Edginess in the subsurface: Microbial diversity of deep subseafloor ecotones

Data Policy Compliance

Identify any published data policies with which the project will comply, including the NSF OCE Data and Sample Policy as well as other policies that may be relevant if the project is part of a large coordinated research program (e.g. GEOTRACES).

Center for Dark Energy Biosphere Investigations data policy (https://www.darkenergybiosphere.org/about-our-center/data/)

Pre-Cruise Planning

If the proposed project involves a research cruise, describe the cruise plans. (Skip this section if it is not relevant to your proposal.) Consider the following questions:

1. How will pre-cruise planning be coordinated? (e.g. email, teleconference, workshop)
2. What types of sampling instruments will be deployed on the cruise?
3. How will the cruise event log be recorded? (e.g. the Rolling Deck to Repository (R2R) event logger application, an Excel spreadsheet, or paper logs)
4. Will you prepare a cruise report?

Cruise plans were made by others. A post-expedition report is available at http://publications.iodp.org/proceedings/349/349title.html.

Description of Data Types

Provide a description of the types of data to be produced during the project. Identify the types of data, samples, physical collections, software, derived models, curriculum materials, and other materials to be produced in the course of the project. Include a description of the location of collection, collection methods and instruments, expected dates or duration of collection. If you will be using existing datasets, state this and include how you will obtain them.

16S Gene Amplicon Sequencing

16S genes from 60 sediment samples were sequenced by Illumina MiSeq, generating .fastq files.

All corresponding sediment geochemical information was collected by IODP Expedition 349.

Data and Metadata Formats and Standards

Identify the formats and standards to be used for data and metadata formatting and content. Where existing

standards are absent or deemed inadequate, these formats and contents should be documented along with any proposed solutions or remedies. Consider the following questions:

1. Which file formats will be used to store your data?
2. What type of contextual details (metadata) will you document and how?
3. Are there specific data or metadata standards that you will be adhering to?
4. Will you be using or creating a data dictionary, code list, or glossary?
5. What types of quality control will be used? How will data quality be assessed and flagged?

Sequencing data files will be stored as .fastq files or as compressed .fastq.gz files. Sequencing files will be described in an Excel spreadsheet according to the core and sediment depth the sediment was collected from, along with additional metadata harvested from existing IODP records.

Data Storage and Access During the Project

Describe how project data will be stored, accessed, and shared among project participants during the course of the project. Consider the following:

1. How will data be shared among project participants during the data collection and analysis phases? (e.g. web page, shared network drive)
2. How/where will data be stored and backed-up?
3. If data volumes will be significant, what is the estimated total file size?

Data was shared via portable hard drive and shared network drive during analysis. Data is stored on multiple external hard drives and computers in different locations. The total file size of sequencing data is approximately 1 Gb.

Mechanisms and Policies for Access, Sharing, Re-Use, and Re-Distribution

Describe mechanisms for data access and sharing, and describe any related policies and provisions for re-use, re-distribution, and the production of derivatives. Include provisions for appropriate protections of privacy, confidentiality, security, intellectual property, or other rights or requirements. Consider the following:

1. When will data be made publicly available and how? Identify the data repositories you plan to use to make data available.
2. Are the data sensitive in nature (e.g. endangered species concerns, potential patentability)? If so, is public access inappropriate and how will access be provided? (e.g. formal consent agreements, restricted access)
3. Will any permission restrictions (such as an embargo period) need to be placed on the data? If so, what are the reasons and what is the duration of the embargo?
4. Who holds intellectual property rights to the data and how might this affect data access?
5. Who is likely to be interested in re-using the data? What are the foreseeable re-uses of the data?

Raw sequencing data was made publicly available via NCBI at accession number PRJNA362622 with no restrictions and is paired with metadata describing the sediments from which each sample originated. Data from this project may be of interest to sediment microbiologists and those studying the sediment biogeochemistry of the South China Sea.

Plans for Archiving

Describe the plans for long-term archiving of data, samples, and other research products, and for preservation of access to them. Consider the following:

1. What is your long-term strategy for maintaining, curating, and archiving the data?
2. What archive(s) have you identified as a place to deposit data and other research products?

Raw sequencing data was made publicly available via NCBI at accession number PRJNA362622 with no restrictions and is paired with metadata describing the sediments from which each sample originated. Data from this project may be of interest to sediment microbiologists and those studying the sediment biogeochemistry of the South China Sea.

Roles and Responsibilities

Describe the roles and responsibilities of all parties with respect to the management of the data. Consider the following:

1. If there are multiple investigators involved, what are the data management responsibilities of each person
2. Who will be the lead or primary person responsible for ultimately ensuring compliance with the Data Management Plan?

Frederick Colwell will be the primary person ultimately responsible for ensuring compliance with the data management plan, including collecting data from all researchers and ensuring it is publicly available with metadata.