to identify the A, C, G, and T extension reactions. Contemporary or Pyrosequencer methods are based on detecting the activity of DNA polymerase (a DNA synthesizing enzyme) with another chemoluminescent enzyme. Essentially, the method allows sequencing of a single strand of DNA by synthesizing the complementary strand along it, one base pair at a time, and detecting which base was actually added at each step.

Dataset-specific Instrument Name	Amicon Ultra 100 kDa centrifugal devices (Millipore)
Generic Instrument Name	Centrifuge
Dataset-specific Description	Used to concentrate viral particles
Generic Instrument Description	A machine with a rapidly rotating container that applies centrifugal force to its contents, typically to separate fluids of different densities (e.g., cream from milk) or liquids from solids.

[table of contents | back to top]

Deployments

1311-SV

Website	https://www.bco-dmo.org/deployment/737543
Platform	R/V Lake Guardian
Start Date	2013-03-29
End Date	2013-04-14
Description	Spring survey

[table of contents | back to top]

Project Information

Ecological impacts and drivers of double-stranded DNA viral communities in the global oceans (GOV)

Coverage: Global oceans

NSF abstract:

Ocean microbes produce half of the oxygen that humans breathe and drive much of the energy and nutrient transformations that fuel ocean ecosystems. Viruses of microbes alter these microbial impacts through mortality, moving genes from one organism to another, and reprogramming a host cell's metabolism during infection. However, current understanding of ocean viruses is limited to just a few specific model systems or study sites. This project will produce a foundational dataset - the Global Ocean Virome (GOV) - for contextualizing newly discovered viruses, and use this dataset to evaluate the environmental conditions that structure viral communities or how viral communities change over time and space. Additionally, the project will illuminate 'viral dark matter' by experimentally identifying viral structural proteins and generating and investigating single-cell genomic datasets to link novel and abundant viruses to their host cells. The project will train 6 researchers, as well as lead to curriculum, seminars and a public exhibit at The Wellington School and the Columbus Center Of Science and Industry that will together reach approximately 500 students and 250,000 members of the public.

The GOV dataset is comprised of 104 viral metagenomes from around the world's surface and deep oceans. This project seeks to analyze the GOV to identify and quantify viral populations globally, then evaluate these data for ecological patterns to determine the ecological drivers of surface and deep ocean viral community structure. These patterns and drivers will be interpreted in the context of (i) viral metaproteomic experiments to maximally annotate unknown viral proteins that are structural, (ii) paired microbial sequence datasets (metagenomes and metatranscriptomes) that will enable assessment of how biotic factors impact viral community structure and (iii) single-cell amplified genomes and phageFISH experiments that will enable identification of hosts for abundant and novel viruses identified in the GOV. In total, this project will further optimize recently developed genome- and population-based viral ecology methods to establish a first available global map of surface and deep ocean viral populations from both free viromes and infected microbes. These analyses will help evaluate and establish myriad hypotheses about viral roles in marine microbial ecology and biogeochemistry, and the dataset will be a foundational resource for microbial and viral ecologists to contextualize new viruses, identify viruses in microbial datasets, and explore virus-host interactions both phenomenologically and through ecosystem modeling. The GOV will be made publicly available through the NSF-funded iPlant Cyberinfrastructure and metaVIR (http://metavir-meb.univ-bpclermont.fr/).

Note: The above NSF abstract references iPlant Cyberinfrastructure which no longer exists. The GOV dataset is publicly available through iVirus at the NSF-funded CyVerse Cyberinfrastructure while contextual data are

available through EMBL and PANGEA (see links below).

This NFS-OCE project was funded to analyze the following pre-existing datasets

Raw and processed Global Ocean Viromes (GOV) dataset
Data Access at iVirus: http://datacommons.cyverse.org/browse/iplant/home/shared/iVirus/GOV/
Results Paper:
Roux, S., Brum, J. R., Dutilh, B. E., Sunagawa, S., Duhaime, M. B., … Sullivan, M. B. (2016). Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. Nature, 537(7622), 689–693. doi:10.1038/nature19366

GOV associated microbial metagenomic datasets
Data Access at EMBL: http://www.ebi.ac.uk/ena/data/view/PRJEB7988

GOV associated oceanographic data
Data Access via PANGEA: https://www.pangaea.de/search?ie=UTF-8&env=All&count=10&q=Tara+Oceans+Expedition

Global Ocean Viromes 2 dataset at
iVirus: http://datacommons.cyverse.org/browse/iplant/home/shared/iVirus/GOV2.0
Results Paper:
Gregory, A. C., Zayed, A. A., Conceição-Neto, N., Temperton, B., Bolduc, B., Alberti, A., … Cruaud, C. (2019). Marine DNA Viral Macro- and Microdiversity from Pole to Pole. Cell, 177(5), 1109–1123.e14. doi:10.1016/j.cell.2019.03.040

Project results in repositories other than BCO-DMO

Viral metagenomes from the Eastern Tropical North Pacific, and archaeal viruses
Data Access: http://datacommons.cyverse.org/browse/iplant/home/shared/iVirus/Vik_et_al_2017_data
Results Paper:
Vik DR., Roux S., Brum JR., Bolduc B., Emerson JB., Padilla CC., Stewart FJ., Sullivan MB. 2017. Putative archaeal viruses from the mesopelagic ocean. PeerJ 5:e3428. DOI: 10.7717/peerj.3428.

Viral mock community and metagenomic datasets
Data Access:
http://datacommons.cyverse.org/browse/iplant/home/shared/iVirus/DNA_Viromes_library_comparison.
Results Paper:
Roux S., Solonenko NE., Dang VT., Poulos BT., Schwenck SM., Goldsmith DB., Coleman ML., Breitbart M., Sullivan MB. 2016. Towards quantitative viromics for both double-stranded and single-stranded DNA viruses. PeerJ 4:e2777. DOI: 10.7717/peerj.2777.

Virophage sequences extracted from previously published freshwater microbial metagenomes
Data Access: http://datacommons.cyverse.org/browse/iplant/home/shared/iVirus/Freshwater_virophages

Results Paper:

Roux, S., Chan, L.-K., Egan, R., Malmstrom, R. R., McMahon, K. D., & Sullivan, M. B. (2017). Ecogenomics of virophages and their giant virus hosts assessed through time series metagenomics. Nature Communications, 8(1). doi:10.1038/s41467-017-01086-2

[These virophages were identified in microbial metagenome datasets from North Temperate Lakes NSF-LTER sites]

Metagenomic data Access: http:// genome.jgi.doe.gov/pages/dynamicOrganismDownload.jsf?organism=TroutBogmetagenomicdata
Non-metagenomic data Access:
https://genome.jgi.doe.gov/portal/ComAssLakMendota_FD/ComAssLakMendota_FD.info.html]

Other publications associated with this project

Bolduc, B., Jang, H. B., Doulcier, G., You, Z.-Q., Roux, S., & Sullivan, M. B. (2017). vConTACT: an iVirus tool to classify double-stranded DNA viruses that infect Archaea and Bacteria. PeerJ, 5, e3243. doi:10.7717/peerj.3243

Emerson, J. B., Roux, S., Brum, J. R., Bolduc, B., Woodcroft, B. J., Jang, H. B., … Sullivan, M. B. (2018). Host-linked soil viral ecology along a permafrost thaw gradient. Nature Microbiology, 3(8), 870–880. doi:10.1038/s41564-018-0190-y

Howard-Varona, C., Hargreaves, K. R., Abedon, S. T., & Sullivan, M. B. (2017). Lysogeny in nature: mechanisms, impact and ecology of temperate phages. The ISME Journal, 11(7), 1511–1520. doi:10.1038/ismej.2017.16

Pasulka, A. L., Thamatrakoln, K., Kopf, S. H., Guan, Y., Poulos, B., Moradian, A., … Orphan, V. J. (2017). Interrogating marine virus-host interactions and elemental transfer with BONCAT and nanoSIMS-based methods. Environmental Microbiology, 20(2), 671–692. doi:10.1111/1462-2920.13996

Roux, S., Adriaenssens, E.M., Dutlith, B.E., Koonin, E.V., Kropinski, A.M., Krupovic, M., … Eloe-Fadrosh, E.A. (2018). Minimum Information about an Uncultivated Virus Genome (MIUViG): a community consensus on standards and best practices for describing genome sequences from uncultivated viruses. *Nature Biotechnology. In press.* http://u.osu.edu/viruslab/files/2015/08/JGI-viralStdsPaper-2g6uqiv.pdf

Roux, S., Chan, L.-K., Egan, R., Malmstrom, R.R., McMahon, K.D., Sullivan, M.B. (2017). Ecogenomics of virophages and their giant virus hosts assessed through time series metagenomics. Nature Communications. 8: 858. doi: 10.1038/s41467-017-01086-2

Roux, S., Emerson, J. B., Eloe-Fadrosh, E. A., & Sullivan, M. B. (2017). Benchmarking viromics: an in silico evaluation of metagenome-enabled estimates of viral community composition and diversity. PeerJ, 5, e3817. doi:10.7717/peerj.3817

Stec, K. F., Caputi, L., Buttigieg, P. L., D'Alelio, D., Ibarbalz, F. M., Sullivan, M. B., … Iudicone, D. (2017). Modelling plankton ecosystems in the meta-omics era. Are we ready? Marine Genomics, 32, 1–17. doi:10.1016/j.margen.2017.02.006

[table of contents | back to top]

Funding

Funding Source	Award
NSF Division of Ocean Sciences (NSF OCE)	OCE-1536989

[table of contents | back to top]