Water quality data, gene copy number counts, and 16S analyses from Rhode River samples collected at SERC during 2017 and 2018

Website: https://www.bco-dmo.org/dataset/757405
Data Type: Other Field Results
Version: 2
Version Date: 2020-12-03

Project
» EAGER: High-throughput, culture-independent technique identifying cyanobacteria infections to improve understanding of carbon biogeochemical cycling (identify cyano infections)

Contributors	Affiliation	Role
Preheim, Sarah	Johns Hopkins University (JHU)	Principal Investigator
Biddle, Mathew	Woods Hole Oceanographic Institution (WHOI BCO-DMO)	BCO-DMO Data Manager
Rauch, Shannon	Woods Hole Oceanographic Institution (WHOI BCO-DMO)	BCO-DMO Data Manager

Abstract
Water quality data, gene copy number counts, and 16S analyses from Rhode River samples collected at SERC during 2017 and 2018.

Table of Contents

Coverage

Spatial Extent: **N**:38.8869 **E**:-76.5402 **S**:38.8869 **W**:-76.5418
Temporal Extent: 2017-05-17 - 2018-12-27

Dataset Description

Water quality data, gene copy number counts, and 16S analyses from water samples collected from the Rhode River at the Smithsonian Environmental Research Center (SERC) in Edgewater, MD during 2017 and 2018.

Acquisition Description

Surface water samples were collected for subsequent microbial community analyses via PCR and shotgun metagenomics. A YSI EXO2 water quality sonde operated by the Smithsonian Environmental Research Center was used to collect physicochemical conditions at time of sample collection.

Shotgun metagenomic DNA and 16S rRNA microbial genes were amplified as described in Sakowski et al. (in revision). Metagenomic DNA was prepped using the Nextera Flex library prep kit. 16S rRNA libraries were prepped by V3-V4 PCR amplification. Both shotgun metagenome and 16S libraries were sequenced on an Illumina MiSeq.

qPCR of 16S rRNA genes and Viral Ribonucleotide Reductase Genes:
Bacterial and viral abundances were estimated by quantitative PCR. For Ribonucleotide Reductase (RNR) quantification, an environmental RNR amplicon was first cloned into chemically competent *Escherichia coli* cells using the Zero Blunt PCR Cloning Kit (Thermo Scientific) following the manufacturer's protocol. Environmental samples were quantitated for 16S rRNA and RNR gene copy numbers by comparing to a serial dilution of the cloned *E. coli* cells at known concentrations. All standards and environmental samples were run in triplicate. Three microliters of sample were combined with UltraPure molecular grade water (Thermo, Inc.), SsoAdvanced Universal SYBR Green Supermix (1x final concentration, Bio-Rad Laboratories, Inc.), Forward primer (0.3 mM final concentration), and Reverse primer (0.3 mM final concentration) to a final volume of 25 mL. Samples were amplified on a CFX96 Real-Time PCR Detection System (Bio-Rad Laboratories, Inc.) with the following conditions: denaturing at 98°C for 10 minutes; 45 cycles of denaturing at 98°C for 10 seconds, annealing at 52°C for 30 seconds, and extension at 72°C for 45 seconds; and a final extension of 72°C for 5 minutes.

16S rRNA amplification and sequencing:
16S rRNA genes were amplified from surface water samples in a 25 mL PCR reaction with the following conditions: three microliters of column-purified DNA were combined with UltraPure molecular grade water (Thermo, Inc.), 10X buffer (1x final concentration), dNTPs (0.1mM each final concentration), 16S forward primer 27F (0.3 mM final concentration), 16S reverse primer PE_16S_V4_E786_R (0.3 mM final concentration), bovine serum albumin (0.02 mg/mL final concentration), and Phusion High-Fidelity DNA Polymerase (0.5U; New England BioLabs, Inc. PCR reactions were combined with 150 mL of 4% UMIL EM90 oil (4% UMIL EM90 oil, 0.05% TritonX-100 v/v in mineral oil; Universal Preserv-A-Chem, Inc.) and emulsified by vortexing at max speed (~2,700 rpm) for one minute on a Vortex Genie 2 (MoBio). Emulsions were loaded as 50 mL aliquots and amplified with the following conditions: denaturation at 94°C for 3 minutes; 33 cycles of denaturation at 94°C for 10 seconds, annealing at 54°C for 30 seconds, and extension at 72°C for 45 seconds; and a final extension of 72°C for 5 minutes (C1000, BioRad Labs., Inc.).

PCR oil emulsions were broken with isobutanol. Briefly, PCR aliquots were pooled in a 1.5mL microcentrifuge tube and combined with 100 uL of sterile 5M NaCl solution and 1 mL of isobutanol. Samples were vortexed briefly to mix and centrifuged at 16,000 x g for 1 minute. The bottom aqueous layer was retained, and DNA was purified by spin column purification (Zymo, Inc.). DNA was eluted in 20 uL of Tris-HCl.

Purified DNA was run on a 1.5% agarose gel (UltraPure Agarose, ThermoFisher Scientific). 16S rRNA gene bands were visualized under blue light excitation, extracted, and gel purified (Zymo, Inc.) Purified DNA was eluted into 20 mL of Tris-HCl. Barcodes and Illumina adapters were added to 16S rRNA gene amplicon products in two subsequent limited PCR steps. Barcodes were added as follows: two microliters of purified DNA were combined with UltraPure molecular grade water (Thermo, Inc.), 10X buffer (1x final concentration), dNTPs (0.1mM each final concentration), 16S forward primer PE_16S_V4_U515F (0.3 mM final concentration), 16S rRNA gene reverse primer with 8-mer barcodes PE_IV_XXX (0.3 mM final concentration), and Phusion High-Fidelity DNA Polymerase (0.5U; New England BioLabs, Inc.). Samples were amplified with the following conditions: denaturing at 98°C for 30 seconds; 8 cycles of denaturing at 98°C for 10 seconds, annealing at 54°C for 30 seconds, and extension at 72°C for 45 seconds; and a final extension of 72°C for 5 minutes. DNA was purified by spin column purification (Zymo, Inc.) and eluted into 20 mL Tris-HCl. Illumina adapters were then added as above with the following primers: Illumina adapter forward primer PE-III-PCR-F (0.3 mM final concentration) and Illumina adapter reverse primer Barcode_Rev (0.3 mM final concentration). Samples were amplified with the following conditions: denaturing at 98°C for 30 seconds; 5 cycles of denaturing at 98°C for 10 seconds, annealing at 54°C for 30 seconds, and extension at 72°C for 45 seconds; and a final extension of 72°C for 5 minutes. DNA was purified by spin column purification (Zymo, Inc.) and eluted into 20 mL Tris-HCl. 16S rRNA gene amplicon libraries were sequenced on an Illumina MiSeq (2 x 300 bp) at the Genetic Core Research Facility at Johns Hopkins University.

Shotgun metagenomic library preparation and sequencing:
Surface water samples were filtered through a 0.2µm PES filter. DNA was extracted from filters for shotgun sequencing from water samples with the DNeasy PowerWater kit (Qiagen) following the manufacturer's protocol with the following amendment: 20 µL of proteinase K was combined with 1 mL of solution PW1 in the bead tube. The bead tube was incubated at 65°C for ten minutes prior to bead beating. Libraries were prepared with the Nextera DNA Flex Library Prep kit (Illumina, Inc.) following the manufacture's protocol and sequenced on an Illumina MiSeq (2 x 300 bp) at the Genetic Core Research Facility at Johns Hopkins University.

epicPCR of environmental samples:
Surface water glycerol samples (25% v/v) were thawed on ice and one mL was added to three replicate 1.5 mL microcentrifuge tubes per sample. Samples were centrifuged at 25,000 x g for 10 minutes and resuspended after supernatant removal to reduce free viral particles. Thirty microliters of each sample was combined with UltraPure molecular grade water (Thermo, Inc.), 10X buffer (1x final concentration), dNTPs (0.1mM each final concentration), Viral Forward primer (1.0 mM final concentration), Viral Reverse_519R fusion primers (R1 and R2 combined, 0.01 mM each final concentration), 16S Reverseeverse primer (1.0 mM final concentration), bovine serum albumin (0.02 mg/mL final concentration), Tween-20 (0.8% v/v final concentration), and Phusion High-Fidelity DNA Polymerase (1.5U; New England BioLabs, Inc.) to a final volume of 75 mL. PCR reactions were combined with 450 mL of 4% UMIL EM90 oil (4% UMIL EM90 oil, 0.05% TritonX-100 v/v in mineral oil; Universal Preserv-A-Chem, Inc.) and emulsified by vortexing at max speed (~2,700 rpm) for one minute. Emulsions were loaded as 50 mL aliquots and amplified with the following conditions: denaturation at 94°C for 3 minutes; 33 cycles of denaturation at 94°C for 10 seconds, annealing at 54°C for 30 seconds, and extension at 72°C for 45 seconds; and a final extension of 72°C for 5 minutes (C1000, BioRad Labs., Inc.).

PCR oil emulsions were broken with isobutanol. Briefly, PCR aliquots were pooled in a 1.5mL microcentrifuge tube and combined with 100 uL of sterile 5M NaCl solution and 1 mL of isobutanol. Samples were vortexed briefly to mix and centrifuged at 16,000 x g for 1 minute. The bottom aqueous layer was retained, and DNA was purified by spin column purification (Zymo, Inc.). DNA was eluted in 20 uL of Tris-HCl.

Amplicons were enriched by nested PCR and subsequently sequenced on a PacBio Sequel with Sequel v3 chemistry (University of Maryland Institute for Genome Sciences). Circular consensus sequences were obtained from raw reads with the following parameters: minimum signal-to-noise ratio (SNR): 3, minimum length: 500bp, minimum passes: 10, minimum read score: 0.75, minimum predicted accuracy: 0.90.

Processing Description

Data Processing:
16S rRNA sequence data were analyzed using the QIIME2 pipeline. Shotgun metagenomes were assembled with metaSPAdes and binned with metaWRAP as described in Sakowski et al. (in revision).

BCO-DMO Processing Notes:
- Split the latitude and longitude values into independent columns;
- converted longitude from degrees West to degrees East;
- reformatted date and time to ISO8601 convention;
- added date-time field in UTC;
- modified parameter names to conform with BCO-DMO naming conventions;
- blank values replaced with no data value 'nd';
- concatenated data files from 2017 and 2018.

[table of contents | back to top]

Related Datasets

IsRelatedTo

Preheim, S. (2020) **Gene copy number counts and 16S analyses of July 2019 Rhode River surface water incubated with or without viral dilution.** Biological and Chemical Oceanography Data Management Office (BCO-DMO). (Version 1) Version Date 2020-08-25 http://lod.bco-dmo.org/id/dataset/821955 [view at BCO-DMO]

[table of contents | back to top]

Parameters

Parameter	Description	Units
Sample	Sample date and identifer	unitless
ISO_DateTime_EST	Date and time (EST) of observation formatted to ISO8601 standard: YYYY-MM-DDThh:mm	unitless
ISO_DateTime_UTC	Date and time (UTC) of observation formatted to ISO8601 standard: YYYY-MM-DDThh:mm	unitless
Temp	Water Temperature	degrees Celsius
Conductivity	Conductivity	millisiemens per centimeter (mS/cm)
DO_pcnt	Water Dissolved Oxygen % air saturation	unitless (percent)
Pressure	Water pressure	psi
pH	Water pH	pH scale
Chlorophyll_RFU	Water Chlorophyll	RFU
BGA_PE_RFU	Water Total Algae	RFU
FDOM_RFU	Fluorescent Dissolved Organic Matter	RFU
Specific_Conductivity	Water Specific Conductivity	millisiemens per centimeter (mS/cm)
Salinity	Water Salinity	psu
DO	Water Dissolved Oxygen	milligrams per Liter (mg/L)
Depth	Probe Depth	meters (m)
Turbidity	Water Turbidity	FNU
Chlorophyll	Water Chlorophyll	micrograms per Liter (ug/L)
BGA_PE	Water Total Algae	micrograms per Liter (ug/L)
FDOM	Fluorescent Dissolved Organic Matter	ppb QSE
Location	Name of sample site	unitless

Lat	Site Latitude with North being positive	decimal degrees
Lon	Site longitude with west being negative	decimal degrees
Instrumentation	Name/type of sensor used for measurements	unitless
Sample_ID	Unique Sample Identifier	unitless
Genes_16S_1	16S qPCR Gene Copy Number Mean	gene copy nmuber mean per milliliter (per mL)
Genes_16S_2	16S qPCR Gene Copy Number Standard Error	gene copy number standard error per mL
RNR_Genes_1	RNR qPCR Gene Copy Number Mean	gene copy number mean per mL
RNR_Genes_2	RNR qPCR Gene Copy Number Standard Error	gene copy number standard error per mL
Library_Accession_Number_16S	NCBI Accession Numbers for 16S Libraries	unitless
Shotgun_Metagenome_Library_Accession_Number	NCBI Accession Numbers for Shotgun Metagenome Libraries	unitless
Fusion_Amplicon_Accession_Numbers	NCBI Accession Numbers for 16S/RNR Fusion Libraries	unitless

[table of contents | back to top]

Instruments

Dataset-specific Instrument Name	Illumina MiSeq
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	YSI EXO2 water quality sonde
Generic Instrument Name	YSI Professional Plus Multi-Parameter Probe
Dataset-specific Description	YSI EXO2 water quality sonde operated by the Smithsonian Environmental Research Center
Generic Instrument Description	The YSI Professional Plus handheld multiparameter meter provides for the measurement of a variety of combinations for dissolved oxygen, conductivity, specific conductance, salinity, resistivity, total dissolved solids (TDS), pH, ORP, pH/ORP combination, ammonium (ammonia), nitrate, chloride and temperature. More information from the manufacturer.

Dataset-specific Instrument Name	CFX96 Real-Time PCR Detection System
Generic Instrument Name	qPCR Thermal Cycler
Generic Instrument Description	An instrument for quantitative polymerase chain reaction (qPCR), also known as real-time polymerase chain reaction (Real-Time PCR).

[table of contents | back to top]

Project Information

EAGER: High-throughput, culture-independent technique identifying cyanobacteria infections to improve understanding of carbon biogeochemical cycling (identify cyano infections)

Coverage: Rhode River waters 38.8869 N, 76.5402 W

NSF Abstract:

Viruses in the ocean are 10 times more abundant than bacteria and kill 10-66% of bacterial cells daily. Viral infections of bacteria, such as of photosynthetic cyanobacteria that form the base of the food web, impact the flow of energy and carbon within the marine ecosystem. Thus, viral infections dramatically alter important biogeochemical and ecological factors in the ocean, such as how much carbon dioxide is respired or how many fish the ocean ecosystem can sustain. Despite their importance in the ocean ecosystem, researchers do not know the answer to the most basic question of viral biology for most environmental viruses: which bacteria do different viruses infect? Identifying these infections could help researchers understand more about how viruses shape the ecosystem through infections of keystone microbial species, infections of microbes with unique characteristics, or infection patterns that promote microbial community stability. This project is to develop a cost-effective method to substantially increase the number of infections identified within natural microbial communities. The researchers are applying this novel method to determine viral infections in cyanobacteria in the Chesapeake Bay and compare the results to standard approaches to determine viral infections. This technique can be widely used to transform our understanding of how viruses impact many ecosystems, since it is cost-effective, does not need specialized equipment and can be adapted to target different viral populations. Additionally, this project provides research opportunities for undergraduate and high school students, including underrepresented minority students and women.

To develop a high-throughput, culture-independent technique to identify infections in the environment, the researchers are optimizing a previously developed method, emulsion paired concatenation-isolation PCR (epicPCR). Adapting epicPCR for viral-host associations will identify interactions by isolating actively infected single cells within a microdroplet to retain the physical proximity of the host and viral DNA during DNA extraction. Next, fusion PCR is done within the microdroplet to allow host rRNA genes to fuse to viral marker genes, such as g20 or ribonucleotide reductase, retained within the same bead. Only rRNA genes successfully fused to viral markers are amplified. Finally, high-throughput sequencing is done on the resulting fusion products. This approach is cultivation-independent, screens a larger fraction of diversity within the sample than traditional approaches, requires little additional equipment compared to microfluidic approaches, and can be scaled up to hundreds of samples because the amount of sequencing required to deeply sample a single environment is low compared to shotgun metagenomic sequencing. Although this technique will be limited to viral marker genes and suffers from the biases of PCR, it still offers great potential to investigate viral-host interactions across a large number of environments. The method is being applied to determine how viral infections influence cynaobacterial blooms in the Chesapeake Bay. The researchers will also compare the results of epicPCR to culture-based, single-cell, and bioinformatics based methods of host-virus associations to identify biases, limitations and caveats of various approaches.

This award reflects NSF's statutory mission and has been deemed worthy of support through evaluation using the Foundation's intellectual merit and broader impacts review criteria.

[table of contents | back to top]

Funding

Funding Source	Award
NSF Division of Ocean Sciences (NSF OCE)	OCE-1820652

[table of contents | back to top]