DATA MANAGEMENT PLAN

All data will conform to NSF Division of Ocean Sciences standards for data access and be made publically available through 1) the Biological and Chemical Oceanography Data Management Office (BCO-DMO), 2) the National Center for Biotechnology Information (NCBI; transcriptomic data), 3) Occidental College's Digital Commons, Institutional Repository, 4) the Dryad Digital Repository, and 5) direct requests to Stubler, DeBiasse, or Kelly. All datasets will be assigned unique digital object identifiers (DOIs) for citation. Upon funding approval of the proposal, PIs will immediately contact the BCO-DMO and register our project by submitting metadata as described by the NSF Division of Ocean Sciences Sample and Data Policy.

Expected Data Products

Environmental Data created in Laboratory:

Laboratory experiments will include discrete measurements of water quality data (pH, temperature, conductivity, dissolved oxygen, and total alkalinity) that will be collected at regular intervals (daily, weekly) throughout the experiments. These discrete data will be clearly logged in notebooks, transferred to digital spreadsheets, and saved as .csv files. Temperature and light (HOBO loggers) will be logged continuously during experimental periods. Continuous data will be downloaded weekly and saved as .csv files. Comparison samples will be taken from different treatments in the experiment and variables compared (sponge growth, carbonate dissolution, bioerosion rates). Any numerical data collected by hand from all analyses will be entered into an electronic database within a week of collection.

Transcriptomic Data:

Raw sequencing reads and their corresponding metadata will be submitted to the NCBI Sequence Read Archive. Assembled transcriptomes in fasta format, read count matrices in .csv format, and direct output from differential gene expression analyses in .csv format will be submitted to BCO-DMO and Dryad. This combination of data storage will ensure our analytical results can be accessed for reinterpretation, and the raw data can be accessed for re-analysis. These data will be deposited in their respective databases within two year of completion of the sequencing effort, allowing sufficient time for quality assurance. PI DeBiasse will be primarily responsible for submission of these data.

Metadata and Standard Operating Procedures:

Metadata will follow BCO-DMO guidelines, including the provision of variable names, units, experimental design details, analysis methods, descriptions of calibration procedures, etc. All ocean acidification data will strictly follow the "*Guide to Best Practices for Ocean Acidification Research and Data Reporting*" by EPOCA to allow for comparisons across different ecosystems and methods. Additionally, methodological and operating procedures will be written and made available for experimental design methods as part of our commitment to open-access science. These documents will be available through PI DeBiasse's GitHub, Dryad, and/or published in open-access formats where applicable.

Short Term Data Storage and Organization:

Short term data storage and organization will be primarily on lab databases/computers, dedicated storage hard drives, and Google Drive cloud storage; these methods will provide accessibility to all persons involved in the project. Throughout the project, transcriptomic sequence data will be available to all team members in cleaned, post-processing format; the raw data will be maintained in the lab where it was generated and will be submitted to NCBI and BCO-DMO.

Data Archiving and Availability:

At the conclusion of the project, all data will be converted and stored in an appropriate format to ensure that the structure of the data is readable by both humans and machines alike. All data and metadata will be contributed to the BCO-DMO and Dryad for archiving and public dissemination. Data and metadata will be replicated and stored at Occidental using external hard drives to ensure that no data set will be lost and can be easily transferred between collaborators and managers; data will also be archived in its entirety in the Digital Commons/Institutional Repository maintained by Occidental College. Proper documentation of the data and metadata will be ensured before archival. Metadata/documentation will include the project title, funding agency, objective, date collected and key words to help lead the user to the correct portion of the data needed, as well as relevant experimental design and activity logs. Raw data and its associated metadata, processed data (e.g., trimmed sequence reads), intermediate analysis files (e.g., gene expression count matrices), along with the command lines and scripts used to analyze the data will be made publically available on PI DeBiasse's GitHub and through Dryad. Manuscripts describing the project will be published in open-access peer reviewed journals.

Expected Schedule for Data Sharing

This project will conform to NSF Division of Ocean Sciences standards for data access and sharing. All project-related data products will be made publicly available free of charge. Continuously logged environmental data will also be made available immediately (within 48 hours) after they are archived. All other data files will be openly and freely available in the BCO-DMO, archives and open-access repositories within one year from the date of collection. Finalized data sets will be versioned to indicate changes since initial release, and a change notification service will be provided to users on request. Primary (raw) data will be archived along with the finalized data. Users will be required to agree to acknowledge this project and cite the dataset in any publications or derivative projects. In addition, synthesized and summary data will be provided via publications in peer-reviewed journals, with data DOI linked to the publications. In all cases we will work with the BCO-DMO staff to ensure that the data are as complete and transparent as possible.