

## **Data Management Plan:**

**We have read and agree to abide by the Division of Ocean Sciences Data and Sample Policy**

### **Products of Research**

At all stages, data inventories and lab notebooks will be carefully generated and stored, and protocols for the lab, field, and analysis work will be generated and shared among all partners of the project using email, FTP servers, dropbox, and teleconferencing.

### **2. Data Storage and Preservation**

Genetic Data: DNA extractions from all samples are archived at -20C in 96-well-plate format, and when available, additional tissue is in long term storage at -80C. Metadata regarding each sample is currently stored in a custom Filemaker database available on the Baums Lab internal server and backed up regularly by the Penn State University ITS backup service. For the samples used in this study, metadata will be associated with genetic sequence data archived on publicly accessible NCBI servers and presented as supplemental information in resulting publications.

The sequences themselves will be stored in raw read format along with relevant barcoding and adapter sequence information on the Baums lab servers. Counts of differentially expressed genes will be submitted to the GEO database (<http://www.ncbi.nlm.nih.gov/geo/>). Gene networks of co-expressed genes will be submitted to DRYAD and/or the PSU ScholarSphere.

Survey data: Survey data recorded in the field is entered into an access database. The datasheets are photographed and hard copies are retained. Temperature loggers are downloaded onto a desktop computer and transferred to an access database. Supporting field images are downloaded to a desktop computer. The database files and images are backed up to a NOAA server onsite in Miami which is further backed up to an off-site NOAA server.

### **3. Data Formats and Metadata**

Genetic Data: Each DNA sample in the Baums Lab database has a unique identifier which is recorded along with relevant metadata regarding collection time, date, depth, GPS coordinates, etc. This data is contained in a Filemaker database for ease of access to lab members. Data tables can be exported in other formats such as tab delimited text files as needed. The raw data for microsatellite genotyping are chromatographs stored in an Oracle relational database in the Baums lab. Scored data is exported as cvs files and imported into the Filemaker database (see above). DNA sequence data will be stored as both raw and quality/adaptor trimmed reads in Fastq format files. Metadata concerning phenotype data will be recorded following the recommendations of Michner et al. (1997) for ecological datasets in the ecological metadata language (EML). All gene expression data will be stored in accordance with expectations for the GEO database formats required by NCBI.

Survey data: Survey and temperature data stored in access can be exported to other formats including text file that are readable in a variety of applications. The raw unmodified temperature logger files are retained in their native (HOBOWare) format as well as converted to a CSV text file and archived prior to transfer to the database. The image files are retained unmodified in their original JPG format.

### **4. Data Dissemination & Policies for Data Sharing and Public Access**

Genetic Data: Because of the limited nature of the DNA extractions, access to the samples used in this study will be by request to PI Baums. All data will be made available to the community

through peer-reviewed publication or 18 months after completion of the study through free-of charge online databases (such as DRYAD data archive), whichever comes first. All data files will be made publicly available via PSU Scholarsphere website, DRYAD (<http://datadryad.org/>). DRYAD and ScholarSphere require users to agree to a Creative Commons license. ScholarSphere's default Creative Commons license is Attribution-Non-Commercial-No-Derivs 3.0, or CC-BY-NC-ND. With this license the data provider (the PI's) share work with others and allow them to download it, provided they attribute the providers as the creators; they must also refrain from changing the content in any way and from using it for commercial means. DRYAD uses the CCO 1.0 Universal version. The providers (the PI's) dedicate the work to the public domain by waiving all of their rights to the work worldwide under copyright law, including all related and neighboring rights, to the extent allowed by law. Users can copy, modify, distribute and perform the work, even for commercial purposes, all without asking permission. Survey data: Survey data is available to the public by request and ultimately available through peer-reviewed publications. Accompanying data will be submitted to appropriate data repositories such as DRYAD (<http://datadryad.org/>) and deposited with the Biological and Chemical Oceanography Data Management Office (BMO-DMO). Reports are submitted to NOAA Office of National Marine Sanctuaries as required under the conditions of the sampling permits.

## **5. Roles and Responsibilities**

The data management plan will be implemented by the PIs. PI Baums will be responsible for submitting genetic data and associated metadata to public databases and maintaining the physical collections. PI Williams will be responsible for compiling past colony data that will be used to inform sampling design, and for cataloging samples and documenting colony condition as the samples are collected.