Data Management Plan for "Collaborative Research: The effects of marine heatwaves on reproduction, larval transport and recruitment in sea urchin metapopulations"

The project investigators will comply with the data management and dissemination policies described in the NSF Award and Administration Guide (AAG, Chapter VI.D.4) and the NSF Division of Ocean Sciences Sample and Data Policy. Data will include those collected from field and experimental samples, computer model output, computer code, as well as molecular sequence data. Non-molecular data include physiological and vital metrics (animal respiration rates, body composition, mass, consumption rates), ecological data (wild source location, ecological source habitat data, collection dates), experimental biophysical data (temperature, salinity, oxygen concentration, flow rate of experimental units) as well as additional body and food composition information for experimental animals (fatty acid profiles including concentration and proportional composition, total lipid content, wet mass, dry mass, ash free-dry mass, and body volume), ocean transport model solutions, and statistical model code. All non-molecular datasets will be stored as ACII files. Sequence data will include RNA-seq, DNA, and chromatin data from experimental animals. Sequencing will be performed at Florida State University following the experiments, will be stored as FASTQ files and uploaded to NCBI and Florida State University public repositories; accession numbers to be provided to BCO-DMO.

Data and Metadata Formats and Standards

Non-molecular data: We will publish datasets using the XML specification adhering to ecological metadata language (EML) specifications, which is adaptable for a variety of data types, and is in wide use among NSF funded ecological and environmental projects, including the LTER Network. We will make use of guidelines for metadata and data developed by and in compliance with compliance with BCO-DMO standards, which are designed to promote long-term usability. These include high-precision geo-referencing, the use of standard keywords, units and measurement descriptions, and the inclusion of methods and/or protocols with all datasets. Because individual organisms from experiments will have numerous different kinds of data associated with them over time, data will include relational unique identifiers for cross referencing experimental units with detailed metadata describing those relationships in related metadata files. Oceanographic model output, statistical and analytical modeling code, and related files will also be submitted to BCO-DMO with appropriate metadata.

Molecular data: All published sequence data will be uploaded to the Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/info/seq.html) with accession numbers provided to BCO-DMO along with metadata for cross referencing samples with biological and physiological data associated with those samples suppled to BCO-DMO. We will upload data to the Gene Expression Omnibus as both raw counts of sequencing reads for the features of interest, and/or normalized abundance measurements. Metadata will include the following: "Samples" (a clear and thorough description of the individual samples); "Protocols" (including: harvest and extraction protocol, library construction protocol, and library strategy); and, "Data Processing Pipelines" (including: include base-calling, alignment, filtering, peak-calling, generation of normalized abundance measurements, as well as the software used in the pipeline).

Data Storage and Access During the Project

The investigators will store project data (including spreadsheets, ASCII files, images, computer code and model output, and genetics data) on FSU, UC Davis, and UCSB servers that are backed up by each University's or Department's IT organization. The Principal Investigator (PI) has also has a central account with Florida State University cloud servers as a project central data repository. Personal computers in all laboratories are backed up daily to these servers.

Mechanisms and Policies for Access, Sharing, Re-Use, and Re-Distribution

Immediately upon completion of the project, all ecological, physiological, oceanographic, and computer model data as well as computer code will be made publicly available through the BCO-DMO data repository. Computer code will also be publicly available on GitHub to allow future versions to be tracked.

Immediately upon completion of the project, all published gene expression and genetics data will be uploaded to the Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/info/seq.html) with accession numbers provided to BCO-DMO along with metadata for cross referencing samples with biological and physiological data associated with those samples suppled to BCO-DMO.

Upon successful completion of the proposed work nucleosome sensitivity and dynamics assays will supply information on chromatin structure and genome organization at an unprecedented level. The chromatin structural annotation of these model organisms will aid researchers in the design and interpretation phases of their own genetic or genomic research projects. To facilitate use by the broader scientific community, we are also working closely with the FSU Research Computing Center (RCC) center and the newly formed FSU Center for Genomics and Personalized Medicine to make the chromatin structural data sets immediately and freely available to the research community by hosting all of the data on a publicly accessible web server. These will include all datasets (nucleosome sensitivity and nucleosome dynamics) as well as our analyses and annotations. We have modeled our server on tree highly successful genomic data websites: replicationdomain.org, genomaize.org and the UCSC genome browser housed at in the Department of Biological Science at FSU. The hardware available to support data acquisition, sharing, and management will be closely aligned and integrated with other key FSU campus computing, network, data acquisition instruments, and data storage assets. In this way we will leverage the existing campus cyberinfrastructure and capitalize on tried and tested procedures for pipelining data acquisition and analyses. Load-balanced file serves located in the Dirac Science Library (DSL) data center currently export FSU's shared-HPC data storage systems across the campus network backbone to the NovaSeq600 instrument located in the FSU College of Medicine. Data will physically reside on a set of robust raided storage arrays connected via QDR Infiniband to web and database servers. To ensure excellent network connectivity to campus computing and to users located around the globe, the DSL building is connected to the 10 Gbps campus backbone via redundant uplinks and the campus network backbone connects to the Florida Lambda Rail (a high speed network covering the State of Florida with connections to Internet2 and the National Lambda Rail). To ensure high availability of the servers used by this project, the DSL data center is equipped with two 40- ton HVAC cooling units, a large format battery UPS, and a 550 KW diesel powered backup generator in the event of a campus-wide power failure. Regular offsite backups will be performed in order to minimize data loose in the event of a catastrophic system failure.

For all datasets, we will adhere to and promote the standards, policies, and provisions for data and metadata submission, access, re-use, distribution, and ownership as prescribed by the BCO-DMO Terms of Use (http://www.bco-dmo.org/termsuse).

Plans for Archiving

BCO-DMO will also ensure that project data are submitted to the appropriate national data archive. Archived samples will be stored at PI Okamoto's lab at Florida State University or at co-PI Rogers-Bennett's lab at UC Davis in a -80C freezer on generator and emergency backup power.

Roles and Responsibilities

Each PI will be responsible for sharing his/her subset of data among the project participants in a timely fashion. The lead PI, D. Okamoto, will coordinate overall data management, sharing, and archiving process. J. Dennis will coordinate submitting molecular data to appropriate repositories and providing accession numbers and metadata to BCO-DMO. BCO-DMO will be responsible for forwarding these data and metadata to the appropriate national archive.