

## DATA MANAGEMENT PLAN

### DATA POLICY COMPLIANCE

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.

### DESCRIPTION OF DATA TYPES

This project will produce observational, experimental, and molecular datasets as detailed below. In addition to these datasets, educational outreach resources will be made available for public use on the University of Massachusetts Amherst digital repository (Scholarworks).

#### Observational Datasets:

1. **Water quality:** temperature data will be collected using Hobo Tidbit temperature loggers deployed at each field site. Data will also include standard water quality measurements (salinity, dissolved oxygen, pH) as well as metadata (deployment date/time, location latitude and longitude, depth).
2. **Abundance and size frequency:** at each site detailed in the project description, we will collect *Urosalpinx* abundance and body size data by hand with vernier calipers.

#### Experimental Datasets:

1. **Development time/growth/consumption/fecundity:** developmental data will be collected by hand, quantified as days between lay date of egg capsule and date of first emergence. Growth rate will be collected with digital imaging and by hand using vernier calipers and a microbalance. Data will include shell height, shell width, and body mass. Consumption will be measured as number of prey consumed per unit time. Fecundity will be recorded as egg capsules laid per mother (count). Hand collected data will be entered on waterproof datasheets and stored in a fireproof cabinet. Digital files will be stored on local computers and UMass Amherst cloud services (Box).
2. **Respiration:** metabolic rates will be inferred by measuring oxygen consumption. Data will include oxygen values in respirometry chambers in addition to metadata (snail body size, water temperature, population). Data will be created digitally using software.
3. **Thermal tolerance:** thermal tolerance data will be measured using lethal tolerance 50% methodology and recorded as binomial survival (yes/no) after thermal trials using an experimental heat bar. Metadata will include the body size, temperature treatment, and population. Data will be entered by hand on waterproof paper and digitized.
4. **Genomic/Transcriptomic sequencing:** All raw sequences will be stored on the Massachusetts Green High Performance Computing Cluster during data analysis, permanently archived on the University of Massachusetts Amherst Network Associated Storage (NAS) maintained by Co-PI Komoroske, and submitted to the NCBI Sequence Read Archive with required metadata prior to submission for publication of results. Resulting genotype data files will be curated in a similar manner as appropriate to the file types. All analyses will be conducted with R, bash and associated genomics data processing programs (e.g., FreeBayes, Samtools), and all associated code will be

additionally stored with the data on the Komoroske Lab NAS and hosted in a publicly accessible Komoroske lab Github repository.

### **Existing datasets:**

1. **Thermal performance and tolerance:** We will complement newly collected empirical data described above with previously collected data from our own experiments. PI Cheng is in possession of hard copy and digital versions of this data on local computers and backed up on cloud servers. Data associated with publications is archived on Dryad.
2. **Environmental data:** We will use complementary physical environmental data from multiple sources, including NOAA National Estuarine Research Program System Wide Monitoring Program and regional Ocean Observing Systems (e.g. NERACOOS).

### **DATA AND METADATA FORMATS AND STANDARDS**

Field observation data will be stored as comma-separated files (.csv). Experimental data will be stored as comma-separated files (.csv) and digital images (.tiff). Hard copies of field and experimental data will be digitally scanned into portable document format (.pdf). Genomic data will be stored as text (.txt), variant call format (.vcf), and binary alignment map (.bam) files as appropriate. Metadata will be stored using Morpho, using the Ecological Metadata Language (EML) standard and will include detailed descriptions of collection and analysis procedures.

### **DATA STORAGE AND ACCESS DURING THE PROJECT**

The investigators will store project data on laboratory computers that are backed up with cloud services (e.g. Carbonite or a local NAS for genomics data). Data will be shared among project participants during collection and analysis with cloud services provided by UMass (Box).

### **DATA SHARING POLICY**

Data and metadata described above will be made available to the public via UMass Scholarworks or NCBI at the time of journal publication or within two years of the completion of the project. This will enable the PIs, and graduate students sufficient time to analyze, interpret, and publish results before data are made public. Prior to being made public, data will be the intellectual property of the PIs, project collaborators and their home institutions. GenBank accession numbers will be provided to the Biological and Chemical Oceanography Data Management Office (BCO-DMO) in an Excel spreadsheet or .csv file and metadata will be provided using the BCO-DMO Dataset Metadata submission form. The investigators will work with BCO-DMO data managers to make data available in compliance with the NSF OCE Sample and Data Policy.

### **PLANS FOR ARCHIVING**

The investigators will work with BCO-DMO to ensure that project data are submitted to the appropriate national data archive, and that the proper and complete documentation are archived along with the data. As appropriate, digital data will also be permanently archived using UMass Amherst Scholarworks and Co-PI Komoroske's local NAS (for molecular data). Hard data copies will be stored in a fireproof cabinet at UMass main campus under the supervision of PI Cheng.

### **ROLES AND RESPONSIBILITIES**

PI Cheng will be responsible for ultimate compliance with the DMP and oversee all data and metadata associated with field observations and experimental data. Co-PI Komoroske will be responsible for all the genomic data and metadata.