DATA MANAGEMENT PLAN

Data Collection, Processing, and Analysis

The data generated by this project will be of several types:

1) Datasets consisting of behavioral responses of juvenile rockfish fish to CO_2 enrichment and low dissolved oxygen levels at the NOAA Santa Cruz lab and Humboldt State. Data include olfactory responses to predator and non-predator odor cues, behavioral lateralization results (relative and absolute lateralization indices), and escape time responses. The data will be recorded on datasheets and transcribed into Microsoft Excel spreadsheets, backed-up on hard drives, and uploaded to dataservers as .csv files.
2) Information on water chemistry parameters from experimental treatments in our aquarium facilities. These will include bi-monthly water samples of carbonate chemistry, daily measures of temperature, pH, dissolved oxygen, and salinity from replicate tanks housing organisms, and continuous records of pH and dissolved oxygen from large reservoir tanks with pH and O_2 sensors. The data from water samples and handheld meters will be recorded on datasheets and transcribed into Microsoft Excel spreadsheets, backed-up on hard drives, and uploaded to dataservers as .csv files. The continuous sensor time series will be processed in Matlab and stored as a .csv or .txt file.
3) Datasets consisting of swimming physiology assays in response to experimental conditions, including hypoxia tolerance, critical swimming speeds, and oxygen consumption rates at rest and during sustained swimming. The data will be recorded on datasheets and transcribed into Microsoft Excel spreadsheets, backed-up on hard drives, and uploaded to dataservers as .csv files.
4) Gene expression data including sample data, de novo assemblies, and BLAST annotations will be created. *De novo* transcriptome assemblies will be uploaded and publicly available in FASTA format on NCBI Sequence Read Archive (SRA), a central repository for next generation sequencing data. Gene expression data will be uploaded to the MIAME-compliant Gene Expression Omnibus through NCBI (publicly available).
5) Oceanographic sensors (SeapHOx units) will record environmental parameters at 4 locations. Data time series of temperature, pH, dissolved oxygen, and salinity will be downloaded from the sensors every 2-3 months and saved in .csv format. Data will be post-processed for quality control (noting breaks and extreme values) using Matlab and uploaded to dataservers as .csv files.
6) Enzyme activities and qPCR data will be available in Microsoft Excel spreadsheets, backed-up on hard drives, and uploaded to dataservers as .csv files.

The IT departments at our collaborating institutions will maintain and back up data servers. General analysis of all data will be conducted using the software packages Matlab, R, JMP, SPSS, Systat, CLC Genomics Workbench, and ErmineJ. PI Logan will access the NSF XSEDE supercomputer clusters for transcriptome assembly and data storage.

Documentation

Metadata will be documented at the time of collection and analysis for each data component described above. Metadata will consist of information on the origin, timing, location, and observer at the time of original data collection; metadata will be updated to include modifications, QA/QC, and transformations and the researcher responsible for these changes.

Products

The data products made available to the public will vary depending on the data type:

1) Behavioral response data will be made available as raw data.
2) Water chemistry data from experimental treatments will be made available as raw data or in summarized forms.
3) Swimming physiology data will be made available as raw data.
4) Gene expression data will be made available in the formats described above.
5) Oceanographic sensor data will be made available in processed form after calibration with carbonate chemistry measures from water samples.
6) Otolith early life history data will be made available as raw data.
7) Enzyme activities and qPCR data will be available as raw data.

In addition, all experimental and field data, environmental data, de novo transcriptome assemblies, and BLAST annotations will also be publicly archived in the Dryad Database Repository (http://datadryad.org/pages/depositing) upon publication.

Data Access Policy

We anticipate the primary users of our data will be academic researchers interested in fish ecology, fisheries, physiological responses to climate change, and ocean chemistry. Data and metadata described above will be made available to the public at the time of journal publication or within 12 months of data collection. This will enable the PIs and graduate students sufficient time to analyze, interpret, and publish results before data are made public. *De novo* transcriptomes will be made available to the public upon completion to aid other interested researchers. Prior to being made public, data will be the intellectual property of the PIs and their home institutions. Data may be shared with collaborators from other institutions and portions of the data may be shared with interested researchers upon request. When sharing data, the PIs will request and encourage the interested users to collaborate with the original data collectors (students, or PIs) on any new projects or publications that use those data.

Outside of formal data archiving, students and researchers at collaborating institutions will periodically report on preliminary findings in blogs, lab Facebook pages, and other informal outreach forums. Data summaries, photos, and videos released in this manner will remain the intellectual property of the PIs and video/photo content will fall under the copyright of the appropriate institution.

Data Curation and Publication

Metadata and raw data will be made available using the Biological and Chemical Oceanography Data Management Office (BCO-DMO). We will register with BCO-DMO when our award begins and submit data/metadata to them on a regular basis, following the schedule outlined below above in our Data Access Plan.