Single amplified genomes (SAGs) of microbial cells isolated from deep-sea hydrothermal vent 'Crab Spa', East Pacific Rise, Pacific Ocean from R/V Atlantis AT15-38 and AT26-10, 2008 and 2014 (Microbial Communities at Deep-Sea Vents project)

Website: https://www.bco-dmo.org/dataset/660802
Data Type: experimental
Version: 3
Version Date: 2016-10-05

Project
» An Integrated Study of Energy Metabolism, Carbon Fixation, and Colonization Mechanisms in Chemosynthetic Microbial Communities at Deep-Sea Vents (Microbial Communities at Deep-Sea Vents)

Programs
» Dimensions of Biodiversity (Dimensions of Biodiversity)
» Center for Dark Energy Biosphere Investigations (C-DEBI)

Contributors	Affiliation	Role
Stepanauskas, Ramunas	Bigelow Laboratory for Ocean Sciences	Principal Investigator
Copley, Nancy	Woods Hole Oceanographic Institution (WHOI BCO-DMO)	BCO-DMO Data Manager

Abstract
Single amplified genomes (SAGs) of microbial cells isolated from deep-sea hydrothermal vent 'Crab Spa', East Pacific Rise, Pacific Ocean from R/V Atlantis AT15-38 and AT26-10, 2008 and 2014 (Microbial Communities at Deep-Sea Vents project)

Table of Contents

Coverage

Spatial Extent: **N**:9.84 **E**:-104.29 **S**:9.8383 **W**:-104.2917
Temporal Extent: 2008-10-01 - 2014-01-15

Dataset Description

The accession links are not yet public. They will be archived at the Integrated Microbial Genomes (IMG), at the Joint Genome Inst. (JGI), US Dept. of Energy (DOE). Please contact the PI for further details.

Acquisition Description

AT15-38: Samples for SAGs were obtained using the so-called 'titanium major samplers' (von Damm et al, 1985). Replicate samples of 1 ml aliquots of water were cryopreserved with 6% glycine betaine (Sigma) or 15% glycerol and stored at -80 ºC for the ”Single Cell” aliquot.

AT26-10: Background samples were obtained from the IGTs, in which the incubations were carried out. Just in this case, an aliquot was removed after sample retrieval and before starting the incubation. During AT26-10, samples were preserved with Gly-TE and stored at -80.

Cells were sorted, identified and sequenced by the Bigelow Laboratory Single Cell Genomics FacilityCenter (SCGC), following SCGCthe facilities’s standard practices: SCGC_Services_Description.pdf

Processing Description

On average, at least 5 million 2x150 bp or longer paired-end reads were generated per SAG using in-house MiSeq and NextSeq (Illumina) instruments. The obtained reads were pre-processed and, de novo, assembled and quality-controlled using algorithms SCGC's standard protocols that are optimized for single cell MDA products . A combination of tetramer homogeneity tests and blast searches against reference databases is used to detect potential DNA contaminants among the assembled contigs. Benchmark data demonstrating SCGC SAG WGS whole genome sequencing pipeline performance is available from the SCGC website http://data.bigelow.org/~scgc/WGS_benchmark_data/.

Genome annotation was performed through IMG (http://img-stage.jgi-psf.org/cgi-bin/submit/main.cgi).

Additional related references (pdf)

[table of contents | back to top]

Related Publications

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. Journal of Molecular Biology, 215(3), 403–410. doi:10.1016/s0022-2836(05)80360-2 https://doi.org/10.1016/S0022-2836(05)80360-2
Methods

McNichol, J., Sylva, S. P., Thomas, F., Taylor, C. D., Sievert, S. M., & Seewald, J. S. (2016). Assessing microbial processes in deep-sea hydrothermal systems by incubation at in situ temperature and pressure. Deep Sea Research Part I: Oceanographic Research Papers, 115, 221–232. doi:10.1016/j.dsr.2016.06.011
Related Research

Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J., & Glockner, F. O. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Research, 35(21), 7188–7196. doi:10.1093/nar/gkm864
Methods

Stepanauskas, R., & Sieracki, M. E. (2007). Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time. Proceedings of the National Academy of Sciences, 104(21), 9052–9057. doi:10.1073/pnas.0700496104
Methods

[table of contents | back to top]

Parameters

Parameter	Description	Units
taxon_oid	Identification number of the Single Amplified Genome (SAG) in IMG	unitless
SAG_id	The identifier of the SAG	unitless
taxon	Taxonomic affiliation of the SAG	unitless
source	The origin of the sample used for the single cell sorting:Background: Cell sorted from the natural water samplesControl: Cell sorted from the natural water samples incubated in isobaric chambers without any amendmentsNO3-/H2, ~24C: Cell sorted from the natural water samples incubated at 24oC in isobaric chambers with the addition of NO3- (nitrate) and hydrogen (H2)H2 only: Cell sorted from the natural water samples incubated at 24oC in isobaric chambers with the addition of hydrogen (H2)NO3- only: Cell sorted from the natural water samples incubated at 24oC in isobaric chambers with the addition of NO3- (nitrate)hydrogen (H2)O2, ~110uM: Cell sorted from the natural water samples incubated at 24oC in isobaric chambers with the addition of 110uM of O2 (oxygen)NO3-/H2, ~50C: Cell sorted from the natural water samples incubated at 50oC in isobaric chambers with the addition of NO3- (nitrate) and hydrogen (H2)	unitless
deployment	J denotes "Jason dive" followed by the nember of dive	unitless
accession_IMG	The accession number in IMG	unitless
GOLD_Analysis_project_id	The project identification number in the GOLD Database; accession number for the metadata	unitless
GOLD_sequencing_strategy	The strategy used for sequencing (whole genome sequencing vs amplicon sequencing)	unitless
latitude	Latitude of the sampling site	unitless
longitude	Longitude of the sampling site	unitless
assembly_size	The total size of the assembled contigs	base pairs (bp)
gene_count	Number of identified genes	unitless
accession_link	Link to taxon_oid at IMG database	unitless
study_name	The name of the study	unitless
site	location where samples were collected	unitless
sequencing_center	the name of the center where the samples were sequenced	unitless
date_collect	date the samples were collected; formatted as month\day\year	unitless

[table of contents | back to top]

Instruments

Dataset-specific Instrument Name	
Generic Instrument Name	ROV Jason
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.

Dataset-specific Instrument Name	
Generic Instrument Name	Automated DNA Sequencer
Dataset-specific Description	MiSeq and NextSeq (Illumina) sequencers
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	IGT
Generic Instrument Name	Isobaric Gas-Tight Sampler
Generic Instrument Description	Isobaric Gas Tight (IGT) samplers, designed and built by scientists and engineers at WHOI, are titanium instruments designed to be used with deep submergence vehicles to sample corrosive hydrothermal vent fluids at high temperature and high pressure. The IGT prevents the sampled fluid from degassing as pressure decreases during the vehicle’s ascent to the surface.

[table of contents | back to top]

Deployments

AT26-10

Website	https://www.bco-dmo.org/deployment/529031
Platform	R/V Atlantis
Report	http://dmoserv3.bco-dmo.org/data_docs/Microbe_Vent_Communities/AT26-10_Cruise_Report_v2_2015-07-09.pdf
Start Date	2013-12-29
End Date	2014-01-27
Description	Samples were collected by ROV Jason II at the 9N deep-sea hydrothermal vent field on the East Pacific Rise, Pacific Ocean

AT15-38

Website	https://www.bco-dmo.org/deployment/660807
Platform	R/V Atlantis
Start Date	2008-10-13
End Date	2008-11-05

[table of contents | back to top]

Project Information

An Integrated Study of Energy Metabolism, Carbon Fixation, and Colonization Mechanisms in Chemosynthetic Microbial Communities at Deep-Sea Vents (Microbial Communities at Deep-Sea Vents)

Deep-sea hydrothermal vents, first discovered in 1977, are poster child ecosystems where microbial chemosynthesis rather than photosynthesis is the primary source of organic carbon. Significant gaps remain in our understanding of the underlying microbiology and biogeochemistry of these fascinating ecosystems. Missing are the identification of specific microorganisms mediating critical reactions in various geothermal systems, metabolic pathways used by the microbes, rates of the catalyzed reactions, amounts of organic carbon being produced, and the larger role of these ecosystems in global biogeochemical cycles. To fill these gaps, the investigators will conduct a 3-year interdisciplinary, international hypothesis-driven research program to understand microbial processes and their quantitative importance at deep-sea vents. Specifically, the investigators will address the following objectives: 1. Determine key relationships between the taxonomic, genetic and functional diversity, as well as the mechanisms of energy and carbon transfer, in deep-sea hydrothermal vent microbial communities. 2. Identify the predominant metabolic pathways and thus the main energy sources driving chemoautotrophic production in high and low temperature diffuse flow vents. 3. Determine energy conservation efficiency and rates of aerobic and anaerobic chemosynthetic primary productivity in high and low temperature diffuse flow vents. 4. Determine gene expression patterns in diffuse-flow vent microbial communities during attachment to substrates and the development of biofilms.

Integration: To address these objectives and to characterize the complexity of microbially-catalyzed processes at deep-sea vents at a qualitatively new level, we will pursue an integrated approach that couples an assessment of taxonomic diversity using cultivation-dependent and -independent approaches with methodologies that address genetic diversity, including a) metagenomics (genetic potential and diversity of community), b) single cell genomics (genetic potential and diversity of uncultured single cells), c) meta-transcriptomics and -proteomics (identification and function of active community members, realized potential of the community). To assess function and response to the environment, these approaches will be combined with 1) measurement of in situ rates of chemoautotrophic production, 2) geochemical characterization of microbial habitats, and 3) shipboard incubations under simulated in situ conditions (hypothesis testing under controlled physicochemical conditions). Network approaches and mathematical simulation will be used to reconstruct the metabolic network of the natural communities. A 3-day long project meeting towards the end of the second year will take place in Woods Hole. This Data Integration and Synthesis meeting will allow for progress reports and presentations from each PI, postdoc, and/or student, with the aim of synthesizing data generated to facilitate the preparation of manuscripts.

Intellectual Merit. Combining the community expression profile with diversity and metagenomic analyses as well as process and habitat characterization will be unique to hydrothermal vent microbiology. The approach will provide new insights into the functioning of deep-sea vent microbial communities and the constraints regulating the interactions between the microbes and their abiotic and biotic environment, ultimately enabling us to put these systems into a quantitative framework and thus a larger global context.

Broader Impacts. This is an interdisciplinary and collaborative effort between 4 US and 4 foreign institutions, creating unique opportunities for networking and fostering international collaborations. This will also benefit the involved students (2 graduate, several undergraduate) and 2 postdoctoral associates. This project will directly contribute to many educational and public outreach activities of the involved PIs, including the WHOI Dive & Discover program; single cell genomics workshops and Cafe Scientifique (Bigelow); REU (WHOI, Bigelow, CIW); COSEE and RIOS (Rutgers), and others. The proposed research fits with the focus of a number of multidisciplinary and international initiatives, in which PIs are active members (SCOR working group on Hydrothermal energy and the ocean carbon cycle, http://www.scorint.org/Working_Groups/wg135.htm; Deep Carbon Observatory at CIW, https://dco.gl.ciw.edu/; Global Biogeochemical Flux (GBF) component of the Ocean Observatories Initiative (OOI), http://www.whoi.edu/GBF-OOI/page.do?pid=41475)

[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.

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-1136488

[table of contents | back to top]