

DATA MANAGEMENT PLAN

The Division of Ocean Science Sample and Data Policy (2011) recommendations (<http://www.nsf.gov/pubs/2011/nsf11060/nsf11060.pdf>) will be followed wherever applicable. All oceanographic, chemical, and hydrographic data acquired by this project will be deposited in the Biological and Chemical Oceanography Data Management Office (BCO-DMO).

Prior data in support of the hypotheses generated

All Imaging FlowCytobot (IFCB) data used in generating the hypotheses in this proposal are publically available through the Gulf of Mexico Coastal Ocean Observing System (GCOOS) data portal- <http://data.gcoos.org/> and our dashboard websites (toast.tamu.edu/IFCB7 or toast.tamu.edu/IFCB111). We will add a link to these data in BCO-DMO for this project. Additional data and analyses are available in the publication of Angles et al. 2015 (see References). All prior transcriptome data have been deposited in the NCBI Short Read Archive and are also available in the iMicrobe database for the Moore Foundation's Marine Microbial Eukaryote Transcriptome project. We will add links to these data in BCO-DMO, also.

Types of data to be produced

1. Digital images from the IFCB
2. Processed data from images
3. Illumina RNA-seq transcriptome short reads
4. Seawater nutrient and CO₂ concentrations
5. Phytoplankton abundance estimates

Standards to be used for data and metadata formatting and content

Written records (lab books) will record all sample information (dates, sample identification numbers, environmental conditions). Recommended metadata (e.g. sequencing platform, RNA sample date/location, environmental conditions) will be recorded for all metatranscriptomes.

Steps to protect privacy, security, confidentiality

Data will be maintained and released in accordance with appropriate standards for protecting privacy rights and maintaining the confidentiality of respondents, if necessary. We do not expect to encounter intellectual property or copyright issues.

Sharing research resources

The data collected on this project will primarily be disseminated through the process of peer reviewed publication in the literature and through the GCOOS data portal. In addition, our findings will be presented at national and international scientific conferences. Information not distributed in this way will be freely available from the PI via written request. Backups of such files will be stored on a local hard disk for at least three years.

Methods for archiving and preserving access to data and materials

Imaging FlowCytometry data will be freely available via a webserver at TAMU/Geosciences (<http://toast.tamu.edu/>). Analysis is performed using software developed by H. Sosik and J. Futrelle (Woods Hole Oceanographic Institution) and freely available on their website. All data will be backed up to external hard drives and stored in a separate facility.

All transcriptome raw sequence data generated in this project will be deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (<http://www.ncbi.nlm.nih.gov/sra>) within a

year of completion of annotation. All data (assemblies, annotations and associated metadata) will be included in published manuscripts that will specifically state how the data can be accessed or reused.