

Data Management Plan:

In accordance with OCE guidelines, all data generated during this project will be organized through (and, if possible, deposited into) the Biological and Chemical Oceanography Data Management Office (BCO-DMO). All project metadata (site descriptions, sampling protocols, environmental physical and chemical data, publications) will be accessible via a project-specific BCO-DMO page, with links to metabolomics data (described below) and to external repositories containing sequence data (e.g., NCBI SRA, described below). Entries to BCO-DMO will be made within 2 months post data generation, with all data products made publicly accessible within 2 years of collection. PI Hay will oversee all aspects of metadata management via BCO-DMO, with Co-PI Stewart managing molecular sequence data archiving. The PIs, or senior members of the PIs' labs (trained in archiving protocols), will share archiving responsibilities as needed if one of PIs leaves the project.

Types of data: This effort will produce data on microbiome composition of various corals growing under differing biotic conditions, the growth, survivorship, and photophysiological health of those corals, and the anti-pathogen activity of mucus from these corals. There will also be video of fish activities impacting our experimental replicates.

Data acquisition, quality control, and dissemination. Before each field expedition, the research team will meet to plan objectives, experimental design, sampling methods, and analyses of samples. Our research will generate detailed physiological (growth, survival, PAM readings, % bleaching), microbiome, and bioassay data. Outcomes of experiments will be published as results (in figures and/or tables) in peer-reviewed journals. Prior to publication, these will be stored as noted below; upon publication, all appropriate data will be deposited on the BCO-DMO data system at <http://www.bco-dmo.org/>. All sequence data will be made publicly accessible within one year of generation (see below).

Sequence data and bioinformatics - All -omic data will be filtered using default protocols associated with the sequencing platform. Additional quality filtering will be imposed using protocols established in the Stewart lab, or via downstream analytical processing (e.g., filtering sequences based on Illumina error profiles using Deblur). Rapid dissemination of sequence data and associated metadata will be a priority. Our proposed amplicon and transcriptome sequencing involves multiple runs on Illumina MiSeq and HiSeq instruments, respectively, ultimately generating hundreds of gigabases of sequence. Rapid dissemination of these data to the broader community will be a priority - all sequence data will be made publicly accessible within one year of generation. We will fully abide to the Minimum Information about a Genome Sequence and Metagenomic Sequence standards that have been recently established by the scientific community (MIGS and MIMS, respectively). Following automatic quality assurance filtering on the Illumina system, demultiplexed raw sequencing data with combined quality scores (FASTQ format) will be archived and stored on servers at Georgia Tech and for public access in the Sequence Read Archive (SRA) at the NCBI. SRA data will be assigned a single BioProject identifier with linked metadata. Our submissions will be annotated with detailed descriptions of the sampled environment or experimental treatments (project description, lat/long, date, habitat type, etc), including brief summaries of any associated biotic variables. Additionally, pdf copies of all protocols used in the generation of sequence data (if not prohibited by manufacturer copyright restrictions) will be linked to the data submissions, either directly or via instructions for accessing copies on the PIs website. Sanger-based sequences will be submitted to NCBI's GenBank and annotated with associated domain characterizations and metadata as above. All analysis results will be stored using internal data formats within the utilized software packages and common formats such as TIFF, ASCII, and Excel. Whenever possible, we will use portable data formats (such as ASCII files) to facilitate data exchange and subsequent analysis. All multivariate analysis data and comparative transcriptomic data will be

encoded both as Excel files and as R/Matlab binary images. We anticipate sharing bioinformatics code for any custom work flows developed as a part of this project. This code will be written up in a data format that is easily sharable, e.g. IPython notebooks, and included with published results.

Field and bioassay data – Field data (coral growth, survivorship, bleaching, PAM readings, disease spread, etc.) will be archived on annotated spread sheets and deposited with Biological and Chemical Oceanography Data Management Office (BCO-DMO) (<http://www.bco-dmo.org/>) upon manuscript acceptance, as has been our common practice for past grants.

Digital copies of all data will be kept in the labs of the PIs for 10 years past the lifetime of the project. The anticipated data occupies a relatively small amount of space compared to our computing capabilities and storage resources; thus, long-term preservation will be easily accomplished by keeping several copies of the data on local computers at Georgia Tech. Should any PI leave the project or employment, they will retain access to their data, as will GT and the other PI via the storage system described above.

Data sharing among partner groups. The Hay and Stewart labs are co-located in the same building and we commonly collaborate and co-publish. We will hold regular meetings to discuss progress and integrate new discoveries. Students and post-docs will present at national and international meetings and submit research to peer-reviewed journals. Raw data, data summaries, and presentations will be maintained on servers in the School of Biology and made available to all participants: <http://omz.biology.gatech.edu/> To provide a back-up and facilitate easy access, we will also archive and share data on Dropbox (Ga Tech has a license providing unlimited, ITAR-compliant cloud storage for faculty and their immediate collaborators).