Data Management Plan

1. Expected Data Types, Formats, and Security

***a) Physical Samples**:* Approximately 1100 samples of the focal species, *Lottia gigantea*, will be collected from across the geographic range. Samples will be stored in ethanol and kept on ice during transport. Once at UC Davis, they will be preserved at -80°C.
***b) Sequence Data**:* We will obtain Illumina short-read data from population-level samples. Genomic sequence data from 12 locations (60-120 individuals each) across the *L. gigantea* will be acquired from an Illumina HiSeq platform at the DNA Technologies Core at UC Davis. Raw data will be obtained in the fastq file format.
c) Trait data: Common garden experiments will produce trait data for individuals from 8 locations across the species range.
d) Field data: We will collect data on growth rates of owl limpets from 8 locations across the species range. For each individual, we will also collection environmental data including temperature, aspect, tidal height, and surrounding biota.
***d) Analytical pipelines*:** We will be adapting existing bioinformatics pipelines for the analysis of genomic data. This includes mapping of resequencing data, identifying genetic variants, analysis of population structure, and genotype-environment associations.

2. Data standards

All students, technicians, assistants, and postdocs will be trained in relevant methods by the project PIs (Bay, Sanford, Grosberg). Datasets collected will include a consistent set of metadata to be agreed upon and finalized at the first PI meeting in year 1. Minimally each dataset will include a title, description, location of source population (latitude & longitude via GPS), date, time (UTC), PI name and contact information and co-PIs. Each data set will also include, as appropriate, standards used for measurements, instrumentation used, analytical methods used, data processing information, sampling procedures, and access restrictions. Data will be entered into a standardized digital format to be agreed upon by the PIs and in consultation with BCO-DMO staff (most likely Excel spreadsheets and then converted into comma delimited files for submission to BCO-DMO for final archiving, as we have done with similar data in the past—see Project Description: Results from Previous NSF Support). A standard file-naming procedure incorporating date as a file extension will be used. Trait and environmental data will typically be recorded in lab notebooks, which will be scanned and stored electronically. Data will be transcribed into Excel spreadsheets. Transcribed data will be spot-checked against either hard or electronic copies to further mitigate data entry errors. Physical samples will be stored in a -80°C freezer at UC Davis. Metadata and protocols will be stored in a general repository on lab computers and backed up in cloud storage (i.e. google drive) for future use by the group. All sequence data and images will be maintained on local servers at UC Davis before for a minimum of 5 years and will be uploaded to public repositories within two years of collection. Analytical pipelines along with bioinformatically processed data will be uploaded to github for version control.

3. Roles and responsibilities

PIs Bay and Sanford will be ultimately responsible for implementing the data management plan, and ensuring that data are properly transcribed, processed, and stored. All project participants will be trained in data management, and granted access to an online data repository that will be created for this project. PIs will subsequently hold one data management meeting each year to ensure that new participants receive appropriate training, and established participants are reminded about internal standards.

PI Bay will oversee the collection and storage of genomic sequencing data and Sanford will be responsible for overseeing appropriate collection and field data and lab data resulting from common garden trait measurement experiments. Compliance with the data management plan will be documented in Annual Project Reports.

4. Dissemination

We will work with the Biological and Chemical Oceanography Data Management Office (BCO-DMO) for data management, including archiving and availability of metadata and linking to other public repositories. Results will be shared via publication in academic journals and presentations at both academic and public forums. Within two years of collection, all data and metadata will be made available online. Raw sequence data will be deposited in NCBI's short read archive and processed data will be available through digital repositories such as DRYAD. Analytical pipelines will be shared on github. Physical collections stored at UC Davis will be made available upon request.

5. Data sharing

All data will be made fully publicly available in online repositories within two years of collection. This will be done through BCO-DMO.

6. Archiving

Physical samples will be stored in a -80°C freezer at the Bay, Grosberg, or Sanford labs at UC Davis for a minimum of 5 years post publication. All electronic data, including sequencing data, images, and metadata, will be deposited in public repositories, but will also be backed up on servers and personal computers at UC Davis. Where possible, processed data and metadata will also be stored in a shared cloud service (google drive or dropbox). Physical laboratory notebooks will be retained in the labs for a minimum of 5 years and scanned electronic copies will be retained for a minimum of 10 years. All spreadsheets will be organized through established naming conventions and a standardized system structure.