Assembled metagenome sequences collected on the R/V Melville (MGLN10MV) and R/V Atlantis (Alvin AT11-07) cruises in the Loihi Seamount, Hawaiian archipelago and East Pacific Rise from 2011-2013 (EPR and Loihi basalt genomes project)

Website: https://www.bco-dmo.org/dataset/616326
Data Type: experimental
Version: 2016-01-14

Project
» Metagenomic signatures in seafloor rocks and subsurface sediments: East Pacific Rise and Loihi Seamount (EPR and Loihi basalt genomes)

Program
» Center for Dark Energy Biosphere Investigations (C-DEBI)

Contributors	Affiliation	Role
Edwards, Katrina	University of Southern California (USC)	Lead Principal Investigator
Singer, Esther	University of Southern California (USC)	Principal Investigator, Contact
Copley, Nancy	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
- Program Information
- Funding

Dataset Description

This dataset includes FASTA formatted files of the metagenomes from the East Pacific Rise seafloor basalt, the Loi'hi Seamount seafloor basalt, and negative control of sterile water.

The sequence data has been submitted to NCBI under Bioproject number PRJNA302671:
- Biosamples SUB1190178 (blank)
- SUB1190153 (Lo'ihi)
- SUB1190159 (EPR)

This dataset is the result of a C-DEBI Graduate Student Fellowship award (2011).

Methods & Sampling

Methodology:

EPR: Rock material was collected from the EPR at 9.725 N, 104.16 W (2,674 m depth) aboard the R/V Atlantis using the submarine Alvin (cruise AT11-07, dive 3968) in 2004. DNA was extracted from basalt chips using a phenol-chloroform extraction including a negative control. DNA was amplified using the illustrate GenomiPhi V2 DNA Multiple Displacement Amplification (MDA) kit (GE Healthcare Life Sciences, Pittsburgh, PA, USA). Final DNA samples and the control were sent to the core genomics center at University of Pennsylvania for whole

genome shotgun sequencing on a Roche GS-FLX Titanium 454 sequencer (454 Life Sciences, Branford, CT, USA).

Loi'hi: Two seafloor basalt samples (J2-243 R2-F, J-246 R2) were collected from the Lō'ihi Seamount (18.47 N, 155.18 W) at a depth of 5,000 m aboard the R/V Melville using the ROV Jason II in 2006. DNA was amplified using the illustrate GenomiPhi V2 DNA Multiple Displacement Amplification (MDA) kit (GE Healthcare Life Sciences, Pittsburgh, PA, USA). Final DNA samples and the control were sent to the core genomics center at University of Pennsylvania for whole genome shotgun sequencing on a Roche GS-FLX Titanium 454 sequencer (454 Life Sciences, Branford, CT, USA).

Negative control: DNA was extracted from sterile MilliQ water using a phenol-chloroform extraction. DNA was amplified using the illustrate GenomiPhi V2 DNA Multiple Displacement Amplification (MDA) kit (GE Healthcare Life Sciences, Pittsburgh, PA, USA). Final DNA was sent to the core genomics center at University of Pennsylvania for whole genome shotgun sequencing on a Roche GS-FLX Titanium 454 sequencer (454 Life Sciences, Branford, CT, USA).

Data Processing Description

All seafloor basalts were stored frozen at -80°C for XRD analysis and DNA extraction. Bulk mineralogy analysis, *i.e.* quantitative determination of rock-forming minerals and total clay minerals, was determined on all three seafloor basalts via X-ray Diffraction (XRD) analysis at KT GeoServices, Inc. Detection limits were at 1-5 wt%. For the two Lō'ihi seafloor basalts, average values were used.

Raw sequence reads were evaluated with FastQC version 0.11.3 (Schmieder and Edwards, 2011a), quality trimmed (minimum quality score – 25, maximum length – 450 bp, maximum homopolymer length – 9 bp, max N-tail – 1 bp), and filtered (removal of technical duplicates, minimum length – 60 bp) with Prinseq 0.20.4 (Schmieder and Edwards, 2011b) and MG-RAST (Meyer et al., 2008). We obtained 1,191,651 sequences in the EPR dataset; 1,102,191 sequences in the Lo'ihi dataset; and 58,188 sequences in the negative control dataset. Quality-filtered reads were assembled denovo using standard 454 settings in mira 3.4.1.1 (Chevreux et al., 1999). Padded (i.e. including potential gaps) contigs > 500 bp were filtered using mira 3.4.1.1 (convert_project) (Chevreux et al., 1999). Seafloor basalt contigs were screened for contamination using a combination of BBMap (bbduk.sh with parameters mcf=0.25, k=31) and the BLASTN algorithm (Altschul et al., 1990). The BBMap algorithm identified 4 potentially contaminant contigs in the EPR metagenome dataset (total of 4,290 bp) and 10 potentially contaminant contigs in the Lo'ihi (total of 12,423 bp). Community richness was estimated using the Chao1 index, diversity analysis was calculated using the Shannon index in QIIME 1.9.1 (alpha_diversity.py) based on BLASTX assignments of contigs. Phylosift was used to assess community diversity using the core molecular marker set of genes, which includes ~40 three-domain protein coding genes, single-copy eukaryote specific nuclear orthologs, ribosomal RNA genes (16S/18S), mitochondrial genes (mtDNA markers), and plastid and viral markers identified through Markov-clustering algorithms applied to genome datasets (Darling et al., 2014).

BCO-DMO Processing:

version 2015-11-06: replaced version 2015-10-23. Added site, lat, lon, and date columns.

[table of contents | back to top]

Data Files

File
metagenomes_fasta3.csv (Comma Separated Values (.csv), 881 bytes) MD5:35cbd4801daa4764e76569d58305fb00 Primary data file for dataset ID 616326

[table of contents | back to top]

Parameters

Parameter	Description	Units
site	Site name: EPR = East Pacific Rise: Loihi = Loi'hi Seamount Hawaii.	unitless
lat	latitude; north is positive	decimal degrees
lon	longitude;east is positive	decimal degrees
year	year of sample collection	YYYY
month	month of sample collection	MM
day	day of sample collection	DD
sample_type	type of sample	unitless
Fasta_file_link	Link to FASTA metagenome files	unitless
NCBI_accession	link to NCBI GenBank accession page	unitless

[table of contents | back to top]

Instruments

Dataset-specific Instrument Name	
Generic Instrument Name	Automated DNA Sequencer
Dataset-specific Description	Roche GS-FLX Titanium 454 sequencer (454 Life Sciences, Branford, CT, USA) at the core genomics center at University of Pennsylvania.
Generic Instrument Description	A DNA sequencer is an instrument that determines the order of deoxynucleotides in deoxyribonucleic acid sequences.

Dataset-specific Instrument Name	Jason II
Generic Instrument Name	ROV Jason
Dataset-specific Description	Jason dives 243 and 246 (18.47 N, 155.18 W).
Generic Instrument Description	The Remotely Operated Vehicle (ROV) Jason is operated by the Deep Submergence Laboratory (DSL) at Woods Hole Oceanographic Institution (WHOI). WHOI engineers and scientists designed and built the ROV Jason to give scientists access to the seafloor that didn't require them leaving the deck of the ship. Jason is a two-body ROV system. A 10-kilometer (6-mile) fiber-optic cable delivers electrical power and commands from the ship through Medea and down to Jason, which then returns data and live video imagery. Medea serves as a shock absorber, buffering Jason from the movements of the ship, while providing lighting and a bird's eye view of the ROV during seafloor operations. During each dive (deployment of the ROV), Jason pilots and scientists work from a control room on the ship to monitor Jason's instruments and video while maneuvering the vehicle and optionally performing a variety of sampling activities. Jason is equipped with sonar imagers, water samplers, video and still cameras, and lighting gear. Jason's manipulator arms collect samples of rock, sediment, or marine life and place them in the vehicle's basket or on "elevator" platforms that float heavier loads to the surface. More information is available from the operator site at URL. https://ndsf.whoi.edu/jason/

Dataset-specific Instrument Name	
Generic Instrument Name	Thermal Cycler
Dataset-specific Description	DNA was amplified using the illustrate GenomiPhi V2 DNA Multiple Displacement Amplification (MDA) kit (GE Healthcare Life Sciences, Pittsburgh, PA, USA)
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)

Dataset-specific Instrument Name	X-ray diffractometer (XRD)
Generic Instrument Name	X-ray diffractometer
Dataset-specific Description	X-ray Diffraction (XRD) analysis at KT GeoServices, Inc.
Generic Instrument Description	Instruments that identify crystalline solids by measuring the characteristic spaces between layers of atoms or molecules in a crystal.

[table of contents | back to top]

Deployments

MGLN10MV

Website	https://www.bco-dmo.org/deployment/616328
Platform	R/V Melville
Start Date	2006-10-24
End Date	2006-11-12
Description	The first of three yearly cruises as part of FeMO with ROV/Jason II to study the microbial environment at Loihi Seamount.

AT11-07

Website	https://www.bco-dmo.org/deployment/616332
Platform	R/V Atlantis
Start Date	2004-01-28
End Date	2004-02-24

AT11-07_Alvin_Dives

Website	https://www.bco-dmo.org/deployment/626019
Platform	HOV Alvin
Start Date	2014-02-03
End Date	2014-02-19
Description	**Methods & Sampling** Dive #396 (9.73 N, 104.16 W)

[table of contents | back to top]

Project Information

Metagenomic signatures in seafloor rocks and subsurface sediments: East Pacific Rise and Loihi Seamount (EPR and Loihi basalt genomes)

Coverage: East Pacific Rise and Lo'ihi Seamount (Hawaii)

The seafloor and subsurface microbial world represents a significant portion of life on our planet. The influence on its proximate ambience and global processes, such as element cycles, has potentially been largely underestimated and not always been precisely evaluated. I am interested in the nature of deep biosphere microorganisms in rocks from the Loihi seamount, Hawai'i, the East Pacific Rise, and the Juan de Fuca Ridge, as well as in sediments from North Pond (Mid-Atlantic). In order to assess microbial diversity, metabolic activity, adaptation strategies and biogeographical signatures in the deep subseafloor biosphere, metagenomics by pyrosequencing will be used to complement previous research efforts with the most in-depth and precise data that is available to date.

This project is a C-DEBI Graduate Student Fellowship award (2011)

[table of contents | back to top]

Program Information

Center for Dark Energy Biosphere Investigations (C-DEBI)

Website: http://www.darkenergybiosphere.org

Coverage: Global

The mission of the Center for Dark Energy Biosphere Investigations (C-DEBI) is to explore life beneath the seafloor and make transformative discoveries that advance science, benefit society, and inspire people of all ages and origins.

C-DEBI provides a framework for a large, multi-disciplinary group of scientists to pursue fundamental questions about life deep in the sub-surface environment of Earth. The fundamental science questions of C-DEBI involve exploration and discovery, uncovering the processes that constrain the sub-surface biosphere below the oceans, and implications to the Earth system. What type of life exists in this deep biosphere, how much, and how is it distributed and dispersed? What are the physical-chemical conditions that promote or limit life? What are the important oxidation-reduction processes and are they unique or important to humankind? How does this biosphere influence global energy and material cycles, particularly the carbon cycle? Finally, can we discern how such life evolved in geological settings beneath the ocean floor, and how this might relate to ideas about the origin of life on our planet?

C-DEBI's scientific goals are pursued with a combination of approaches:
(1) coordinate, integrate, support, and extend the research associated with four major programs—Juan de Fuca Ridge flank (JdF), South Pacific Gyre (SPG), North Pond (NP), and Dorado Outcrop (DO)—and other field sites;
(2) make substantial investments of resources to support field, laboratory, analytical, and modeling studies of the deep subseafloor ecosystems;
(3) facilitate and encourage synthesis and thematic understanding of submarine microbiological processes, through funding of scientific and technical activities, coordination and hosting of meetings and workshops, and support of (mostly junior) researchers and graduate students; and
(4) entrain, educate, inspire, and mentor an interdisciplinary community of researchers and educators, with an emphasis on undergraduate and graduate students and early-career scientists.

Note: Katrina Edwards was a former PI of C-DEBI; James Cowen is a former co-PI.

Data Management:
C-DEBI is committed to ensuring all the data generated are publically available and deposited in a data repository for long-term storage as stated in their Data Management Plan (PDF) and in compliance with the NSF Ocean Sciences Sample and Data Policy. The data types and products resulting from C-DEBI-supported research include a wide variety of geophysical, geological, geochemical, and biological information, in addition to education and outreach materials, technical documents, and samples. All data and information generated by C-DEBI-supported research projects are required to be made publically available either following publication of research results or within two (2) years of data generation.

To ensure preservation and dissemination of the diverse data-types generated, C-DEBI researchers are working with BCO-DMO Data Managers make data publicly available online. The partnership with BCO-DMO helps ensure that the C-DEBI data are discoverable and available for reuse. Some C-DEBI data is better served by specialized repositories (NCBI's GenBank for sequence data, for example) and, in those cases, BCO-DMO provides dataset documentation (metadata) that includes links to those external repositories.

[table of contents | back to top]

Funding

Funding Source	Award
NSF Division of Ocean Sciences (NSF OCE)	OCE-0939564

[table of contents | back to top]