16S rRNA gene datasets, metagenomes, and metatranscriptomes from cores collected on R/V Knorr cruise KN223 in the North and West Atlantic Ocean from October to December 2014

Website: https://www.bco-dmo.org/dataset/798144
Data Type: Cruise Results, experimental
Version: 1
Version Date: 2020-02-28

Project
» Microbial activity in oxygenated subseafloor sediment (Microbial activity subseafloor sediment)

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

Contributors	Affiliation	Role
Orsi, William D.	University of Munich	Principal Investigator
Rauch, Shannon	Woods Hole Oceanographic Institution (WHOI BCO-DMO)	BCO-DMO Data Manager

Abstract
16S rRNA gene datasets, metagenomes, and metatranscriptomes from cores collected on R/V Knorr cruise KN223 in the North and West Atlantic Ocean from October to December 2014.

Table of Contents

Dataset Description

All data is publicly available as described in NCBI at the following link (NCBI BioProject PRJNA473406): https://www.ncbi.nlm.nih.gov/bioproject/PRJNA473406

Methods & Sampling

Methodology: We optimized a DNA extraction protocol for ultra-organic lean sediments that provided increased yields of DNA, enabling sequencing of 16S rRNA and ammonia monooxygenase [amoA] genes, as well as metagenomes, from two deep oxic subseafloor sediment cores reaching up to ~15 million years old. In addition, we demonstrate the viability of uncultivated microbial populations via ^{18}O-labeling in long-term (18-month) incubations in the presence of $H_2{}^{18}O$, a method for identifying actively growing microbes.

Sampling and analytical procedures: All samples were taken by Cruise KN223 of the R/V Knorr, in the North Atlantic, from October 26th 2014 - December 2014 (Woods Hole, MA - Woods Hole, MA). At both Site 11 (22° 47.0'N, 56° 31.0' W, water depth ~5,600 m) and Site 12 (29° 40.6' N, 58° 19.7' W, water depth ~5,400 m), successively deeper subseafloor cores were taken with a multicorer [to ~0.4 meters below seafloor (mbsf)], a gravity corer (to ~3 mbsf), and the 45-m WHOI "long corer" piston-coring device

(https://www2.whoi.edu/site/longcore/) (to ~28 mbsf). Subsamples from the core sections for DNA extraction were sampled on board the ship immediately after retrieval with sterile 60 mL syringes with the Luer taper end cut off and frozen immediately at -80 °C prior to DNA extraction. Subsamples for the 18O-labeling experiment were sampled from the core sections in the same manner, but were stored at +4 °C prior to incubation set up.

DNA extraction: Subcores sampled aseptically with sterile syringes were sampled aseptically in a UV-sterilized DNA/RNA clean HEPA filtered laminar flow hood. To reduce contamination, the paraffin was removed and the outer 3 cm of sediment pushed out of the syringe, which was then sliced off with a red-hot, sterile spatula. A second unused, sterile, spatula was used to carefully sample the uncontaminated center of the remaining core sample inside the syringe. In brief, 10 g of sediment was transferred into 50 mL Lysing Matrix E tubes (MP Biomedicals) containing silica glass beads and homogenized for 40 sec at 6 m/s using a FastPrep 24 5-G homogenizer (MP Biomedicals) in the presence of 15 mL preheated (65 C) sterile filtered extraction buffer (76 vol% vol 1M $NaPO_4$ pH 8, 15 vol% 200 proof ethanol, 8 vol% MoBio's lysis buffer solution C1, and 1 vol% SDS). The samples were incubated at 99°C for 2 minutes and frozen overnight at -20°C, thawed, and frozen again at -20°C overnight, followed by additional incubation at 99°C for 2 minutes and a second homogenization using the settings described above. The additional freeze thaw steps, particularly freezing overnight, was found to increase DNA yield 2–10 fold. After the second homogenization, the samples were centrifuged for 15 minutes, and the supernatants concentrated to a volume of 100 uL using 50 KDa Amicon centrifugal filters (Millipore). Co-extracted PCR-inhibiting humic acids and other compounds were removed from the concentrated extract using the PowerClean Pro DNA Clean-up Kit (MoBio). Extraction blanks were performed alongside the samples to assess laboratory contamination during the extraction process.

qPCR: DNA was quantified fluorometrically using a Qubit with a dsDNA high-sensititivity kit (Life Technologies). Quantitative PCR (qPCR) was performed using the custom primer dual indexed approach that targets the V4 hypervariable region of the 16S rRNA gene using updated 16S rRNA gene primers 515F/806R (515F: 5' –GTGYCAGCMGCCGCGGTAA– 3', 806R: GGACTACNVGGGTWTCTAAT) that increase coverage of ammonia oxidizing Thaumarchaea and other marine strains(*59*). To measure the abundance of amoA genes from archaea, the primers Arch amoA-1F (STAATGGTCTGGCTTAGACG) and Arch amoA-2R (GCGGCCATCCATCTGTATGT) were used. qPCR reactions were prepared using an automated liquid handler (pipetting robot), the EpMotion 5070 (Eppendorf), was used to set up all qPCR reactions and standard curves. The efficiency values of the qPCR was <90% and R^2 values >0.95. qPCR was performed using white 96-well plates as this was found to increase the signal to noise in the SYBR green assay 2-fold compared to clear plates. The technical variability of 16S rRNA gene qPCR measurements was determined to be consistently <5% under the the EpMotion 5070.

16S rRNA and amoA gene sequencing: Barcoded V4 hypervariable regions of amplified 16S rRNA genes were sequenced on an Illumina MiniSeq. This yielded a total of >20,000,000 raw sequencing reads that were then subjected to quality control. In order to quality control the OTU picking algorithm for the data, we also sequenced a "mock community" alongside our environmental samples. The mock communities contained a defined number of species (n=18) all containing 16S rRNA genes >3% difference. USEARCH version 10.0.240 was used for quality control and OTU picking (*61*), OTUs were clustered at 97% sequence identity. The taxonomic relationship of OTU representative sequences were identified by BLASTn searches against SILVA database (www.arb-silva.de) version 128. To identify contaminants, 16S rRNA genes from extraction blanks and dust samples from the lab were also sequenced. These 16S rRNA gene sequences from contaminants were used to identify any contaminating bacteria in our oxic abyssal clay samples. All OTUs containing sequences from these 'contaminant' samples were removed prior to downstream analysis.

qPCR of 16S rRNA genes in DNA extraction blanks were consistently $<10^2$ copies per extraction, and thus we used 10^2 copies to define our detection limit for the abyssal clay samples. Consistent with this, high-throughput sequencing of amplicons with qPCR values $<10^2$ copies per g sediment had up to 50% sequence representation from contaminant taxa, whereas samples with values $>10^2$ copies per g sediment had <5% representation from contaminant taxa. This further supported our definition of $<10^2$ as a realistic detection limit. Using samples that had 16S rRNA gene copies $>10^2$ copies per g sediment, we were able to analyze microbial communities down to ca. 15 mbsf at Site 11 and ca. 8 mbsf at Site 12.

Thaumarchaeal amoA genes amplified via qPCR using the method described above were cloned using the TOPO TA cloning kit, and Sanger sequenced at the LMU Munich Sequencing Service at the Faculty of Biology (http://www.gi.bio.lmu.de/sequencing). Prior to phylogenetic analysis, the reads were quality trimmed in CLC Genomics using the default settings for quality control.

Experimental setups: DNA-SIP experiments with $H_2{}^{18}O$ were set up at two North Atlantic coring sites: Site 11 and Site 12, from depths 2.8 and 3.5 mbsf, respectively. Prior to setting up the incubations, the subcores were sampled with sterile syringes using the sample aseptic technique used for the DNA extraction. For each sample depth, seven grams of abyssal clay was placed into sterile 20-mL glass flasks and incubated with 4 mL of

sterile artificial seawater composed of either $H_2{}^{18}O$ (97% atomic enrichment) or unlabeled artificial seawater. Vials were crimp sealed, with an oxygenated headspace of approximately 10 mL, and incubated at 8 C. The water content of the clay was measured to be approximately 40% (+/- 5%) of the total weight. This initial water content diluted the final concentration of added $H_2{}^{18}O$ to be ca. 60% of the total water content of the sample. The artificial seawater was different from the porewater at depth because there was no added nitrate, but there was also no added ammonia which should be similar to the in situ conditions where ammonia is generally below detection. Oxygen was measured continuously throughout the incubations using non-invasive fiberoptic measurements as described previously. Small fluctuations in the oxygen measurements in the killed control, and experimental incubations (Fig S3), were likely due to temperature fluctuations of the incubator itself (1°C), since the non-invasive fiber optic oxygen sensor spots are temperature sensitive.

To assess the preservation potential of extracellular DNA (eDNA), and its ability to bias our study that is based on DNA from living organisms, we monitored microbial growth in the presence and absence of added eDNA over a 210-day incubation experiment. eDNA extracted from a culture of *Rhodococcus erythropolis* was added to sediment slurries from 2.8 mbsf at Site 11 at a concentration of 5 ng g^{-1}. Microbial growth was measured over time with 16S rRNA gene qPCR, and also in a control that did not receive the eDNA. As a second control, we added eDNA to autoclaved (dead) sediment. DNA was extracted from each timepoint and measure with qPCR using the methods described above.

Density gradient fractionation, qPCR: We used Tag-SIP to measure the atom % ^{18}O-enrichment of actively growing microbial taxa. In brief, after 7 and 18 months incubations DNA was extracted and subjected to Cesium Chloride (CsCl) density gradient centrifugation as described previously. The same 16S 515F/806R primers (described above) were used in qPCR (described above) to determine density shifts in the peak DNA of buoyant density (BD) for each incubation. 16S rRNA gene amplicons from each fractions resulting from the density gradient fractionation were Illumina sequenced as described above. To identify contaminants that may have entered during the fractionation process, we also included in the sequencing run extraction blanks from the SIP fractionation. OTUs containing sequences from extraction blanks were removed.

Metagenome library preparation, sequencing and bioinformatics analysis: Prior to library preparation, whole genome amplifications were performed on DNA extracts through a multiple displacement amplification (MDA) step of 6 to 7 hours, using the REPLI-g Midi Kit (QIAGEN) and following the manufacturer's instructions. We monitored amplification using SYBR green I (Invitrogen) on a CFX-Connect qPCR machine, stopping amplifications once the exponential phase was reached. Metagenomic libraries were prepared using the Nextera XT DNA Library Prep Kit (Illumina). Quality control and quantification of the libraries were obtained on an Agilent 2100 Bioanalyzer System, using the High Sensitivity DNA reagents and DNA chips (Agilent Genomics). was used to normalize library DNA concentrations. Metagenomic libraries were diluted to 1 nM using the Select-a-Size DNA Clean and Concentrator MagBead Kit (Zymo Research), and pooled for further sequencing on the Illumina MiniSeq platform.

Data Processing Description

Data processing: Contigs were assembled on CLC Genomics Workbench v. 9.5.4 (QIAGEN), using a word size = 20, bubble size = 50, and a minimum contig length of 300 nucleotides. Reads were then mapped to the contigs using the following parameters (mismatch penalty = 3, insertion penalty = 3, deletion penalty = 3, minimum alignment length = 50% of read length, minimum percent identity = 95%). We then performed even further stringency controls, by removing any contig that had less than 5x coverage, e.g. reads per kilobase mapped (RPKM). The final resulting dataset of contigs was then used for open reading frame (ORF) searches and BLAST analysis. Protein encoding genes and open reading frames (ORFs) were extracted using FragGeneScan v. 1.30. Cut-off values for assigning hits to specific taxa were performed at a minimum bit score of 50, minimum amino acid similarity of 60, and an alignment length of 50 residues.

For phylogenetic analyses, OTUs of AOAs were aligned with SINA online v.1.2.11 and plotted in ARB against the SILVA 16S rRNA SSU NR99 reference database release 132. Closest environmental sequences with nearly full-length sequences (>1400 bp) were selected as taxonomic references and used to calculate trees using the Maximum Likelihood algorithm RAxML implemented with the archaeal filter and advanced bootstrap refinement selecting the best tree among 100 replicates. Partial OTU sequences were added to the tree using the maximum parsimony algorithm without allowing changes of tree topology. Statistical analyses of beta-diversity were performed using R.Studio Version 3.3.0 with the vegan package.

[table of contents | back to top]

Related Publications

Vuillemin, A., Wankel, S. D., Coskun, Ö. K., Magritsch, T., Vargas, S., Estes, E. R., ... Orsi, W. D. (2019). Archaea dominate oxic subseafloor communities over multimillion-year time scales. Science Advances, 5(6), eaaw4108. doi:10.1126/sciadv.aaw4108
Results

[table of contents | back to top]

Parameters

Parameters for this dataset have not yet been identified

[table of contents | back to top]

Instruments

Dataset-specific Instrument Name	Illumina MiniSeq
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	FastPrep 24 5-G homogenizer (MP Biomedicals)
Generic Instrument Name	Homogenizer
Generic Instrument Description	A homogenizer is a piece of laboratory equipment used for the homogenization of various types of material, such as tissue, plant, food, soil, and many others.

Dataset-specific Instrument Name	
Generic Instrument Name	Multi Corer
Generic Instrument Description	The Multi Corer is a benthic coring device used to collect multiple, simultaneous, undisturbed sediment/water samples from the seafloor. Multiple coring tubes with varying sampling capacity depending on tube dimensions are mounted in a frame designed to sample the deep ocean seafloor. For more information, see Barnett et al. (1984) in Oceanologica Acta, 7, pp. 399-408.

Dataset-specific Instrument Name	
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]

Deployments

KN223

Website	https://www.bco-dmo.org/deployment/567408
Platform	R/V Knorr
Start Date	2014-10-25
End Date	2014-12-02

[table of contents | back to top]

Project Information

Microbial activity in oxygenated subseafloor sediment (Microbial activity subseafloor sediment)

Website: https://www.darkenergybiosphere.org/award/microbial-activity-in-oxygenated-subseafloor-sediment/

Coverage: North Atlantic, Sargasso Sea Basin

The subseafloor sedimentary biosphere is the largest ecosystem on Earth, where microbes subsist under energy-limited conditions over long timescales. It is poorly understood how activity is converted into microbial reproduction and survival under these conditions. Here, we examine this question in deep-sea subseafloor communities subsisting in oxic and anoxic abyssal sediments for over multimillion year timescales. Ammonia-oxidizing Archaea (AOA) dominate oxic abyssal microbial communities by up to an order of magnitude for 15 million years in the oxic subseafloor of the North Atlantic Ocean. Rates of nitrification correlated with the abundance of these dominant AOA populations, whose metabolism is characterized by ammonia oxidation, mixotrophic utilization of organic nitrogen, deamination, and the energetically efficient chemolithoautotrophic hydroxypropionate/ hydroxybutyrate carbon fixation cycle. These AOA thus have the potential to couple mixotrophic and chemolithoautotrophic metabolism via mixotrophic deamination of organic nitrogen, followed by oxidation of the regenerated ammonia for additional energy to fuel carbon fixation. This metabolic feature likely reduces energy loss and improves AOA fitness under energy starved, oxic conditions, thereby allowing them to outcompete other taxa for millions of years. In abyssal anoxic sites, a single population affiliated with the candidate Phylum "Candidatus Atribacteria" dominates the subseafloor community for up to 8 million years. Expression of genes encoding proteins for cell division were detected only in the upper 10 mbsf, indicating increased abundances of "Ca. Atribacteria" were due to actively reproducing microbes. Mean net sulfate reduction rate is relatively high over the upper 10-meter interval. At greater depths, the ecosystem is subject to net death, with mean net sulfate reduction rate below detection, microbial abundance steadily declining, and no detectable expression of cell division genes. Even in this net-death interval, "Ca. Atribacteria" dominates the subseafloor community. The transcriptomes indicate that "Ca. Atribacteria" is homoacetogenic, utilizing electron bifurcation to couple fermentative H2 production from sugars with the Wood-Ljungdahl carbon fixation pathway. Additional reducing power for ATP synthesis appears to be gained by secondary fermentations via a bacterial micro-compartment. This energy-efficient metabolism apparently improves the fitness of "Ca. Atribacteria" in this energy-limited setting, allowing this group to dominate communities over multimillion-year timescales.

Additional project information is available from C-DEBI: https://www.darkenergybiosphere.org/award/microbial-

activity-in-oxygenated-subseafloor-sediment/

[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]