NSF OCE-2049386 - Shunt or shuttle? Nutrient-driven biogeochemical consequences of diatom host-virus interactions

Data Management Plan

The PIs will conform to NSF policy on all aspects of data management, including the archiving, dissemination, and sharing of all research results. Data management will be coordinated by the PI, leveraging existing systems wherever possible.

Types of data and samples to be collected

We are proposing a laboratory and field based approach to understanding the role of nutrient availability on diatom host-virus interactions and subsequent biogeochemical consequences. Laboratory data will consist of a multitude of cellular and physiological parameters including: cell and virus abundance, photophysiology, particulate and dissolved nutrient analysis, chlorophyll, quantitative PCR, enzyme kinetics, sinking rates, flow cytometry, spectroscopy data, and FlowCam images. Biological samples will include fixed cells, extracted DNA, and filtered or concentrated cell biomass. Field data will consist of broad biogeochemical dataset such as chlorophyll, dissolved and particulate nutrients, FlowCam images, biomass proxies, etc.

Standards and formats for metadata and data

Raw and processed data will be maintained in the most appropriate and relevant format (i.e. laboratory notebooks, excel spreadsheets, standard flow cytometry files). When possible, data will be archived in cross-platform compatible tabular formats (e.g. .csv files). Field-based data will conform to the metadata standards of the Biological and Chemical Oceanography Data Management Office (BCO-DMO).

Data access and sharing policies

Data collected under the project will be made available to the public with as few restrictions as possible. Under these policies, all PIs plan for publication of most data with metadata, with manuscripts submitted during year 3 of the study and all major results published at a maximum of two years after the completion of the study. Once data are uploaded, PIs will confirm to BCO-DMO convention for making data publicly available within 1 year (or less).

Laboratory protocols will be posted, updated, and shared using the Protocols.io repository. A project page where all data (lab and field-based) will be submitted will be set up at the Biological and Chemical Oceanography Data Management Office (BCO-DMO).

To further facilitate data sharing as well as archiving requirements, experimental data and observations will be maintained in Thamatrakoln and Bidle's custom-designed, laboratory information management system (LIMS). Through a partnership with Big Rose Web Design LLC, Thamatrakoln and Bidle have built an integrated, query-based database management system that allows for data entry, secure data storage, visualization, and analysis. Big Rose Web Design LLC specializes in application and database development specifically for scientific data. The goal of the LIMS is integration of diverse datasets, obtained across multiple projects, in an easily-accessible format that allows for rigorous and robust interpretation and identification of potential linkages between experimental findings that may otherwise go unnoticed. A project within the LIMS specific to the work proposed here will be created and all relevant personnel will be granted authorized, secure access through our secure, password-protected lab website. Users will receive relevant training and will be able to add/edit/query/analyze/export data, which are stored in a MySQL database. MySQL data tables are created using standards based data schemas, ensuring that experimental data can be shared with other laboratories and exported to

public databases in uniform formats. This will greatly facilitate ease of data sharing between groups and organize our data in a standardized format. Throughout the duration of the project, Big Rose will provide support and maintenance of the database, as well as any new customizations that may be deemed necessary. The database is housed internally within DMCS on secure-servers maintained by the Information Technology group and backed up daily. As part of the LIMS, a public facing website will allow members of the community to access our curated data that will be made freely available.

Policies for re-use and re-distribution

All data publically available will be free to the community for re-use.

Plans for archiving and preserving samples and data

Biological samples will mostly be consumed during analysis (e.g. chlorophyll extracts, purified DNA, particulate and dissolved nutrients), but any remaining samples will be stored at -80°C for at least five years after the end of the project. As per NSF requirements, data will be archived at BCO-DMO.