Data Management Plan

Data Policy Compliance: The project investigators will comply with the data management and dissemination policies described in the NSF-OCE guidelines for samples and data (document num. NSF 17-037) Description of Data Types:

1. Chemical and environmental data: The proposed research will generate time-series data from multiple sensors in the mesocosms and from periodic bottle sampling including; dissolved O2, pH, salinity, temperature, dissolved inorganic carbon, and total alkalinity. Data (chemical parameters and profiles) will be organized and grouped across experiments, and additional treatments within each mesocosm. All of these data will have associated metadata and, after preliminary processing, will also have associated statistics and quality assurance metrics. Chemical conditions of manipulations a key data product to disseminate to the scientific community. The PI and the sponsoring scientist will discuss their data management needs for the chemical data generated in this proposal with the Biological and Chemical Oceanography Data Management Office (BCO-DMO), as this is the most logical data repository for this type of oceanographic time-series data. Data will be quality checked and then be transferred to the BCO-DMO office for archiving and sharing during each year of the project.

2. High-throughput sequence data: Several data types will be collected, including RNA-seq and methyl-seq. Raw sequence files and associated metadata will be uploaded to the NCBI short-read archive (SRA) as fastqfiles. Data products derived from raw sequence data will be similarly uploaded to appropriate NCBI database, including the Transcript Shotgun Archive (TSA, fasta file format) for assembled transcriptomes, or the Gene Expression Omnibus (GEO) for methyl-seq patterns. BioProject codes will be included in associated publications, and the links will be added to BCO-DMO.

3. Physiological assays: Other data types may include qPCR expression patterns, gene activity assays, and other physiological data including survival, growth, oxygen consumption and critical temperature and oxygen assays. Physiological data will be analyzed and quality controlled as described in the research plan. a. File types: simple text formats such as csv or txt files. Repository: Dryad, FigShare, or similar. These will be stored in simple text formats such as csv or txt files and archived in stable public repositories, such as Dryad or FigShare within two years of collection.

4. Curricular materials: Any materials developed for K-12 education programs will be saved as word documents, pdf files, and web-based materials. They will be archived on the Woods Hole Open Access Server(WHOAS).

Metadata Formats and Standards: Chemical metadata will be prepared in accordance with BCO-DMO conventions and will include detailed descriptions of collection and analysis procedures. Sequence metadata will comply with NCBI conventions including MINSEQE (Minimum Information About a Next-generation Sequencing Experiment) standards.