Data management plan

Data description: The proposed experiments will generate a suite of physiological data (see project description). The fieldwork proposed during a BATS cruise will generate new DDA isolates for further experimental work in the lab. From these experiments we will harvest biomass for genome and transcriptome analysis. We will produce modeling codes in Python, which will be made freely available to the public upon publication at GitHub (https://github.com/ag105020), as has been done for most of Inomura's published studies. In short there are three major data types **1) experiment-derived physiological, genomic and transcriptomic data, 2) 2) biological cultured isolates and 3) modeling code.**

Data release: 1) If awarded, upon receipt of the award we will contact the Biological & Chemical Oceanography Data Management Office (BCO-DMO) to register our project on their website (http://www.bco-dmo.org/). All physiological and biogeochemical data will be uploaded to this site. 2) Genome and transcriptome sequences from experiments will be uploaded to NCBI's GenBank, Gene Expression Omnibus (GEO) and/or the Short Read Database (SRA) as appropriate and linked to both BCO-DMO and a single bio project number. Submission to GEO will include annotations and differential expression data from the transcriptome comparisons, as well as a link to the raw data in the SRA. 3**)** Novel cultured isolates will be a) submitted to anyone who places a request pending sufficient material and b) archived in the Provosoli-Guillard National Center for Marine Algae and Microbiota (NCMA—formerly CCMP).

Data archiving: **1)** Physiological and biogeochemical data will be assembled and organized in electronic spreadsheets and stored on local and backup servers, prior to submission to BCO-DMO. We will submit all data upon publication to BCO-DMO for archiving in a searchable project format. We will keep NSF abreast of our compliance with data management through our annual reports and all data will be made available as expeditiously as possible. **2)** Through the genome and transcriptome sequencing and subsequent analysis, we will be generating and storing significant amounts of sequence data and the associated analytical files. The data will be stored on redundant external hard drives, in the URI supported google drive space in the cloud (unlimited storage capacity) and one two local computers. In this manner there is redundancy in preserving the raw data and the associated analytical files. Data analysis will be performed on a custom pipeline run via PI Rynearson's access to both the Brown University compute cluster and the NSF supported XSEDE network access (see facilities statement). We have several strategies for data archival. Raw data will be included as supplementary material to publications when applicable, and will be available to the public upon publication through BCO-DMO as well. Further data will be archived with GEO and the SRA as appropriate with a link to BCO-DMO metadata. We will explore additional avenues for long-term archiving the data at other venues where there are appropriate metadata repositories as alternatives become available. **3)** New DDA isolates will be deposited in the Provosoli-Guillard National Center for Marine Algae and Microbiota (NCMA—formerly CCMP) whenever possible.