Data Management Plan

This Data Management Plan was prepared in accordance with NSF Data Inventory guidelines for Biological Oceanography. The PIs will work with the Biological and Chemical Oceanography Data Management Office (BCO-DMO) to archive and integrate all data from this project and make it available for use. BCO-DMO will assist with archiving ecological data on host density and disease prevalence, and molecular data on genetic diversity.

Types of samples, data, physical collections, and software code

In addition to BCO-DMO, we will utilize the Knowledge Network for Biocomplexity (KNB: http://knb.ecoinformatics.org/index.jsp), a national network whose purpose is to provide an efficient way to access complex ecological and environmental research data. The KNB network provides access to a database that allows development of metadata specification, Ecological Metadata Language (EML: http://knb.ecoinformatics.org/software/eml/), which was developed by the Ecological Society of America and associated efforts (Michener et al. 1997). EML uses XML documents to organize ecological data into individual modules that describe the project's overall metadata. Our proposed research will utilize EML to facilitate data sharing. To submit data to EML, we will use Morpho Data Management software (http://knb.ecoinformatics.org/morphoportal.jsp), which will allow us to manage our data to create suitable metadata modules. These databases are easily accessible as they are based on Java, which works with Windows, Macintosh, and UNIX operating systems, allowing access to multiple operating system users and increasing collaboration efforts with other researchers overall. We will also deposit our metadata in the Biological and Chemical Oceanography Data Management Office (BCO-DMO).

1) Genetic data - Our field surveys will generate SNP and DNA microsatellite data across sites. All data tables will contain columns indicating location (GPS coordinates), date, and time of collection, as well as unique site and grid coordinate identifiers. All microsatellite data will be archived as GenePop-formatted files (http://genepop.curtin.edu.au/help_input.html). We will use microsatellite sequences previously submitted to GenBank (http://www.ncbi.nlm.nih.gov/genbank/), and microsatellite primers and PCR conditions will be published in peer-reviewed journals. Millions of genetic sequences will be produced through HiSeq Illumina sequencing at a 3rd party sequencing center (e.g., University of Texas GSAF) These genetic data are stored at Illumina's online cloud storage service BaseSpace, and will be archived on external hard-drives while we analyze and publish the datasets. Once articles are published, we will upload all sequences to NCBI's Sequence Read Archive database (https://www.ncbi.nlm.nih.gov/sra).

2) Ecological data - Field surveys will generate datasets describing for each of our 6 sites. Data tables will be created for each survey site, and all tables will include columns indicating location (GPS coordinates), sampling date and time of collection, and grid location, to match the morphological data, etc., collected at each of these locations. HOBO logger data series will generate tables of temperature and light for each of the 6 geographical sites for the duration of all experimental outplants.

Field, greenhouse and incubator experiments will generate datasets including site/source location, treatment (as appropriate), sampling date, and the individual responses measured (e.g., % germination, survival, morphology) for each experimental replicate.

3) Software code: Software code used in analysis for publications will be made available via the metadata or other appropriate web repositories (e.g., Dryad) upon publication.

Data Collection, Processing, and Archiving:

Short term: Data sheets will be scanned or photocopied on the day that data are collected, with copies stored in two separate locations. Electronic files will be backed up the day that data are entered, with at least one copy in a remote location.

Long term: All ecological data will be aggregated on a shared cloud server managed by PI Hays. This server will be automatically linked to BCO-DMO and the KNB. Any remaining seagrass tissue samples will be stored in the Hays lab and made available by request.

Standards to be used for data and metadata format and content

We will use standards-based metadata for all measured variables to 1) publicize project datasets through the relevant repositories (e.g. BCO-DMO), 2) classify data according to our data access policy, 3) maintain knowledge of the composition, organization, and quality of the data, and 4) document detailed data schema to ensure easy and complete understanding of the dataset.

Policies for accessing and sharing data

All data will be uploaded to BCO-DMO within two years of acquirement. When quality control and analysis of the data are not complete within that time frame, the PIs will provide the raw data to BCO-DMO and work with the office to temporarily embargo the data until the final data products are ready for final dissemination.

Policies and provisions for re-use, re-distribution, and the production of derivatives

Users will be required to provide their name, affiliation, email address and contact information prior to receiving data, and agree to acknowledge this project and cite the dataset in any publications or derivative projects.