DATA MANAGEMENT PLAN

Data policy compliance

The project investigators will comply with the data management and dissemination policies described in the NSF Proposal & Award Policies & Procedures Guide (PAPPG - Chapter XI.D.4) and the NSF Division of Ocean Sciences Sample and Data Policy (NSF 17-037).

Description of data types

The project will produce several experimental datasets which are described below:

1. NirA Enzyme Kinetics: The results of in vitro biochemical assays will include rates of nitrite reduction at several substrate (nitrite) concentrations for both versions of the nitrite reductase (NirA) enzyme found in LLI *Prochlorococcus* and NirA in high-light adapted *Prochlorococcus* and a co-occurring heterotroph (e.g. *Marinobacter* sp.). File type: tab-delimited text file; Repository: BCO-DMO.

2. Nitrite production and consumption: The results of culture experiments using axenic strains will include cell concentrations and nitrite concentrations over the course of the growth curve for batch cultures and during steady state conditions for continuous cultures. Derived data will include net cell specific nitrite production or consumption rates (depending on strain) for each time point during the growth curve or during steady state conditions. File type: tab-delimited text file; Repository: BCO-DMO.

3. Fast repetition rate fluorometry (FRRF): FRRF profiles will be obtained for several of the physiology and field experiments (File type: tab-delimited text file; Repository: BCO-DMO and FigShare). Data derived from analysis of the FRRF profiles will include standard photosystem parameters including the quantum efficiency of photochemistry in photosystem II, initial fluorescence, maximum fluorescence, variable fluorescence, function absorption cross-section of photosytem II, and rates of electron transfer between PSII and PSI File type: tab-delimited text file; Repository: BCO-DMO and FigShare.

4. Co-culture strain frequencies: The results of continuous co-culture experiments will include qPCR data (strain specific cell concentration), flow cytometry data (total population cell concentration), and derived proportional frequencies of each strain in the co-culture during steady state and during acclimation to imposed perturbations. File type: tab-delimited text file; Repository: BCO-DMO.

5. Nucleic acid sequences: DNA (metagenomics) and RNA (transcriptomics and metatranscriptomics) sequencing will be performed at the MIT BioMicro Center using the Illumina platform. DNA samples were previously collected on NSF C-MORE funded cruises in the North Pacific Subtropical Gyre in the summer of 2012 (cruise identifiers KM1215, KM1217, and KM1219) and archived at -80ºC at MIT. RNA samples will be collected as part of physiology experiments to assess the transcriptional response of cells that do and do not excrete nitrite when acclimating to increased irradiance. Additional DNA and RNA samples will be collected and sequenced as part of our proposed field work in the North Pacific Gyre. File type: tab-delimited text file and fastq file; Repository: NCBI-SRA and NCBI-GEO.

Data and metadata formats and standards

For sequence data, we will comply with the Genomic Standards Consortium (GSC) recommendations for reporting with metadata stored in machine readable text files. For metadata submitted to the Biological and Chemical Oceanography Data Management Office (BCO-DMO), metadata will be prepared in accordance with BCO-DMO conventions and will include detailed descriptions of collection and analysis procedures. NIST certified nitrite and nitrate standards for colorimetric measurements will be obtained from a commercial company (e.g. Sigma) and detection limits and standard deviations will be reported with the accompanying nutrient analysis data. PCR efficiency and limits of detection will be reported with the strain specific cell concentration data for co-culture experiments.

Data storage and access during the project

The project data (e.g. scanned laboratory notebooks, spreadsheets, colorimetric nutrient acquisition data, FRRF profiles, raw flow cytometry and raw qPCR data) will be stored on laboratory computers which are backed up daily to a centralized institute server using CrashPlan PROe. Raw sequencing data (approximately 550 GB compressed) will be stored in two physical locations: on storage servers at the MIT BioMicro Center and on storage servers at the Massachusetts Green High Performance Computing Center (MGHPCC). The PI maintains weekly backup copies of his personal computer and stores these backups at both on-site and off-site locations.

Mechanisms and policies for access, sharing, reuse, and redistribution

The expected data products are not sensitive in nature and do not include human or animal subjects. Berube and Chisholm are both covered the MIT Faculty Open Access Policy which allows authors to legally make their final, peer-reviewed manuscripts freely accessible through the open access repository DSpace@MIT. All axenic *Prochlorococcus* strains and nucleic acids samples will be made available to other researchers upon request at no more than incremental cost and within a reasonable time. Nucleic acid sequence data will be published in open access journals (e.g. Nature Scientific Data or ASM Microbiology Resource Announcements) to facilitate access and use of these datasets. Any custom code derived from this project will distributed through a GitHub repository.

Plans for archiving

All sequencing data will be permanently archived in NCBI and further distributed by International Nucleotide Sequence Database Collaboration (INSDC) partners. The PI will work with BCO-DMO personnel to ensure that all data resulting from the project are appropriately archived and include proper documentation for methods and data processing pipelines.

Roles and responsibilities

The Lead PI, P.M. Berube, will coordinate all aspects of the data management plan including the supervision of project personnel to ensure appropriate data collection and archiving practices. Berube will be responsible for submission of sequencing data to the National Center for Biotechnology Information's (NCBI) sequence read archive (SRA) and gene expression omnibus (GEO) databases, as well as submission of final data products (co-culture strain frequencies and nitrite production/consumption rates for laboratory cultures) to BCO-DMO.