

DATA MANAGEMENT PLAN

Collaborative Research: Investigating iron-binding ligands in Southern Ocean diatom communities: The role of diatom-bacteria associations

The proposed work will result in macronutrient (nitrate+nitrite, nitrite, phosphate, silicate), micronutrient (dissolved iron, ^{57}Fe ; leachable particulate Fe, ^{57}Fe), iron speciation (iron-binding ligand concentrations and conditional stability constants), chlorophyll *a*, photosynthetic efficiency (Fv/Fm), dissolved organic carbon (DOC), exopolymeric substances (EPS), diatom and bacteria community composition, diatom and bacterial transcriptome, and quantitative PCR data from a suite of field- and laboratory-based experimental bottle incubations of Southern Ocean diatom communities.

Experimental data will be submitted to the data archive managed by the **Biological & Chemical Oceanography Data Management Office (BCO-DMO)**; these data sets will be available online from the BCO-DMO data system (<http://bco-dmo.org/data/>) and archived permanently at the **National Oceanographic Data Center (NODC) database** (www.nodc.noaa.gov). Co-PI Buck has prior experience with this data system, which maintains the data from her U.S. GEOTRACES and CoFeMUG iron speciation work. PI Jenkins and Co-PI Chappell received training on BCO-DMO data management practices at the 2012 Ocean Carbon Biogeochemistry Workshop in Woods Hole, MA. Within the BCO-DMO system, our data will be linked to previous data generated by the co-PIs.

We will isolate diatom species from our field studies and experiments. Isolates that are established as axenical for lab experiments **will be deposited in the Provasoli-Guillard National Center for Marine Algae and Microbiota (NCMA—formerly CCMP)**.

We will be generating multiple diatom transcriptomes, as well as eukaryotic and bacterial metatranscriptomes. We will also be generating additional bacterial and diatom population data using high throughput sequencing methods. Sequence data will be submitted to the **Sequence Read Archive (SRA) at the National Center for Biotechnology Information** (<http://www.ncbi.nlm.nih.gov/sra>), which stores sequencing data from highthroughput sequencing platforms. 18S and 16S sequence data will also be submitted to the **European Molecular Biology Laboratory's European Bioinformatics Institute (EMBL EBI) metagenomics database** (<https://www.ebi.ac.uk/metagenomics>) that has a pipeline for both SRA submission and comparison to other samples. The EBI Metagenomics service is an automated pipeline for the analysis and archiving of metagenomic data and includes analysis for the phylogenetic diversity as well as the functional and metabolic potential of a sample. All of the public data in the repository can be freely browsed. The EMBL EBI metagenomics database will allow others to access our data and perform taxonomic analysis using Qiime. We note that this database already contains metagenomic data from Antarctic lakes, Arctic Winter marine ecosystems, and bacterioplankton associated with a spring phytoplankton bloom.

We will keep NSF abreast of our compliance with data management through our annual reports.