Sub-seafloor single amplified genomes (SAGs) from anaerobic Peru Margin sediment collected on R/V JOIDES Resolution cruise JRES-201 in 2002

Website: https://www.bco-dmo.org/dataset/637878
Data Type: Cruise Results
Version: 04 Feb 2016
Version Date: 2016-02-04

Project
» Studying genomic and population biology of dehalogenating Chloroflexi in deep sea sediments by single cell sorting and single cell genome amplification (Chloroflexi in deep sea sediments)

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

Contributors	Affiliation	Role
Spormann, Alfred M.	Stanford University	Principal Investigator
Rauch, Shannon	Woods Hole Oceanographic Institution (WHOI BCO-DMO)	BCO-DMO Data Manager

Table of Contents

Dataset Description

Sub-seafloor single amplified genomes (SAGs) from anaerobic Peru Margin sediment collected on the JOIDES Resolution Leg 201 at Ocean Drilling Program Site 1230.

Further data can be found at:
Kaster, A.K., et al. 2014. Single cell genomic study of Dehalococcoidetes species from deep-sea sediments of the Peruvian Margin. ISME J. 2014 Sep; 8(9): 1831–1842. doi: 10.1038/ismej.2014.24

Methods & Sampling

A frozen deep-sea sediment sample of the Peruvian Margin drill site 1230 (ODP 201), collected 7.3 meters below seafloor (mbsf) and stored at −80 degrees C without glycerol preservation for 8 years, was used for single cell genome analysis. Physical isolation of the single cells was performed by Fluorescent Activated Cell Sorting in two 384-well plates (630 single cells, 6 positive controls and 132 negative controls). The sample processing was performed at the Bigelow Laboratory Single Cell Genomics Center. Single cells were lysed, and the DNA was amplified by MDA. In all, 250 wells showed good amplification with a Cp value of <10 h (~40%). DNA was screened with broad eubacterial (27F-M13: 5'-AGRGTTYGATYMTGGCTCAG-3'/907R_degen-M13: 5'-CCGTCAATTCMTTTRAGTTT-3') and archaeal (Arc_344F-M13: 5'-ACGGGGYGCAGCAGGCGCGA-3'/Arch_915R-M13R: 5'-GTGCTCCCCCGCCAATTCCT-3') 16S rRNA primers and Sanger sequenced. Analysis with the RDB (Ribosomal Database) yielded 33 hits (5.2% of all single cells sorted, 13.2% of successful MDA reactions), including three *Chloroflexi* single cells that showed a 16S rRNA sequence by Sanger most similar to *Dehalogenimonas*. The first MDA products yielded 500–900 ng of DNA after clean up with the QIAamp DNA kit (Qiagen). The first MDA products of the three single cells were re-amplified in a second MDA. To avoid

additional bias, the second MDA was performed in four separate reactions that were subsequently combined at the end.

The first MDA products of the single cells were sequenced separately on an Illumina HighSeq platform (San Diego, CA, USA) using Nextera library preparation with an average yield of 15 000 Mb and 150 000 000 reads with 2 × 100 bp read length. The second MDA products were sequenced using the PacBio *RS*Magbead CLR sequencing technique (Menlo Park, CA, USA), resulting in a mean read length of over 2.5 Kb and ~100 Mb raw sequence data. Sequencing was carried out according to the manufacturer's instructions and resulted in 12 Mb raw sequence data for single cell 1 and 190 Mb for single cells 2 and 3.

Data Processing Description

Different strategies were applied to assemble the reads of the individual cells, and later to combine single cells Dsc # 2 and # 3, in order to get the most out of the sequencing data. Statistics were checked with assemblathon. Since good assembly statistics do not automatically hold true that the assembly is optimal assemblies were always run through the RAST pipeline to check for misassemblies. SAGs were assembled by:

A. CLC bio
B. spades 2.3
C. spades-n
D. velvet-sc, kmer=37
E. velvet-sc n
F. Celera (CA)
G. Hybrid error correction method using CA assembled Illumina® data to correct long PacBio® reads
H. velvet assembly using Euler correction, kmer=55
I. spades assembly of Illumina®-only combined via PCAP with CA assembly of PacBio corrected by PacBio only
J. velvet-sc assembly of Illumina®-only combined via PCAP with CA assembly of PacBio® corrected by PacBio® only
K. velvet-sc assembly of Illumina®-only combined via PCAP with CA assembly of PacBio® corrected by Illumina®-only
L. spades assembly of Illumina®-only combined via PCAP with CA assembly of PacBio corrected by Illumina® only n = Normalization of the Illumina® reads

Single cells 2 and 3 were assembled together since they showed almost 100% identity at the nucleotide level after individual assembly. At this stage, a 0.32-Mb assembly was contained in 126 contigs for single cell 1 (Dsc1) and a 1.38-Mb assembly in 327 contigs for the co-assembly of single cells 2 and 3 (DscP2).

Assembled contigs were submitted to the Integrated Microbial Genomes database annotation pipeline (IMG, version 4.1) and to the Rapid Annotations using Subsystems Technology pipeline (RAST, version 4.0) in 2013. Some computationally assigned annotations were manually changed based on the inspection of evidence for the assigned annotations, orthologs in related genomes and gene neighborhoods. Pathways were predicted using RAST, IMG and KEGG (Kyoto Encyclopedia of Genes and Genomes). Nucleotide and amino-acid sequences of genes were blasted as query sequences against the NCBI databases.

[table of contents | back to top]

Data Files

File
SAGs.csv (Comma Separated Values (.csv), 777 bytes) MD5:9924b5c9c8609956fcceb9dcef79ed59 Primary data file for dataset ID 637878

[table of contents | back to top]

Parameters

Parameter	Description	Units
organism	Organism studied.	dimensionless
cruise_id	Cruise identifier.	dimensionless
BioProject_ID	NCBI BioProject identification number.	dimensionless
accession_num	NCBI accession number.	dimensionless
BioProject_link	Hyperlink to NCBI BioProject.	dimensionless
accession_link	Hyperlink to NCBI accession number.	dimensionless

[table of contents | back to top]

Deployments

JRES-201

Website	https://www.bco-dmo.org/deployment/626163
Platform	R/V JOIDES Resolution
Report	http://dmoserv3.whoi.edu/data_docs/C-DEBI/cruise_reports/201PREL-1.pdf
Start Date	2002-01-27
End Date	2002-03-29
Description	Leg 201 Controls on Microbial Communities in Deeply Buried Sediments, Eastern Equatorial Pacific and Peru Margin Sites 1225-1231 27 January-29 March 2002 Cruise report obtained from http://www-odp.tamu.edu/publications/pubs.htm

[table of contents | back to top]

Project Information

Studying genomic and population biology of dehalogenating Chloroflexi in deep sea sediments by single cell sorting and single cell genome amplification (Chloroflexi in deep sea sediments)

Project description from C-DEBI:
Dehalogenating Chloroflexi, such as Dehalococcoidites (Dhc) were originally discovered as the key microorganisms mediating reductive dehalogenation of the prevalent groundwater contaminants tetrachloroethene and trichloroethene. Molecular and genomic studies on their key enzymes for energy conservation, reductive dehalogenases (rdh), have provided evidence for ubiquitous horizontal gene transfer. A pioneering study by Futagami et al. discovered novel putative rdh phylotypes in sediments from the Pacific, revealing an unknown and surprising abundance of rdh genes in pristine habitats. The frequent detection of Dhc-related 16S rRNA genes from these environments implied the occurrence of dissimilatory dehalorespiration in marine subsurface sediments. Despite being ubiquitous in those environments, metabolic life style or ecological function of Dhc in the absence of anthropogenic contaminants is still completely unknown. We therefore analyzed a non-contaminated deep sea sediment sample of the Peru Margin 1230 site

by a single cell genomic (SGC) approach. We were able get for the first time data on three single Dhc cells, helping to elucidate their role in the poorly understood oligotrophic marine sub-surface environment. Although all three single cells show the majority of their best Blast hits to Dhc species only one putative reductive dehalogenase was discovered, with very weak similarity to other known sequences. One of the reasons might be the incompleteness of the genome and rdh genes might have been missed. Another possibility is that deep sea Dhc are not halorespirers like their terrestrial relatives. Interestingly, when screening the DNA of other single cells, PCR shows a positive match for a rdh sequence in Firmicutes. This was quite an unexpected twist of the project.

This project was funded by a C-DEBI Postdoctoral fellowship awarded to Anne-Kristin Kaster.

[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]