NCBI BioSample accession numbers for microbes and viruses collected from samples in Canoe Cove, Nahant, MA during 2010. (Marine Bacterial Viruses project)

Website: https://www.bco-dmo.org/dataset/658586
Data Type: experimental
Version: 1
Version Date: 2016-09-09

Project
» How can bacterial viruses succeed in the marine environment? (Marine Bacterial Viruses)

Contributors	Affiliation	Role
Kelly, Libusha	Yeshiva University	Principal Investigator
Polz, Martin	Massachusetts Institute of Technology (MIT)	Co-Principal Investigator
Ake, Hannah	Woods Hole Oceanographic Institution (WHOI BCO-DMO)	BCO-DMO Data Manager

Abstract
NCBI BioSample accession numbers for microbes and viruses collected from samples in Canoe Cove, Nahant, MA during 2010. (Marine Bacterial Viruses project)

Table of Contents

Dataset Description

This dataset contains NCBI Biosample IDs and associated data. Vibrio and associated phage genomes isolated off the coast of Nahant, MA, USA.

Related Datasets:

Isolation culturing and sequencing of bacteria and viruses

Methods & Sampling

Bacteria and viruses were collected from the littoral marine zone at Canoe Cove, Nahant, MA, USA, on August 22 [ordinal day 222], September 18 [261], and October 13 [286], 2010.

Bacteria were collected using a previously described size-fractionation method[1]. Bacterial strain naming convention is described using the example of 10N.286.54.E5: the first position (here "10N") indicates the year (2010) and location (Nahant) of isolation, the second position (here "286") indicates the ordinal day of isolation, the third position (here "54") is a code representing the size-fraction of origin (0.2um: 45,46,47; 1um: 48,49,50; 5um: 51,52,53; 63um: 54,55,56), and the fourth position is the storage plate well identifier. Multiple codes within the size-fraction identifier reflect independent water samples for the 63um fraction, and independent water sample fractionation series for the other size classes (water sample A: 45,51,54; sample B: 46, 52, 55; sample C: 47, 53, 56).

Bacterial genome libraries were prepared for sequencing using a tagmentation-based approach and 1-2ng input DNA per isolate, as previously described[2]. Genomes were sequenced in multiplexed pools of 50-60 samples per Illumina HiSeq lane. Accession numbers for all bacterial genomes associated with this study are provided in Supplementary Table S1.

Bacterial phylogenetic relationships were determined by extracting ribosomal proteins from 278 genomes with hmmsearch[3] and aligning with MAFFT[4] as described in Hehemann et al. (2016)[5]. These strains were added to the Vibrionaceae ribosomal phylogeny used in Hehemann et al., 2016 and taxonomy was assigned using manual inspection. Full-length *hsp60* sequences were also extracted from these genomes using hmmsearch with default parameters and the Cpn60 hmm (PF00118) from Pfam[6]. The *hsp60* sequences were aligned using the mafft-fftnsi algorithm. Sanger-sequenced *hsp60* fragments from 40 strains lacking genome sequences were added to this alignment using the mafft-fftnsi algorithm with the --addfragments option. The *hsp60* alignment was concatenated to the ribosomal protein alignment and used to create a phylogeny using RAxML under a partitioned general time reversible (GTR) model (options: –q, -m GTRGAMMAX) [7]. SH-like supports were calculated using RAxML.

Viruses were collected using a previously described iron flocculation approach[8], using 4L sample volumes, 0.2um pre-filtration to remove bacteria, 0.2um filters for floc-capture, and oxalate solution for resuspension to maintain virus viability. Isolation of viruses was performed as follows. Iron-oxalate concentrate volumes equivalent to 15mL of seawater were mixed into agar overlays of 1,334 potential host *Vibrio.* The agar overlays were performed by combining 150uL overnight host culture and virus concentrate directly on a bottom agar (1% agar, 5% glycerol, 125mL/L of chitin supplement [40g/L coarsely ground chitin, autoclaved, 0.2um filtered] in 2216MB), directly pipetting 2mL of molten top agar (52 degrees C, 0.4% agar, 5% glycerol, in 2216MB) onto the bottom agar, and rapidly swirling to mix. Plates were incubated for 2 weeks, and plaques were archived for later purification. Sequencing and genome analysis of viruses is described briefly, as follows. High titer lysates of serially purified viruses were concentrated using 30kD centrifugal filter units (Millipore, Amicon Ultra Centrifugal Filters, Ultracel 30K, UFC903024) and washed with 1:100 Marine Broth 2216 to reduce salts for nuclease treatment. Concentrates were brought to approximately 500uL using 1:100 diluted 2216MB and then treated with DNase I and RNase A for 65 min at 37 degrees C to digest unencapsidated nucleic acids. Nuclease treated viral lysates were extracted by addition of 1:10 final volume of SDS mix (0.25M EDTA, 0.5M Tris-HCl (pH9.0), 2.5% sodium dodecyl sulfate), 30 min incubation at 65C; addition of 0.125 volumes 8M potassium acetate, 60 min incubation on ice; addition of 0.5 volumes phenol-chloroform; recovery of nucleic acids from aqueous phase by isopropanol and ethanol precipitation. Genomes were sequenced in multiplexed pools using Illumina MiSeq and HiSeq technologies, assembled using CLC assembly cell, and manually curated to standardize genome start positions for the Caudovirales.

Viral strain naming convention is described using the example of 1.008.O._10N.286.54.E5, with specific identifiers separated by a period. The first position (here "1") represents a unique identifier for each independent plaque isolated from a given host from the initial exposure of a given host to an environmental virus concentrate. The second position (here "008") represents a unique working ID for a host strain. The third position (here "O") indicates a unique sublineage generated from a single plaque during viral serial purification, for example due to the emergence of multiple plaque morphologies. Following the underscore is the full strain ID of the host of isolation, as described above.

Accession numbers for all viral genomes associated with this study are included under NCBI BioProject PRJNA328102.

References:

1. Hunt DE, David LA, Gevers D, Preheim SP, Alm EJ, Polz MF. Resource Partitioning and Sympatric Differentiation among Closely Related Bacterioplankton. Science. 2008;320(5879):1081–1085.

2. Baym M, Kryazhimskiy S, Lieberman TD, Chung H, Desai MM, Kishony R. Inexpensive Multiplexed Library Preparation for Megabase-Sized Genomes. PLoS ONE. 2015 May 22;10(5):e0128036.

3. Eddy SR. Accelerated Profile HMM Searches. PLOS Comput Biol. 2011 Oct 20;7(10):e1002195.

4. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol. 2013 Apr 1;30(4):772–780. PMID: 23329690

5. Hehemann J-H, Arevalo P, Datta M, Yu A, Corzett C, Henschel A, Preheim S, Timberlake S, Alm EJ, Polz MF. Adaptive Radiation by Waves of Gene Transfer Leads to Fine-Scale Resource Partitioning in Marine Microbes. Nat Commun Submitt. 2016;

6. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 2016 Jan 4;44(D1):D279–D285. PMID: 26673716

7. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014 May 1;30(9):1312–1313. PMID: 24451623

8. John SG, Mendez CB, Deng L, Poulos B, Kauffman AKM, Kern S, Brum J, Polz MF, Boyle EA, Sullivan MB. A simple and efficient method for concentration of ocean viruses by chemical flocculation. Environ Microbiol Rep. 2011 Apr 1;3(2):195–202.

Data Processing Description

These are accession numbers for sequenced, assembled, annotated genomes for 282 viruses and 301 Vibrio strains.

Data Management Office Notes:

-Data was initially separated into two files for microbes and viruses. These files were combined.
-All columns which contained no data for either viruses or microbes were removed.
-"nd" was entered into columns with no recorded data.
-Column names were reformatted to comply with BCO-DMO naming standards.

[ ]

Data Files

File
NCBI_IDs.csv(Comma Separated Values (.csv), 108.62 KB) MD5:22a3d34c297f307856603170bde15ca5
Primary data file for dataset ID 658586

[ ]

Parameters

Parameter	Description	Units
organism_type	Type of organism described	unitless
BioProject_ID	The accession number of the BioProject(s) to which the BioSample belongs	unitless
BioSample	Accession number of the BioSample	unitless
accession	The accession number of the sample described.	unitless
sample_name	Sample name in source database.	unitless
organism	Organism associated with sample. Identitified to species when possible.	unitless
tax_ID	Taxonomic ID	unitless
isolate	Identification or description of the specific individual from which this sample was obtained	unitless

[ ]

Deployments

CanoeCove_2010

Website	https://www.bco-dmo.org/deployment/658581
Platform	shoreside Massachusetts
Start Date	2010-08-22
End Date	2010-10-13

[table of contents | back to top]

Project Information

How can bacterial viruses succeed in the marine environment? (Marine Bacterial Viruses)

Coverage: Coastal waters off Nahant, MA

Description from NSF award abstract:
Microbes make up the majority of the biomass in the ocean and viral mortality is one of the main ecological factors determining the diversity, abundance and turnover of microbial taxa. Yet, in spite of the known overall importance of viruses, the dynamics of their interactions with their specific microbial hosts remain poorly understood. This project will characterize viral strategies for survival and interaction with their hosts in the ocean, with the goal of enabling a better understanding of the conditions under which viruses can effectively control bacterial populations. The work will generate and integrate diverse data types, ranging from quantification of specific interactions, environmental dynamics of microbial hosts and their viruses, and comparative genome analysis. While the project focuses on the coastal ocean of New England, the approaches and findings will be applicable to the larger field of marine microbial ecology, to other virus/host systems in nature and to engineered systems. This project will fill a gap in current microbial ecology curricula by creating a bioinformatics module to provide training in large-scale sequence data collection and analysis. The module will be refined through testing during an annual course in Nicaragua and will be broadly accessible in the US and internationally. The close collaboration, throughout this project and its associated outreach, between two laboratories with complementary research strengths will provide highly interdisciplinary training for undergraduate students as well for two graduate students.

Viruses and their microbial hosts have co-evolved over billions of year and shape the ecology of the ocean in many ways. Broadly, understanding the mechanisms and emergent properties of virus-host interactions will allow for better understanding and modeling of biogeochemical cycles and the diversity of microbes at the population and genomic level. The guiding hypothesis of this project is that the prevalence of each of different viral strategies is probabilistic and linked to host availability, environmental parameters, and frequency-dependent competition with other virus strains for available hosts. This research will address four aims that characterize how viruses interact with their hosts in the dilute ocean environment by:
(1) quantification of ecological tradeoffs between specialist and generalist viral strategies,
(2) estimation of the prevalence of dual lytic/lysogenic strategy in marine viruses,
(3) identification of host surface receptors of particular viruses and examination of genetic signatures of distinct viral strategies in virus and host genomes, and
(4) identification of genetic and metabolic interactions between virus and host genomes.

This study takes advantage of a model system with the largest available collection of viruses and hosts for which host range and genome sequences have been determined. This work will provide fine-scale analysis of host and phage genomic diversity and abundance in this model system, while at the same time estimating host-range and co-infection, all of which represent important, poorly constrained parameters in virus-host interactions. Finally, this project complements the large number of studies that have looked at single host-virus interactions, metagenome sequencing, and assessment of viral impact on microbial production.

[table of contents | back to top]

Funding

Funding Source	Award
NSF Division of Ocean Sciences (NSF OCE)	OCE-1435868