

Data management plan:

CAREER: Assessing local adaptation in the chemosynthetic symbionts of hydrothermal vent animals

Roxanne Beinart (University of Rhode Island)

Data Policy Compliance: The PI will comply with the data management and dissemination policies described in the *NSF Award and Administration Guide* (AAG, Chapter VI.D.4) and the *NSF Division of Ocean Sciences Sample and Data Policy*.

Pre-Cruise Planning: Pre-cruise planning will occur through email, teleconferences, and virtual planning workshops with all involved collaborators and participants. The cruises will utilize a remotely operated vehicle (ROV JASON, Woods Hole Oceanographic Institution). Preliminary dive plans will be written prior to each cruise, and updated and amended prior to each dive based on cruise events.

Cruise event logging: Detailed dive reports for each ROV dive will be digitally compiled as Microsoft Word documents during the cruise. Each dive report will include site information, times of launch and recovery, sampling events, and watch-stander logs. These, along with each dive plan, participant information, and samples logs will be compiled into a cruise report (PDF document) at the conclusion of the cruise.

Description of Data Types

This project will produce observational datasets through the ship's underway sensors, ROV sensors and cameras, the *in situ* mass spectrometer, and derived from analysis of collected biological specimens. Observational data will be collected on two Lau back-arc basin research cruises.

Cruise underway data: Standard underway data collected along the ship's track (e.g., sea surface temperature, salinity, etc.). File types: .csv ASCII files. Repository: BCO-DMO and the Rolling Deck to Repository (R2R).

ROV data: Routine sensor data (e.g., temperature), video, and images collected by the ROV Jason, as well as dive event logs, are made available through the Jason Virtual Van and archived at the National Deep Submergence Facility (NDSF) at Woods Hole Oceanographic Institution. File types: .csv ASCII, .jpg, .mpeg Repository: Virtual Van and NDSF archive

in situ mass spectrometer data: This includes raw mass spectra, calibrated mass spectra, and final calibrated data. File types:.txt files. Repository: BCO-DMO.

Stable isotopic composition data: Carbon stable isotopic composition data will be derived from *Alviniconcha* tissue and experimental seawater samples. Analysis of stable isotopic composition will be performed at the Yale Analytical and Isotope Center. File types: Excel file(s). Repository: BCO-DMO.

Genomic and gene sequences: Genomic and gene sequences will be derived from *Alviniconcha* tissue. Genomic sequencing will be performed with Psomagen Inc., and individual gene sequencing will be performed at U. Rhode Island's Genomic Sequencing Center. The resulting assembled and raw data will be made publicly available. File types: Short-read archive .sra, .fastq read files, .fasta files. Repository: NCBI; accession numbers and associated sample metadata will be submitted to BCO-DMO.

Data and Metadata Formats and Standards: Genetic sequence data will be prepared in accordance with the minimum information about a metagenome sequence (MIMS) specifications and minimum information about a marker gene sequence (MIMARKS) developed by the Genomic Standards Consortium. All other data will be stored in Excel spreadsheets, which can be submitted to the BCO-DMO. In addition, metadata will be prepared using BCO-DMO metadata forms and will include detailed descriptions of sampling and analysis protocols.

Data Storage and Access During the Project:

The investigator will store project data (including Excel spreadsheets, PDFs of dive logs, etc.) on a shared Google Drive directory that will be hosted through the University of Rhode Island's unlimited space Google Suite. In addition, this data will be stored on the investigator laboratory computers that are backed up by their institution's central IT organizations. Genetic sequence data will be kept on laboratory network attached storage and on the Brown University large data storage with automatic off-site backups associated with its high-performance computing cluster, which URI researchers have access to through an EPSCoR agreement.

Mechanisms and Policies for Access, Sharing, Re-Use, and Re-Distribution: Immediately after completion of the research cruise, underway data and metadata will be submitted to the Rolling Deck to Repository. Within four months of the conclusion of the cruise, the final cruise report will be submitted to the BCO-DMO. All other data will be submitted to the appropriate repository (detailed previously) upon submission of manuscripts, or 2 years post-cruise, whichever comes first. Genbank accession numbers will be submitted to the BCO-DMO after sequences are submitted to Genbank.

Data produced by this project may be of interest to chemical and biological oceanographers, particularly deep-sea biologists studying hydrothermal vent microbial and invertebrate communities. We will adhere to and promote the standards, policies, and provisions for data and metadata submission, access, re-use, distribution, and ownership as prescribed by the BCO-DMO Terms of Use (<http://www.bco-dmo.org/terms-use>), the minimum information about a metagenome sequence (MIMS) specifications and Minimum information about a marker gene sequence (MIMARKS) developed by the Genomic Standards Consortium.

Plans for Archiving: Rolling Deck to Repository will ensure that the original underway measurements are archived permanently at NODC and/or NGDC as appropriate. NDSF maintains permanent archive of ROV sensor data, images, and video. The NCBI Genbank maintains a permanent archive of sequence data. The PI will work with R2R, NDSF, BCO-DMO to ensure data are archived appropriately and that proper and complete documentation are archived along with the data.

Roles and Responsibilities: The PI will be responsible for sharing the data among the project participants in a timely fashion. Collaborator P. Girguis will be responsible for collecting and analyzing the *in situ* and experimental geochemical data. The PI will oversee the stable isotopic analyses and genetic and genomic sequencing work and will submit the resulting sequences to the BCO-DMO and the National Center for Biotechnology Information's (NCBI) GenBank database. The PI will coordinate the overall data management and sharing process and will submit the project data, including GenBank accession numbers, and metadata to the Biological and Chemical Oceanography Data Management Office (BCO-DMO) who will be responsible for forwarding these data and metadata to the appropriate national archive.