

Data Management Plan

Products of the Research and data formats. The primary data products for this project will include quantitative and rate-based DMSP, DMSO and DMS measurements, RT-qPCR gene abundance data for key genes involved with biogenic sulfur cycling, and both 16S rRNA (bacterial) and 18S rRNA (protistan) diversity data revealed by rRNA gene sequencing.

DMS(O)(P) data will reflect measurements of standing stock, synthesis and catabolism of these compounds under various experimental treatments of oceanic acidification off the Canary Islands and coastal Gulf Maine. RT-qPCR of sulfur-responsive genes (e.g., DMSP Lyases: *ddd(D)(L)(P)(Q)(W)(Y)* and those coding for DMSP demethylation pathways: *dmd(A)(B)(C)(D)*) will result in gene-specific abundances. Approximately 50,000 diversity-based sequencing reads will be generated for bacteria and 50,000 reads for protists to permit diversity estimation of the prokaryotic and eukaryotic microbial assemblages and provide a detailed taxonomic survey for each experiment. DNA sequences will be curated and quality checked using established bioinformatic pipelines to remove low-quality and/or chimeric sequences and will be maintained in the FASTA format.

Notebooks containing raw data generated by this project, detailed procedures and methodological approaches, deviations from protocols, specific equipment, and chemical reagents utilized for this project will be cataloged by the PIs, the Postdoc and both research technicians in their respective labs. Digital copies of laboratory notebooks, including related digital media (e.g., spreadsheets and data summaries), will be exchanged among the labs to promote data security and prevent physical loss via water or fire damage. Notebooks and digital copies thereof will serve as permanent records of the project and will be available upon request.

Teaching and public lecture material will result from the proposed research, primarily in the form of PowerPoint presentations and large-format research posters. Any such materials will be made freely available on the websites of each investigator for educational use. All PIs expect to incorporate results from the proposed research into their future undergraduate lesson plans as part of the new affiliation between Bigelow Laboratory and Colby College.

Access to Data and Data Sharing Practices and Policies. All aspects of our research will be well-documented in both physical (Lab Notebooks) and digital formats. The project will be highlighted on the websites of all three PIs, which will provide links to all protocols and digital data following the NSF guidelines for the timeline of both data and sample release specified in the '2004 NSF Division of Ocean Sciences Data and Sample Policy'. A running log will be kept by the PIs for all samples collected and analyses performed and will be published on the project websites – this will ensure maximum transparency of both progress and data/sample availability. Data resulting from the proposed work will be shared openly after publication or at the end of the project following the NSF Policies on Dissemination and Sharing of Research Results: [Award Administration Guide, Section VI.D.4](#). The PIs will work closely on NSF annual reports to provide updates to NSF program managers, and to the general public. The PIs will collaborate on all aspects of data management during field deployments and back at home. To the extent they are still available after analyses, archived and frozen (-80 °C) sub-samples of extracted nucleic acids and unextracted backup samples will be available to interested parties for their independent analysis. In order to facilitate interpretation of the two field experiments, we will submit our processed data to the PANGAEA database (<http://www.pangaea.de>) in accord with other datasets generated during the BIOACID experiment (also see plans below for preserving/archiving/accessing data from BCO-DMO). PANGAEA data will include the basic environmental and biological metadata (temperature, salinity, carbonate chemistry, major

nutrient concentrations, HPLC pigments including chlorophyll *a*) and microbial abundance (e.g., counts of bacteria, picoeukaryotes, heterotrophic and phototrophic protistan nanoplankton, and microplankton). The information system PANGAEA is operated as an Open Access library aimed at archiving, publishing and distributing geo-referenced data from earth system research. The system guarantees long-term availability of its content through a commitment of the operating institutions. Data submitted by BIOACID partners and participants in the mesocosm experiments is restricted to the project partners for an embargo period of 2 years. After the embargo period, both metadata and data become free and open access. Participants are thus encouraged to publish in peer review journals during the embargo period in order to benefit from first use of their data. All experimental and field data published in PANGAEA are assigned a Digital Object Identifier (DOI) to make them citable and retrievable by library catalogues, scientific journal publishers or Google Scholar, which guarantees the availability of the unique BIOACID data to the global scientific community beyond the life time of the project.

Plans for Preserving and Archiving Research Products. All DNA sequences will be submitted to public databases such as GenBank: <http://www.ncbi.nlm.nih.gov/genbank/> and CAMERA: <http://camera.calit2.net/>. Environmental metadata will be included for all sequence submissions. We will work with staff at the Biological and Chemical Oceanography Data Management Office (BCO-DMO) to insure that all of our data are archived at the NODC facility and that DNA sequence and gene abundance data are linked back from international databases to NODC data portals. Permission will be requested from the NSF Program Managers to deposit data associated with biogenic sulfur measurements in the international DMS database (<http://saga.pmel.noaa.gov/dms/>), as these parameters are not yet accepted by the National Oceanographic Data Center <http://www.nodc.noaa.gov/> to comply with the '2004 NSF Division of Ocean Sciences Data and Sample Policy' for biological data. Additionally, our data sets will be made available online from the BCO-DMO data system (<http://bco-dmo.org/data/>) where they will be managed. We will work with BCO-DMO personnel to ensure that all data components are linked among the different databases including those hosting: **1)** DNA sequences and data generated from sequence analysis (e.g., sequence alignments, rarefaction curves, diversity estimates), **2)** biogeochemical process and rate measurements, **3)** environmental metadata, **4)** and microbial abundance and imagery data. Images will be submitted to Microbe Library <http://www.microbelibrary.org/>, a site created and hosted by ASM to generate a database of microbial images for use in undergraduate education and will also be submitted to Micro*Scope <http://starcentral.mbl.edu/microscope/>. All data will be archived as soon as it becomes available on storage devices located in each of the PIs labs and on their institutional servers, which are backed up both locally and remotely to ensure data security and prevent loss. Samples in both extracted and unextracted forms will be archived at -80 °C and divided among at least two freezers until analysis; biogeochemical samples are destroyed during analysis. We are presently working to implement a lab-wide digital tracking program for logging all collected samples. Our new laboratory facilities in East Boothbay have secure, reliable and redundant data backup facilities and emergency power generators to ensure both data and sample security. Data resulting from the project will be preserved by the Bigelow Laboratory for a period of at least 10 years according to institutional policy.