EAGER: Tracking marine diazotrophy with isotope-labeling proteomics

Data Management Plan

Data Policy Compliance

The project investigators 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 (NSF 17-037).

Pre-Cruise Planning

Pre-cruise planning will be coordinated with the HOT program at the University of Hawaii using established procedures for participation of visiting researchers on HOT program cruises.

Description of Data Types

The primary type of data to be generated by the proposed research is proteomic mass spectrometry data from analyses of on-deck $^{15}N_2$ incubation experiments on the two proposed HOT cruises. This will include both raw mass spec data as well as interpreted data products such as peptide sequences, protein identifications, and relative protein quantification among and between samples. Additional data types generated by the project include 16S and nifH sequencing data and flow cytometry data. We also anticipate utilizing and refining in-house developed software tools, such as our peptide atom% 15N determination and functional/taxonomic assignment pipelines, that will be of use to the broader metaproteomics community.

Data and Metadata Formats and Standards

All raw proteomics mass spectral data will be deposited in the public MassIVE repository operated by the University of California San Diego (https://massive.ucsd.edu) concurrent with publication. MassIVE is a member of the international proteomeXchange consortium (http://www.proteomexchange.org) that facilitates proteomics data sharing. Proteomics data submissions will conform to the Minimum Information About a Proteomics Experiment – Mass Spectrometry (MIAPE-MS) guidelines maintained by the Proteomics Standards Initiative of the Human Proteome Organization (www.psidev.info/miape/MIAPE_MS_2.24.pdf). Data products generated by computational analysis of mass spec data, including tables of peptide sequences, annotations and atom% ^{15}N values, as well as sampling metadata, will be submitted to BCO-DMO (bco-dmo.org) in accordance with BCO-DMO conventions (i.e. using the BCO-DMO metadata forms) and will include detailed descriptions of collection and analysis procedures. Amplicon sequence data will be deposited in NCBI's Sequence Read Archive. Flow cytometry data will be deposited in FlowRepository (http://flowrepository.org). Software tools will be made available via the public code repository GitHub and the PI's website (biogeolabs.uchicago.edu/jwal).

Data Storage and Access During the Project

Data will be backed up and shared among project participants during data collection and analysis phases using UChicago Box, a cloud-based storage service that provides unlimited free online space for storing and sharing files. Data files will also be archived on the dedicated Proteome Destroyer proteomics analysis workstation in the Waldbauer lab.

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

Proteomics mass spectral data will be deposited in MassIVE and DNA sequences will be deposited in the National Center for Biotechnology Information (NCBI) database GenBank upon submission of manuscripts. MassIVE and GenBank accession numbers will be provided to the Biological and Chemical Oceanography Data Management Office (BCO-DMO) in an Excel spreadsheet or .CSV file and metadata will be provided using the BCO-DMO Dataset Metadata submission form. Data sets produced by the science party will be made available through the BCO-DMO data system within two-years from the date of collection. The project investigators will work with BCO-DMO data managers to make project data available online in compliance with the NSF OCE Sample and Data Policy. Data, samples, and other information collected under this project can be made publicly available without restriction once submitted to the public repositories.

Data produced by this project may be of interest to chemical and biological oceanographers, and climate scientists interested in the role of biogeochemistry in the global climate system. 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).

Plans for Archiving

The PI will work with the data repositories (MassIVE and GenBank) and BCO-DMO to ensure data are archived appropriately and that proper and complete documentation are archived along with the data. BCO-DMO will also ensure that project data are submitted to the appropriate national data archive.

Roles and Responsibilities

The PI, J. Waldbauer, will coordinate the data management and sharing process and will submit the project data, including MassIVE and GenBank accession numbers, and metadata to the Biological and Chemical Oceanography Data Management Office (BCO-DMO).