

DATA MANAGEMENT PLAN

Data Policy Compliance

The project investigators will comply with the NSF OCE Data and Sample Policy and the NSF Award and Administration Guide (Chapter VI.D.4).

Description of Data Types

This project will produce experimental datasets, evolved laboratory cultures, and computer code, described in the list below.

Experimental Datasets:

Laboratory experimental data: A number of different metrics will be used to quantify the physiology and ecology of evolved cell lines. This will generate a series of datasets reporting (1) growth rates, (2) photophysiology including chlorophyll, photosynthetic efficiency, and photosynthetic rates, (3) respirometry, (4) cellular stoichiometry, (5) grazing rates, and (6) palatability to higher trophic level consumers. These data and their metadata will be tabulated in .csv and .txt files as appropriate, and uploaded to BCO-DMO and/or Dryad data repositories.

Transcriptomes: Transcriptome sequences will be compiled into .fasta files; unique genetic sequences will be uploaded to the NCBI repository, with accession numbers provided to BCO-DMO. Because the raw datasets will be large, they will be submitted to the iMicrobe platform (where the MMETSP transcriptomes are also available). Metadata will be provided to BCO-DMO.

Living Organisms:

Evolved cell lines: The evolution experiment will generate 405 independent lineages of *Ochromonas* across all experimental treatments and replicates. These lineages will be archived twice per year by cryopreserving a subset of culture in the Moeller Lab -80°C freezer, and will be available to other researchers on request. During Year 3 of the proposed work, PI Moeller will contact the National Center for Marine Algae and Microbiota (NCMA) at Bigelow Labs to ask whether she can submit representative evolved lineages to their collection.

Computer Code and Model Output:

Statistical analysis code: Analyses of the above datasets will require statistical code (e.g., using the programs R or MATLAB). After manuscripts have been accepted for publication, code will be made available either as a supplement (depending on the journal) or by deposition on the first author's GitHub page.

COBALT ocean model output: The COBALT model simulates surface ocean biogeochemistry at the global scale and outputs a number of abiotic (e.g., nutrient, DOC) and biotic (e.g., phytoplankton, microzooplankton, larger grazers) variables. Detailed logs of model runs will be kept, raw model data will be stored on local data servers in PI Moeller's lab, and key output data will be stored as Netcdf (.cf) files. Metadata and specific outputs (e.g., DCM depth, C export) will be uploaded to the BCO-DMO repository as manuscripts are published. Additional output from model runs will be available to members of the community on request.

Data and Metadata Formats and Standards

For all datasets, we will strive to deposit data in non-proprietary, open data formats (e.g., .csv, .cf, .txt, .shp, ASCII, etc.) appropriate to the dataset in question (see above). Metadata preparation will follow BCO-DMO conventions (using BCO-DMO metadata forms), and will include detailed descriptions of station location and properties, data collection procedures, and data analysis pipeline.

Data Storage and Access During the Project

Investigators will store project data on laboratory computers at UC Santa Barbara. Any hard-copy data (e.g., lab notebooks and data sheets) will be photographed and/or scanned, and PDF copies will also be saved to these data repositories. PI Moeller has established a Google Drive by which her lab maintains cloud storage for data sharing. This cloud storage will be used for data sharing across project locations. PI

Moeller also requires that all project participants routinely back up computer hard drives daily using Apple Time Machine or another onsite external hard drive, and weekly to an offsite hard drive.

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

All datasets will be made available through the appropriate data system (see above) within two years of data collection or immediately following data publication, whichever comes soonest. Publication lead authors will work with BCO-DMO data managers to ensure that project data are in compliance with the NSF OCE Sample and Data Policy. Data may be of interest to biological oceanographers, microbial ecologists, and biogeochemists, and will be available without restriction once placed in public repositories. 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

By placing data in BCO-DMO, Dryad, NCBI, and iMicrobe repositories, PI Moeller will help to ensure that data are appropriately archived with appropriate and complete documentation, in addition to maintaining independent laboratory archives during her career.

Roles and Responsibilities

As the project lead, PI Moeller is responsible for sharing and managing collected data across project participants; this will be facilitated by cloud storage solutions. C. Laufkötter will collaborate with PI Moeller to ensure timely publication and open access of model output. In general, we expect initial data generation and curation from the empirical side of the project to be led by the graduate student supported by this work, and the modeling side to be led by the postdoctoral investigator. PI Moeller is ultimately responsible for ensuring that high-quality data products are published in a timely fashion.