

DATA MANAGEMENT PLAN

I. Types of data

The proposed research project will generate data from the analysis of culture samples collected in the laboratory and field as well as field incubation samples collected during the proposed cruise.

Data from the cultures will include growth rates under different conditions described in the proposal text, nitrogen isotopic ($\delta^{15}\text{N}$) measurements of dissolved inorganic (i.e., nitrate+nitrite, nitrite), dissolved organic, and particulate nitrogen and particulate carbon, chlorophyll *a*, photophysiological data (e.g. Fv/Fm, Ek, alpha, non-photochemical quenching, absorption cross sectional area of PSII, electron transport rates), abundance of proteins involved in N assimilation and photosynthesis and quantitative reverse transcription polymerase chain reaction (qRT-PCR) data.

Data from the field will include water column concentration and nitrogen isotopic ($\delta^{15}\text{N}$ and $\delta^{18}\text{O}$) measurements of dissolved inorganic, dissolved organic, and particulate nitrogen (i.e., nitrate+nitrite, nitrite, ammonium, dissolved organic nitrogen, and suspended particulate organic nitrogen). Additionally, size-fractionated particulate nitrogen concentration and $\delta^{15}\text{N}$ samples from the water column will be collected and measured. Samples in the field will also be collected from manipulative incubation experiments conducted during the cruise, including: concentration and $\delta^{15}\text{N}$ measurements of particulate carbon, particulate nitrogen, nitrate+nitrite, nitrite, ammonium, and dissolved organic nitrogen, chlorophyll *a*, abundance of proteins involved in N assimilation and photophysiology. Amplicon sequencing data for diatom community composition and qRT-PCR gene expression data will be generated for water column and incubation samples.

II. Data and Metadata Standards

The water column nutrient data generated from the cruises will be integrated with supporting hydrographic (CTD and bottle) data to create a merged water column data set. All water column, shipboard incubation, laboratory data, and associated metadata will be submitted by project PIs to the Biological and Chemical Oceanography Data Management Office (BCO-DMO), to be made publicly available online from the BCO-DMO data system following standard NSF requirements upon publication. All PI's have prior experience with this data system, which maintains the data from Knapp's work quantifying nitrogen fixation rates in the Western and Eastern Tropical South Pacific, work Kranz contributed to in the Palmer Station LTER, and Chappell's measurements from diatom cultures.

The project will also generate nucleic acid sequence data (18S amplicon sequencing), which is not directly housed by BCO-DMO, though BCO-DMO will provide links to sequence data stored elsewhere. The sequence information will be captured by Illumina sequencing, to be conducted at Old Dominion University (amplicon sequencing). Nucleic acid sequence information will be stored in FASTA formatted files or SFF files, which integrate sequence quality information with the sequence itself. Nucleic acid sequence data and complementary metadata will be made available through multiple avenues. Primarily they will be submitted to the NCBI short read archive (SRA), which houses most high-throughput data obtained to date. The appropriate metadata to make the sequence information meaningful include physical and chemical conditions of incubations, biomass of microorganisms, and their relative productivity. These are consistent with genomics standards consortium (GSC) standards for metagenomic studies. For redundancy, sequence data will also be submitted to the European Molecular Biology Laboratory's European Bioinformatics Institute (EMBL EBI) metagenomics database that has a pipeline for both SRA submission and comparison to other samples. We will keep NSF apprised of our compliance with data management through our annual reports.

III. Policies for access and sharing and provisions for appropriate protection/privacy

Water column nutrient concentration and isotopic data as well as shipboard and laboratory incubation data will be deposited to BCO-DMO as described above. Nucleic acid sequence

DATA MANAGEMENT PLAN

data and complementary metadata will be submitted to the NCBI SRA and the EMBL EBI metagenomics database. Metadata is also submitted at the same time. Links to this public sequence data will be made available through BCO-DMO. Data will be released upon publication. These databases are accessible to the public, and all data will be published in peer-reviewed journals. The PIs have a history of reporting data in figures as well as in supplementary tables for ease of use by other researchers, and will continue to publish data in this format.

IV. Policies and provisions for re-use, re-distribution

There will be no permission restrictions for these data. The data may be of interest to chemical and biological oceanographers. The intended and foreseeable users of the data are oceanographers, phytoplankton ecologists, and modelers within academia and government labs.

V. Plans for archiving and preservation of access

Initially, all data will be archived on computers in the respective labs of the PIs, and backed up on remote servers and/or external hard drives. Data will be submitted to public databases (BCO-DMO, NCBI and EMBL) where they will be permanently archived to preserve access to the public. A hard copy of all notes (i.e., lab notebooks) will be retained in the laboratory. All relevant metadata associated with genomic libraries will be submitted along with the nucleic acid sequences themselves. Research publications generated from this work will include all relevant data and refer readers to public databases where data is permanently archived.