

DATA MANAGEMENT PLAN

Data management and sharing are essential to the goals of the proposed research. The PI is committed to working with the Biological and Chemical Oceanography Data Management Office (BCO-DMO), an NSF-funded data repository, to archive and make all data sources publically available. The PI has corresponded with BCO-DMO and is assured that they will be able to assist with archiving and storing data generated from this proposed project. During all years of this project, the PI has budgeted salary time to organizing, managing, and publishing data from this project.

This project will produce large amounts of data in the form of: (i) genomic amplicon sequence raw and processed data, (ii) metadata from the field and aquaria experiments (e.g., latitude/longitude, temperature, PAR, salinity, nutrients, etc.) and (iii) peer-reviewed publications.

Raw data management, archival and storage

Data collection and analysis processes as well as contextual details will be documented in individuals' laboratory notebooks with daily photo back-ups of these pages. All field and experimental metadata will be electronically recorded and managed by data type using the software Microsoft Excel, stored on laptop computers and backed up daily using external hard-drives (in the field) or to a remote location at the Woods Hole Oceanographic Institution (WHOI).

The raw and processed next-generation sequencing data will be stored on a server located in Apprill's laboratory, which is backed-up nightly. Raw amplicon sequence data will be submitted to the National Center for Biotechnology Information (NCBI) Sequence Read Archive depository (<http://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi>). Processed amplicon data, with and without the spike-in controls, will be submitted to the Coral Microbiome Portal, a data visualization and sharing platform for coral and reef-associated microbial data (https://vamps.mbl.edu/portals/coral_microbe/coral.php). This submission will include the raw fastq files, and metadata about the sequences that conforms to MIMARKS, the Genomic Standards Consortium (<http://gensc.org>). Accession numbers for all sequences will be made available in respective publications.

BCO-DMO archival and integration of project data

The BCO-DMO project website will serve as the primary link to all project data. All project metadata will be submitted to BCO-DMO under the award number of this project. PI Apprill will work with BCO-DMO to archive, integrate, and link the genomic repository data and make it available for use.

Publications

Project results will be published in open access, peer reviewed publications with links available to the data (BCO-DMO, NCBI). The publications will also be linked to the BCO-DMO project website.

Policies and provisions for re-use, re-distribution

There will be no embargo periods for political/commercial/patent reasons. Further, there will be no permission restrictions placed on the data. Biological data will be made available following collection and analysis. Data dissemination will be noted in the publications within the Materials and Methods section to inform the scientific community of the data availability and accessibility. All nucleic acid sequence data will be available through NCBI or ENA. These data sources are free of charge and open to the public. We will retain the right to hold data prior to publication only if a conflict of interest seems warranted.

The dissemination of the biological data to be collected for this proposed research will not be restricted by any ethical or privacy issues, copyright concerns or restrictive licenses. As discussed above, all the data collected will be made readily available to the scientific community through various datacenters, published manuscripts in open access peer-reviewed journals, and upon request to the affiliated researchers.