Supplementary Table 4B: Annotations for contigs within transcriptome libraries for the eleven samples that were manually curated for selected metabolic processes

Website: https://www.bco-dmo.org/dataset/812997
Data Type: Other Field Results
Version: 1
Version Date: 2020-06-22

Project
» Collaborative Research: Delineating The Microbial Diversity and Cross-domain Interactions in The Uncharted Subseafloor Lower Crust Using Meta-omics and Culturing Approaches (Subseafloor Lower Crust Microbiology)

Program
» International Ocean Discovery Program (IODP)

Contributors	Affiliation	Role
Edgcomb, Virginia P.	Woods Hole Oceanographic Institution (WHOI)	Principal Investigator, Contact
Soenen, Karen	Woods Hole Oceanographic Institution (WHOI BCO-DMO)	BCO-DMO Data Manager

Abstract

Supplementary Table 4B: Metatranscriptome data summary for cellular activities presented and statistics on sequencing and removal of potential contaminant sequences: Annotations for contigs within transcriptome libraries for the eleven samples that were manually curated for selected metabolic processes. Samples taken on board of the R/V JOIDES Resolution between November 30, 2015 and January 30, 2016.

Table of Contents

Coverage

Spatial Extent: **Lat**:-32.70567 **Lon**:57.278183
Temporal Extent: 2015-11-30 - 2016-01-30

Dataset Description

Supplementary Table 4B: Metatranscriptome data summary for cellular activities presented and statistics on sequencing and removal of potential contaminant sequences: Annotations for contigs within transcriptome libraries for the eleven samples that were manually curated for selected metabolic processes. Samples taken on board of the R/V JOIDES Resolution between November 30, 2015 and January 30, 2016.

Acquisition Description

Frozen rock material was crushed as above, and then ground quickly into a fine powder using a precooled sterilized mortar and pestle, and then RNA extraction started immediately. The jaw crusher was cleaned and rinsed with 70% ethanol and RNaseZap™ RNase Decontamination Solution (Invitrogen, USA) between samples. About 40 g of material was extracted for each sample using the RNeasy PowerSoil Total RNA Isolation Kit (Qiagen, USA) according to the manufacturer's protocol with the following modifications.

Each sample was evenly divided into 8 Bead Tubes (Qiagen, USA) and then 2.5 mL of Bead Solution were added into the Bead Tube followed by 0.25 mL of Solution SR1 and 0.8 mL of Solution SR2. Bead Tubes were frozen in liquid nitrogen and then thawed at 65°C in a water bath three times. RNA was purified using the MEGAclear Transcription Clean-up Kit (Ambion, USA) and concentrated with an overnight isopropanol precipitation at 4 °C. Trace amounts of contaminating DNA were removed from the RNA extracts using TURBO DNA free™ (Invitrogen, USA) as directed by the manufacturer.

To ensure DNA was removed thoroughly, each RNA extract was treated twice with TURBO DNase (Invitrogen, USA). A nested PCR reaction (2 x 35 cycles) using bacterial primers was used to confirm the absence of DNA in our RNA solutions. RNA was converted to cDNA using the Ovation® RNA-Seq System V2 kit (NuGEN, USA) according to the manufacturer's protocol to preferentially prime non-rRNA sequences. The cDNA was purified with the MinElute Reaction Cleanup Kit (Qiagen, USA) and eluted into 20 µL elution buffer. Extracts were quantified using a Qubit Fluorometer (Life Technologies, USA) and cDNAs were stored at -80 °C until sequencing using 150 bp paired-end Illumina NextSeq 550.

To control for potential contaminants introduced during drilling, sample handling, and laboratory kit reagents, we sequenced a number of control samples as above. Two samples controlled for potential nucleic acid contamination; a "method" control to monitor possible contamination from our laboratory extractions, which included ~ 40 g sterilized glass beads processed through the entire protocol in place of rock, and a "kit" control to account for any signal coming from trace contaminants in kit reagents, which received no addition. In addition, 3 more controls were extracted: a sample of the drilling mud (Sepiolite), and two drilling seawater samples collected during the first and third weeks of drilling. cDNA obtained from these controls were sequenced together with the rock samples and co-assembled. Trimmomatic (v. 0.32) was used to trim adapter sequences (leading=20, trailing=20, sliding window=04:24, minlen=50). Paired reads were further quality checked and trimmed using FastQC (v. 0.11.7) and FASTX-toolkit (v. 0.014). Downstream analyses utilized paired reads. After co-assembling reads with Trinity (v. 2.4.0) from all controls (min length 150 bp), Bowtie2 (v. 2.3.4.1, 50) was used (with the parameter 'un-conc') to align all sample reads to this co-assembly. Reads that mapped to our control co-assembly allowing 1 mismatch were removed from further analysis (23.5-68.5% of sequences remained in sample data sets, see Supplementary Table 4).

Trinity (v. 2.4.0) was used for de novo assembly of the remaining reads in sample data sets (min. length 150 bp). Bowtie aligner was used to align reads to assembled contigs, RSEM was used to estimate the expression level of these reads, and TMM was used to perform cross sample normalization and to generate a TMM-normalized expression matrix. Within the Trinotate suite, TransDecoder (v. 3.0.1) was used to identify coding regions within contigs and functional and taxonomic annotation was made 622 by BLASTx and BLASTp against UniProt, Swissprot (release 2018_02) and RefSeq non-redundant protein sequence (nr) databases (e-value threshold of 1e-5). BLASTp was used to look for sequence homologies with the same e-values. HMMER (v. 3.1b2) was used to identify conserved domains by searching against the Pfam (v31.0) database. SignalP (v. 4.1) and TMHMM (2.0c) were used to predict signal peptides and transmembrane domains. RNAMMER (v.1.2) was used to identify rRNA homologies of archaea, bacteria and eukaryotes.

Because the Swissprot database does not have extensive representation of protein sequences from environmental samples, particularly deep-sea and deep biosphere samples, annotations of contigs utilized for analyses of selected processes were manually cross checked by BLASTx against GenBank nr database. Aside from removing any reads that mapped well to our control co-assembly (1 mismatch), as an extra precaution, any sequence that exhibited ≥ 95% sequence identity over ≥ 80% of the sequence length to suspected contaminants (e.g., human pathogens, plants, or taxa known to be common molecular kit reagent contaminants, and not described from the marine environment) as in Salter et al. and Glassing et al. were removed. This conservative approach potentially removed environmentally relevant data that were annotated to suspected contaminants due to poor taxonomic representation from environmental taxa in public databases, however it affords the highest possible confidence about any transcripts discussed.

Additional functional annotations of contigs were obtained by BLAST against the KEGG, COG, SEED, and MetaCyc databases using MetaPathways (v. 2.0) to gain insights into particular cellularprocesses, and to provide overviews of metabolic functions across samples based on comparisons of FPKM-normalized data. All annotations were integrated into a SQLite database for further analysis.

Processing Description

BCO-DMO processing notes:

- Reformatted table structure
- Adjusted column header names to comply with database requirements
- Replaced , with ; in data table to comply with database requirements

[<u>table of contents</u> | <u>back to top</u>]

Data Files

File	Version
Data file associated with dataset 812997 - Annotations filename: annotations.csv (Comma Separated Values (.csv), 29.22 KB) MD5:54f6798a5157d0b8bbdfa05f8e48e697 *A .csv data file associated with dataset 812997: Supplementary Table 4B - Annotations for contigs within transcriptome libraries for the eleven samples that were manually curated for selected metabolic processes.*	1

[<u>table of contents</u> | <u>back to top</u>]

Related Publications

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics, 30(15), 2114–2120. doi:<u>10.1093/bioinformatics/btu170</u>
Methods

Cole, J. R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R. J., … Tiedje, J. M. (2009). The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Research, 37(Database), D141–D145. doi:<u>10.1093/nar/gkn879</u>
Methods

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., … Bateman, A. (2015). The Pfam protein families database: towards a more sustainable future. Nucleic Acids Research, 44(D1), D279–D285. doi:<u>10.1093/nar/gkv1344</u>
Methods

Glassing, A., Dowd, S. E., Galandiuk, S., Davis, B., & Chiodini, R. J. (2016). Inherent bacterial DNA contamination of extraction and sequencing reagents may affect interpretation of microbiota in low

bacterial biomass samples. Gut Pathogens, 8(1). doi:10.1186/s13099-016-0103-7
Methods

Krogh, A., Larsson, B., von Heijne, G., & Sonnhammer, E. L. . (2001). Predicting transmembrane protein topology with a hidden markov model: application to complete genomes11Edited by F. Cohen. Journal of Molecular Biology, 305(3), 567–580. doi:10.1006/jmbi.2000.4315
Methods

Lagesen, K., Hallin, P., Rødland, E. A., Stærfeldt, H.-H., Rognes, T., & Ussery, D. W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Research, 35(9), 3100–3108. doi:10.1093/nar/gkm160
Methods

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. Nature Methods, 9(4), 357–359. doi:10.1038/nmeth.1923
Methods

Li, B., & Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics, 12(1). doi:10.1186/1471-2105-12-323
Methods

Petersen, T. N., Brunak, S., von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature Methods, 8(10), 785–786. doi:10.1038/nmeth.1701
Methods

Salter, S. J., Cox, M. J., Turek, E. M., Calus, S. T., Cookson, W. O., Moffatt, M. F., … Walker, A. W. (2014). Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biology, 12(1). doi:10.1186/s12915-014-0087-z
Methods

[table of contents | back to top]

Parameters

Parameter	Description	Units
Cycle	Cycle of the biosynthetic pathway	unitless
Biosynthetic_pathway	Name of biosynthetic pathway	unitless
ID_2R1	Taxonomic annotations per pathway for sample 2R1	unitless
ID_19R1	Taxonomic annotations per pathway for sample 19R1	unitless
ID_26R2	Taxonomic annotations per pathway for sample 26R2	unitless
ID_31R1	Taxonomic annotations per pathway for sample 31R1	unitless
ID_42R2	Taxonomic annotations per pathway for sample 42R2	unitless
ID_51R3	Taxonomic annotations per pathway for sample 51R3	unitless
ID_62R1	Taxonomic annotations per pathway for sample 62R1	unitless
ID_68R4	Taxonomic annotations per pathway for sample 68R4	unitless
ID_71R1	Taxonomic annotations per pathway for sample 71R1	unitless
ID_81R2	Taxonomic annotations per pathway for sample 81R2	unitless
ID_84R6	Taxonomic annotations per pathway for sample 84R6	unitless

[table of contents | back to top]

Instruments

Dataset-specific Instrument Name	Illumina NextSeq 550 platform
Generic Instrument Name	Automated DNA Sequencer
Dataset-specific Description	RNA sequencing was performed using the Illumina NextSeq 550 platform (Univ. of Georgia).
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.

Deployments

IODP-360

Website	https://www.bco-dmo.org/deployment/810905
Platform	R/V JOIDES Resolution
Report	http://publications.iodp.org/scientific_prospectus/360/index.html
Start Date	2015-11-30
End Date	2016-01-30

Project Information

Collaborative Research: Delineating The Microbial Diversity and Cross-domain Interactions in The Uncharted Subseafloor Lower Crust Using Meta-omics and Culturing Approaches (Subseafloor Lower Crust Microbiology)

Coverage: SW Indian Ridge, Indian Ocean

NSF abstract: The lower ocean crust has remained largely unexplored and represents one of the last frontiers for biological exploration on Earth. Preliminary data indicate an active subsurface biosphere in samples of the lower oceanic crust collected from Atlantis Bank in the SW Indian Ocean as deep as 790 m below the seafloor. Even if life exists in only a fraction of the habitable volume where temperatures permit and fluid flow can deliver carbon and energy sources, an active lower oceanic crust biosphere would have implications for deep carbon budgets and yield insights into microbiota that may have existed on early Earth. This is all of great interest to other research disciplines, educators, and students alike. A K-12

education program will capitalize on groundwork laid by outreach collaborator, A. Martinez, a 7th grade teacher in Eagle Pass, TX, who sailed as outreach expert on Drilling Expedition 360. Martinez works at a Title 1 school with ~98% Hispanic and ~2% Native American students and a high number of English Language Learners and migrants. Annual school visits occur during which the project investigators present hands on-activities introducing students to microbiology, and talks on marine microbiology, the project, and how to pursue science related careers. In addition, monthly Skype meetings with students and PIs update them on project progress. Students travel to the University of Texas Marine Science Institute annually, where they get a campus tour and a 3-hour cruise on the R/V Katy, during which they learn about and help with different oceanographic sampling approaches. The project partially supports two graduate students, a Woods Hole undergraduate summer student, the participation of multiple Texas A+M undergraduate students, and 3 principal investigators at two institutions, including one early career researcher who has not previously received NSF support of his own. Given the dearth of knowledge of the lower oceanic crust, this project is poised to transform our understanding of life in this vast environment. The project assesses metabolic functions within all three domains of life in this crustal biosphere, with a focus on nutrient cycling and evaluation of connections to other deep marine microbial habitats. The lower ocean crust represents a potentially vast biosphere whose microbial constituents and the biogeochemical cycles they mediate are likely linked to deep ocean processes through faulting and subsurface fluid flow. Atlantis Bank represents a tectonic window that exposes lower oceanic crust directly at the seafloor. This enables seafloor drilling and research on an environment that can transform our understanding of connections between the deep subseafloor biosphere and the rest of the ocean. Preliminary analysis of recovered rocks from Expedition 360 suggests the interaction of seawater with the lower oceanic crust creates varied geochemical conditions capable of supporting diverse microbial life by providing nutrients and chemical energy. This project is the first interdisciplinary investigation of the microbiology of all 3 domains of life in basement samples that combines diversity and "meta-omics" analyses, analysis of nutrient addition experiments, high-throughput culturing and physiological analyses of isolates, including evaluation of their ability to utilize specific carbon sources, Raman spectroscopy, and lipid biomarker analyses. Comparative genomics are used to compare genes and pathways relevant to carbon cycling in these samples to data from published studies of other deep-sea environments. The collected samples present a rare and time-sensitive opportunity to gain detailed insights into microbial life, available carbon and energy sources for this life, and of dispersal of microbiota and connections in biogeochemical processes between the lower oceanic crust and the overlying aphotic water column. About the study area: The International Ocean Discovery Program (IODP) Expedition 360 explored the lower crust at Atlantis Bank, a 12 Ma oceanic core complex on the ultraslow-spreading SW Indian Ridge. This oceanic core complex represents a tectonic window that exposes lower oceanic crust and mantle directly at the seafloor, and the expedition provided an unprecedented opportunity to access this habitat in the Indian Ocean.

Program Information

International Ocean Discovery Program (IODP)

Website: http://www.iodp.org/index.php

Coverage: Global

The International Ocean Discovery Program (IODP) is an international marine research collaboration that explores Earth's history and dynamics using ocean-going research platforms to recover data recorded in seafloor sediments and rocks and to monitor subseafloor environments. IODP depends on facilities funded by three platform providers with financial contributions from five additional partner agencies. Together, these entities represent 26 nations whose scientists are selected to staff IODP research expeditions conducted throughout the world's oceans. IODP expeditions are developed from hypothesis-driven science proposals aligned with the program's science plan Illuminating Earth's Past, Present, and Future. The

science plan identifies 14 challenge questions in the four areas of climate change, deep life, planetary dynamics, and geohazards. IODP's three platform providers include: The U.S. National Science Foundation (NSF) Japan's Ministry of Education, Culture, Sports, Science and Technology (MEXT) The European Consortium for Ocean Research Drilling (ECORD) More information on IODP, including the Science Plan and Policies/Procedures, can be found on their website at http://www.iodp.org/program-documents.

Funding

Funding Source	Award
NSF Division of Ocean Sciences (NSF OCE)	OCE-1658031