

Data Management Plan

I. Types of data

This research project will generate tissue resources for multiple species of coral and their dinoflagellate symbionts from the Caribbean, Hawaii and Australia. These tissue resources include archived whole samples, and extracted DNA and/or RNA from individuals collected from the field and experimental samples. During the course of these studies we will also generate up to tens of Mbases of metagenetic, transcriptomic and metatranscriptomic sequence data.

II. Data and Metadata Standards

Metagenetic, transcriptomic and metatranscriptome sequences produced by Illumina next-generation sequencers will be automatically converted to FASTA format for archiving and analysis. All data from this multi-disciplinary study will be made available through the institutional web pages of the respective principal investigators. These web sites will be an important portal of entry for the data, reports, photographs, and publications collected during this study. The sequencing data from our field collections and experiments will be deposited for broader dissemination at established data portals (e.g., NCBI, DRYAD, EMBL, DDBJ, CAMERA, SEED, MG-RAST and IMG/M) for short read archives. Additionally, DRYAD, NCBI and the European Nucleotide Archive will be the primary repository for all metagenetic, transcriptomic metatranscriptomic data allowing for data sharing of high quality data much easier.

Metadata, including information on individual samples, such as species identification, sampling location, and other pertinent data, will be digitized using Microsoft Excel. Relevant publications describing the samples and analysis will be stored as .pdf files.

III. Policies for access and sharing and provisions for appropriate protection/privacy

Biological samples, including coral specimens and extracted DNA/RNA will be shared freely upon request. We will also provide the necessary information and protocols for our next generation sequencing methodologies and validation protocols. Processed or annotated data will be made available as soon as possible or right after publication.

Sequence data will be made publically available in a timely manner. Next generation sequence data will be posted to NCBI GenBank, and other archival databases described above, prior to publication. We already have scripts in place to generate and submit sequence data to the NCBI Sequence Read Archive (SRA) in an automated process. Scripts written during the course of developing our bioinformatics pipeline will be made open source and available upon validation.

IV. Provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements

We do not anticipate any intellectual property, or ethical or privacy issues stemming from this research.

V. Policies and provisions for re-use, re-distribution

Access to the data produced by this research will generally be unrestricted. We expect that this data will be of interest to other researchers, to population geneticists, and to genomics researchers. The data could be reused in comparative studies or reviews of the response of marine organisms to thermal stress and in studies comparing coral genomes to other published genomes. We do not see any ethical or intellectual property reasons preventing sharing or reuse of the data.

VI. Plans for archiving and Preservation of access

All tissue resources and extracted DNA/RNA samples will be stored at -80C at the University of New Hampshire for a minimum of five years after the publication of results of this project. Information on individual samples, including species identification, sampling location, and other pertinent data, will be digitized using Microsoft Excel and at least two backups stored away from the master copy. Backups will be stored on the laboratory computers/servers of the collaborative PIs of this project.

The project will be registered with the Biological and Chemical Oceanography Data Management Office [BCO-DMO] and all oceanographic data (and other data types as appropriate) will be formatted for broader dissemination, and redundancy.