DATA MANAGEMENT PLAN

Data Description

The proposed project will generate many types of data and resources. Data types include sequence data (Illumina sequencing), environmental data (temperature), physiological measurements, a collection of genotyped natural populations to be curated in the Reitzel lab, and a taxonomic survey of potential prey species (invertebrates). Raw data from high throughput sequencing will create two types of file, one for base pair information and the second for sequence quality. These data will be made freely available through NCBI Sequence Read Archive (SRA) and on the PI's website. Data from sequencing will be primarily analyzed with CLC Genomics software and publicly available software (edgeR). Data analysis and interpretation will be disseminated through peer-reviewed publications and processed alignments, polymorphisms, and mapped loci will be made available by the Biological and Chemical Oceanography Data Management Office (BCO-DMO). Toxin gene data from each individual clone line and in each experimental condition will be deposited and made available by BCO-DMO. All processed data (identification and quantification of genes in each sample; lists of differentially expressed toxins and associated statistics) will be included in the database submission as well as published manuscripts. Temperature data collected by deployed loggers and species identified in field samples will be made freely available through BCO-DMO.

Data and Metadata Standards

Raw data from high throughput sequencing will be transferred from the sequencing core lab at UNC Charlotte to the PI's lab for analysis and storage. All data will be stored on computers in the lab and office of the PI as well as back-up servers at UNC Charlotte. Sequence data will be annotated with essential information for each sample, which will include description of the sample (e.g., location of origin for population, date, clone identification number), method for library generation, and the sequencing method. These data tags for description will be included with data submitted to SRA and BCO-DMO as well as hosted on the PI's website. Live animal cultures will be labeled with precise site of origin (latitude, longitude), date of collection, and genotype. Field temperature data will be stored in flat ASCII files, which can be read easily by different software packages. Field data will include date, time, latitude, longitude, and temperature, as appropriate. Metadata will be prepared in accordance with BCO-DMO using the BCO-DMO metadata forms and will include detailed descriptions of collection and analysis procedures.

Data Storage and Access During the Project

All raw and minimally processed sequence data (sequence and quality files) and temperature data will be backed up on hard drives in the laboratory and university servers. These data will also be uploaded to SRA, and BCO-DMO. All other digital data (e.g., qPCR, images) will be similarly stored on hard drives and the university network through shared drives on the network. Laboratory notebooks are stored in the lab. Use of electronic notebooks will be encouraged for those interested.

Data Access, Sharing, Re-Use, and Re-Distribution

All data will be made available through NCBI GenBank, NCBI SRA, BCO-DMO, and the PI's website as appropriate. Sequence-based datasets will be linked with the environmental data through BCO-DMO. GenBank accession numbers will be provided to BCO-DMO in an Excel spreadsheet and metadata will be provided using the BCO-DMO Dataset Metadata submission form. Sequence data will also be included with previous sequence data for *Nematostella* as part of a community effort for continued improvement to sequence resources through publicly available sites (Stellabase, JGI, future plans for UCSC browser). Data will also be shared via publication and presentations. Papers will be published in open access journals when possible using funds designated in the proposed budget. The PI will continue sharing living animals with other biologists for developing collaborative projects or to facilitate their independent projects. The PI has a long history of sharing animals for both research and teaching applications. Data sets produced by the science party will be made available through the BCO-DMO data system within two-years from the date of collection. The project investigators will work with BCO-DMO data managers to make project data available online in compliance with the NSF OCE Sample and Data Policy. Data, samples, and other information collected under this project can be made publically available without restriction once submitted to the public repositories.

Plans for Archiving

Sequence data will be archived at NCBI GenBank or SRA. BCO-DMO ensures that the data are archived properly at the appropriate National Data Archive for long-term archive preservation. The PI will work with BCO-DMO to ensure data are archived appropriately and that proper and complete documentation are archived along with the data.