

Whole genome sequence data from bacterial isolates from venting fluids at NW Rota Seamount, collected on R/V Thomas G. Thompson and R/V Kilo Moana cruises TN232 and KM1005 in the Mariana arc of the western Pacific in 2009 and 2010

Website: <https://www.bco-dmo.org/dataset/632784>

Data Type: Other Field Results

Version: 18 Jan 2016

Version Date: 2016-01-18

Project

» [Functional gene diversity and expression in ocean crust microbial communities](#) (NP Functional Gene Div)

Program

» [Center for Dark Energy Biosphere Investigations](#) (C-DEBI)

Contributors	Affiliation	Role
Huber, Julie	Marine Biological Laboratory (MBL)	Principal Investigator
Rauch, Shannon	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](#)

Dataset Description

Whole genome sequence data from bacterial isolates from venting fluids at NW Rota Seamount, collected in 2009 and 2010 on cruises TN232 and KM1005.

Methods & Sampling

Diffuse hydrothermal vent fluids were collected at several vent sites on NW Rota-1 seamount in 2009 and 2010 using the ROV Jason 2 and the hydrothermal fluid and particle sampler. Anaerobic enrichment media previously used for the isolation of *Caminibacter profundus* was inoculated with 1 ml of unfiltered diffuse flow fluids and incubated at 55 degrees C. Enrichments with positive microbial growth were isolated by three sets of dilution-to-extinction. The growth of *Lebetimonas* under varying conditions including alternative electron donor/acceptor pairs and with N₂ gas as the sole nitrogen source was evaluated as described in the Supplementary Material of Meyer & Huber (2014). Growth of *Lebetimonas* strain JH369 with N₂ gas as the sole nitrogen source was evaluated using anaerobic seawater media without yeast extract or ammonia and containing formate and elemental sulfur with an 80% N₂ and 20% CO₂ headspace.

Genomic DNA was extracted from pure cultures at log phase using a CTAB extraction. Libraries were prepared using Nextera DNA sample prep kits (Illumina, San Diego, CA, USA) and sequenced by Roche 454 GS FLX Titanium (454 Life Sciences, Branford, CT, USA) and/or using Illumina HiSeq 2000 paired reads (Illumina). In the case of strains sequenced with multiple platforms, the same genomic DNA extraction was used for all library preparations, with the exception of strain JS085. Genomes were assembled using several tools as described in the Supplementary material of Meyer and Huber 2014.

Related references:

Meyer, J.L. and J.A. Huber. 2014. Strain-level genomic variation in natural populations of *Lebetimonas* from an erupting deep-sea volcano. ISME Journal. 8:867–880. doi:[10.1038/ismej.2013.206](https://doi.org/10.1038/ismej.2013.206)

Data Processing Description

Prior to assembly, Illumina sequences were quality filtered using adaptive window trimming and a quality threshold of 30 using the script Trim.pl (<http://wiki.bioinformatics.ucdavis.edu/index.php/Trim.pl>). All reads were screened for adaptor, barcorde, primer, and transposon sequences and trimmed as needed using FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/index.html). De novo genome assembly was performed with several assembly programs. Sequences generated through the 454 platform were first assembled with Roche's GS De Novo Assembler v 2.6 ("Newbler") 2 using default parameters. De novo assemblies of 454 reads were also performed using mira 3 with the default settings for normal quality de novo genome assembly. De novo assembly of subsets of Illumina reads was performed with velvet 4, using an estimated coverage of 1000x, kmer size of 21, and a coverage cutoff of 5). Large contigs from Newbler, mira, and velvet were consolidated using Geneious Pro v 5.6.6 (Biomatters, Ltd, <http://www.geneious.com>) and aligned with progressiveMauve 5 to visualize the relationship of large contigs from different assemblies and to identify gaps to close. Primers were designed at the ends of contigs using either Geneious Pro or CLC Genomics Workbench v 5.1 (CLCbio, <http://www.clcbio.com>) to amplify gaps between contigs. Positive PCR amplification products linking contigs were cleaned using a Min-Elute PCR Purification kit (Qiagen) and Sanger sequenced. A nearly complete draft genome from strain JS085 served as a reference genome for the remaining five strains. Both Illumina and 454 reads were mapped to the reference genome with CLC Genomics Workbench. Unmapped reads were then assembled de novo to ensure that novel genomic content in the mapped strains was not overlooked. De novo assembly of 454 and/or Illumina reads for each strain was also performed in CLC Genomics Workbench and compared to the mapped assemblies using progressiveMauve.

Four of the strains were sequenced using both 454 and Illumina and two strains were sequenced only with Illumina. The sequencing coverage depth of quality-filtered reads ranged from 22X to 50X for 454 and up to 3618X for Illumina. *Lebetimonas* strain JS085 had the highest coverage of 454 reads and was assembled into 33 large contigs with Newbler and 1747 contigs with mira. The 20 largest contigs from each of these assemblies were consolidated using de novo assembly in Geneious to 10 contigs. An additional round of assembly in Geneious with the 10 consolidated contigs and velvet contigs greater than 10 Kbp further consolidated the draft genome to 6 contigs. Primers were designed for all possible combinations between the 6 contigs. One gap was closed using Sanger-sequenced positive pcr products. Finally, all 454 and Illumina reads for strain JS085 were mapped to the draft genome consisting of 5 contigs and the resulting consensus was used as the final draft genome. The five remaining genomes were assembled by mapping 454 and Illumina reads to the JS085 reference genome in CLC Genomics Workbench. Hybrid de novo assemblies in CLC Genomics Workbench of each strain did not extend contigs or close gaps between the 5 contigs of the draft genomes. Assemblies of unmapped reads produced only short contigs with no significant similarities using nucleotide BLAST 6.

BCO-DMO Processing:

- modified parameter names to conform with BCO-DMO naming conventions;
- added hyperlinks;
- removed "m" (meters) in depth column.

[[table of contents](#) | [back to top](#)]

Data Files

File
Lebetimonas_genomes.csv (Comma Separated Values (.csv), 7.69 KB) MD5:69335cf3ef035c17b08cae1173d69106
Primary data file for dataset ID 632784

[[table of contents](#) | [back to top](#)]

Parameters

Parameter	Description	Units
sequencing_center	Name of sequencing center.	dimensionless
domain	Domain of sample.	dimensionless
phylum	Taxonomic phylum.	dimensionless
class	Taxonomic class.	dimensionless
order	Taxonomic order.	dimensionless
family	Taxonomic family.	dimensionless
genus	Taxonomic genus.	dimensionless
study_name	Name of study.	dimensionless
sample_name	Name/identifier of the sample.	dimensionless
taxon_oid	Taxon identifier (OID).	dimensionless
species	Species identifier.	dimensionless
NCBI_accession_num	NCBI accession number.	dimensionless
accession_url	Hyperlink to NCBI for the accession number.	dimensionless
IMG_genome_ID	IMG database (http://img.jgi.doe.gov/) genome identifier.	dimensionless
NCBI_taxon_ID	NCBI taxon identifier.	dimensionless
IMG_submission_ID	IMG database (http://img.jgi.doe.gov/) submission identifier.	dimensionless
GOLD_study_ID	GOLD database (https://gold.jgi.doe.gov/) study identifier.	dimensionless
GOLD_study_url	Hyperlink to GOLD database (https://gold.jgi.doe.gov/) for the study.	dimensionless

GOLD_project_ID	GOLD database (https://gold.jgi.doe.gov/) project identifier.	dimensionless
GOLD_project_url	Hyperlink to GOLD database (https://gold.jgi.doe.gov/) for the project.	dimensionless
GOLD_analysis_project_ID	GOLD database (https://gold.jgi.doe.gov/) analysis project identifier.	dimensionless
GOLD_analysis_project_url	Hyperlink to GOLD database (https://gold.jgi.doe.gov/) for the analysis project identifier.	dimensionless
GOLD_analysis_project_type	GOLD database (https://gold.jgi.doe.gov/) project type.	dimensionless
gene_model_QC	Gene model QC? (yes/no)	dimensionless
submission_type	Submission type.	dimensionless
strain	Strain.	dimensionless
is_public	Is the dataset public? (yes/no)	dimensionless
high_quality	Is it a high quality dataset? (yes/no)	dimensionless
add_date	?	dimensionless
biotic_relationships	Description of the biotic relationships.	dimensionless
cell_shape	Description of the cell shape.	dimensionless
contact_email	Contact email address.	dimensionless
contact_name	Contact name.	dimensionless
culture_type	Culture type.	dimensionless
cultured	Cultured? (yes/no)	dimensionless
depth	Depth.	dimensionless

ecosystem	Description of ecosystem.	dimensionless
ecosystem_category	Description of ecosystem category.	dimensionless
ecosystem_subtype	Description of ecosystem sub-type.	dimensionless
ecosystem_type	Description of ecosystem type.	dimensionless
energy_source	Energy source.	dimensionless
GOLD_sequencing_strategy	GOLD database (https://gold.jgi.doe.gov/) sequencing strategy.	dimensionless
gram_staining	Type of gram staining.	dimensionless
habitat	Description of habitat.	dimensionless
isolation	Description of isolation.	dimensionless
lat	Latitude.	decimal degrees
longhurst_code	Longhurst code.	dimensionless
longhurst_descrip	Longhurst description.	dimensionless
lon	Longitude.	decimal degress
motility	Motility.	dimensionless
O2_requirement	O2 requirements.	dimensionless
project_name	Project name.	dimensionless
relevance	Relevance.	dimensionless
sporulation	Type of sporulation.	dimensionless
temp_range	Description of temperature range.	dimensionless

gene_count	Gene count.	dimensionless
------------	-------------	---------------

[[table of contents](#) | [back to top](#)]

Instruments

Dataset-specific Instrument Name	
Generic Instrument Name	Automated DNA Sequencer
Dataset-specific Description	Libraries were prepared using Nextera DNA sample prep kits (Illumina, San Diego, CA, USA) and sequenced by Roche 454 GS FLX Titanium (454 Life Sciences, Branford, CT, USA) and/or using Illumina HiSeq 2000 paired reads (Illumina).
Generic Instrument Description	A DNA sequencer is an instrument that determines the order of deoxynucleotides in deoxyribonucleic acid sequences.

Dataset-specific Instrument Name	Jason 2
Generic Instrument Name	ROV Jason
Generic Instrument Description	<p>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.who.edu/jason/</p>

[[table of contents](#) | [back to top](#)]

Deployments

TN232

Website	https://www.bco-dmo.org/deployment/568188
Platform	R/V Thomas G. Thompson
Start Date	2009-04-03
End Date	2009-04-17
Description	Data expected from C-DEBI investigator, Julie Huber. Additional cruise information and original data are available from the NSF R2R data catalog.

KM1005

Website	https://www.bco-dmo.org/deployment/567993
Platform	R/V Kilo Moana
Start Date	2010-03-16
End Date	2010-03-30
Description	Data expected from C-DEBI investigator, Julie Huber. Additional cruise information and original data are available from the NSF R2R data catalog.

[[table of contents](#) | [back to top](#)]

Project Information

Functional gene diversity and expression in ocean crust microbial communities (NP Functional Gene Div)

Coverage: North Pond

Project description from [C-DEBI](#):

The objective of this project is to determine the diversity, phylogeny, and expression of functional genes involved in carbon, hydrogen, and sulfur cycling in North Pond crustal fluids. These formation fluids are expected to be representative of the ubiquitous cold ocean crust habitat, where reactions between the water and mineral rock surfaces create substrates suitable for sustaining a potentially large reservoir of microbial life. Information regarding crustal microbial communities and the energy sources available for microbial metabolism has been limited by the inaccessibility of samples. IODP Expedition 336 will provide a unique opportunity to access deep subsurface formation fluids from North Pond, including sampling from multiple depth horizons within oceanic crust. My goal is to develop quantitative polymerase chain reaction assays to determine the expression of functional genes in order to increase our understanding of microbial metabolisms in deep subsurface environments.

This project was funded by a C-DEBI Postdoctoral Fellowship to Julie Meyer (formerly at the Marine Biological Laboratory).

[[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](#)]