Protein sequences from diversity-generating retroelements in groundwater microorganisms collected near Rifle, Colorado between 2011 and 2012 (Viruses in Methanotrophic Marine Ecosystems project)

Website: https://www.bco-dmo.org/dataset/687822
Data Type: experimental
Version: 1
Version Date: 2017-04-11

Project
» Dimensions: The Role of Viruses in Structuring Biodiversity in Methanotrophic Marine Ecosystems (Viruses in Methanotrophic Marine Ecosystems)

Program
» Dimensions of Biodiversity (Dimensions of Biodiversity)

Contributors	Affiliation	Role
Valentine, David L.	University of California-Santa Barbara (UCSB)	Principal Investigator, Contact
Paul, Blair	University of California-Santa Barbara (UCSB-MSI)	Student
York, Amber D.	Woods Hole Oceanographic Institution (WHOI BCO-DMO)	BCO-DMO Data Manager

Table of Contents

Coverage

Spatial Extent: **Lat**:39.5291 **Lon**:-107.7721

Dataset Description

This dataset includes links to diversity-generating retroelement sequences files in .fasta format. There are separate .fasta files for reverse transcriptase protein sequences and variable protein sequences. These sequences were derived from previously generated sequence accessions at the National Center for Biotechnology Information (NCBI). Original sampling took place near Rifle, Colorado between 2011 and 2012.

To access the .fasta files and the list of source sequences, click the "Get Data" button at the top of this page.

These data are published in the following journal article:
Paul, B.G., Burstein, D., Castelle, C.J., Handa, S., Arambula, D., Czornyj, E., Thomas, B.C., Ghosh, P., Miller, J.F., Banfield, J.F. and Valentine, D.L. (2017) Retroelement-guided protein diversification abounds in vast lineages of Bacteria and Archaea. Nature Microbiology, 2, p.17045. doi: 10.1038/nmicrobiol.2017.45

Methods & Sampling

A bioinformatic analysis of metagenome-assembled genomes to identify retroelements was performed. This study used data from several previously described sampling efforts (Brown et al. 2015, Castelle et al. 2015, and Anantharaman et al. 2016). Original sampling was conducted within an unconfined aquifer at the Rifle Integrated Field Research Challenge (IFRC) site, which is adjacent to the Colorado River, near Rifle, Colorado, USA (39.5291, -107.7721). Samples were collected between August 25th to December 12th of 2011, and August 2nd to December 12th of 2012. Sequencing of these samples took place at the Joint Genome Institute using the Illumina HiSeq 2000 platform to generate 2 × 150 paired-end reads.

For more information about sampling methodology see Paul et al. 2017.

Data Processing Description

These sequences were processed using Python v2.7.12, and Geneious v8.1.4.

[table of contents | back to top]

Data Files

File
DGR_RT_aa_Sequences.fasta (FASTA, 259.50 KB) MD5:5be2cc58038583355868895efc9ca52f Reverse transcriptase protein sequences from diversity-generating retroelements identified in groundwater metagenomes
DGR_VP_aa_Sequences.fasta (FASTA, 59.38 KB) MD5:6f7f4b5dd0bcd3ed38b78142e3f6bdff Variable protein sequences from diversity-generating retroelements identified in groundwater metagenomes
sequence_sources.csv (Comma Separated Values (.csv), 59.14 KB) MD5:0476fc0310ac3a07fe51ae80a77a0f86 # List of previously generated sequence accessions used for the groundwater microorganism retroelement protein sequence dataset # PI: David Valentine # N.B. This list of accessions has various authors, see accession links for full metadata at NCBI

[table of contents | back to top]

Related Publications

Anantharaman, K., Brown, C. T., Hug, L. A., Sharon, I., Castelle, C. J., Probst, A. J., … Banfield, J. F. (2016). Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. Nature Communications, 7, 13219. doi:10.1038/ncomms13219
Methods

Brown, C. T., Hug, L. A., Thomas, B. C., Sharon, I., Castelle, C. J., Singh, A., … Banfield, J. F. (2015). Unusual biology across a group comprising more than 15% of domain Bacteria. Nature, 523(7559), 208–211. doi:10.1038/nature14486
Methods

Castelle, C. J., Wrighton, K. C., Thomas, B. C., Hug, L. A., Brown, C. T., Wilkins, M. J., … Banfield, J. F. (2015). Genomic Expansion of Domain Archaea Highlights Roles for Organisms from New Phyla in Anaerobic Carbon Cycling. Current Biology, 25(6), 690–701. doi:10.1016/j.cub.2015.01.014
Methods

Paul, B. G., Burstein, D., Castelle, C. J., Handa, S., Arambula, D., Czornyj, E., … Valentine, D. L. (2017). Retroelement-guided protein diversification abounds in vast lineages of Bacteria and Archaea. Nature Microbiology, 2, 17045. doi:10.1038/nmicrobiol.2017.45
Results
,

Methods

[table of contents | back to top]

Parameters

Parameter	Description	Units
sequence_id		unitless
accession_id		unitless
accession_link		unitless

[table of contents | back to top]

Instruments

Dataset-specific Instrument Name	Illumina HiSeq 2000
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.

[table of contents | back to top]

Deployments

Valentine_IFRC_site

Website	https://www.bco-dmo.org/deployment/692458
Platform	Colorado_River
Description	Original sampling was conducted within an unconfined aquifer at the Rifle Integrated Field Research Challenge (IFRC) site, which is adjacent to the Colorado River, near Rifle, Colorado, USA.

[table of contents | back to top]

Project Information

Dimensions: The Role of Viruses in Structuring Biodiversity in Methanotrophic Marine Ecosystems (Viruses in Methanotrophic Marine Ecosystems)

Marine methanotrophic ecosystems are responsible for consuming around 75 Tg of methane annually, preventing this potent greenhouse gas from entering the atmosphere. These microbial ecosystems thus play a vital role in the global climate system. The nature of these communities depends on the presence or absence of oxygen: methanotrophy is a bacterial lifestyle in aerobic shallow sediments, but in deeper anaerobic sediments it is the exclusive province of archaea, in syntrophy with sulfate-reducing bacteria. It is known which phyla are most commonly found in methanotrophic environments. However, because of these environments' physical inaccessibility and because nearly all microbes from these systems have resisted cultivation, understanding of these communities lags far behind their importance. The cultivation-resistance of microbial hosts from these systems has additionally prevented the use of classical methods to study the viral community. Thus, to date science is largely unable to fill in the broad outlines of marine methanotrophic biodiversity, to fully describe the microbial communities or determine what shapes them.

This project seeks to define the importance of viruses in structuring functional, genetic, and taxonomic diversity in methanotrophic marine ecosystems. The underlying assertion is that viruses structure the diversity of archaeal and bacterial communities in these ecosystems by causing both mortality and horizontal gene transfer. To establish viral contributions to biodiversity of aerobic and anaerobic marine methanotrophic ecosystems, this project combines biogeochemical, genomic, and metagenomic approaches, in both field and laboratory settings.

The project first seeks to assess viral activity in situ by extending established stable isotope probing techniques to quantify rates of viral production at sea floor methane seeps. The same techniques will be used to track the flow of carbon from methane to microbes to viruses and to isolate genetic material from just those organisms that actively cycle methane-derived carbon, enabling the production of microbial and viral metagenomes that are anchored in ecosystem function. Comparisons among these metagenomes will reveal any functional sequences in transit between organisms, providing the basis for an evaluation of the relationships between functional and genetic diversity. At the same time, single-cell whole-genome amplification will pinpoint individual cells for comparison with the microbial and viral assemblages, permitting assessment of the relationships between taxonomic and genetic diversity. Last, the comparison of genomic and metagenomic data both within and across distinctive marine methanotrophic ecosystems will enable analysis of the relationship between functional and taxonomic diversity.

[table of contents | back to top]

Program Information

Dimensions of Biodiversity (Dimensions of Biodiversity)

Website: http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=503446

Coverage: global

(adapted from the NSF Synopsis of Program)
Dimensions of Biodiversity is a program solicitation from the NSF Directorate for Biological Sciences. FY 2010 was year one of the program. [MORE from NSF]

The NSF Dimensions of Biodiversity program seeks to characterize biodiversity on Earth by using integrative, innovative approaches to fill rapidly the most substantial gaps in our understanding. The program will take a broad view of biodiversity, and in its initial phase will focus on the integration of genetic, taxonomic, and functional dimensions of biodiversity. Project investigators are encouraged to integrate these three dimensions to understand the interactions and feedbacks among them. While this focus complements several core NSF programs, it differs by requiring that multiple dimensions of biodiversity be addressed simultaneously, to understand the roles of biodiversity in critical ecological and evolutionary processes.

[table of contents | back to top]

Funding

Funding Source	Award
NSF Division of Ocean Sciences (NSF OCE)	OCE-1046144

[table of contents | back to top]