Data Management Plan

We will use the Biological and Chemical Oceanography Data Management Office (BCO-DMO) as our primary data management source. After our award has been initialized, we will contact BCO-DMO and register our project, and each year submit dataset metadata followed by the data themselves.

The PIs have a strong track-record of making their data available through timely publication in peer-reviewed, ISI-rated, scientific journals. Raw datasets are published as Supporting Information whenever possible (see Loh & Pawlik 2014). When fiscally prudent, articles are published open-access; otherwise, they are available upon request.

(1) Types of data, samples, and other products.

The primary *product* of this research program will be publications that fully present, analyze and discuss the data listed below:

- Flow data
- Flow cytometry data
- Nutrient analysis data
- Isotopic mass spec data
- Next generation (Illumina platform) DNA sequence data – bacterial and archaeal 16S rRNA from sponge symbionts

In addition, video footage for the purposes of public outreach will be posted on the platforms listed in the Broader Impacts section of this proposal.

(2) Standards for data format and content.

Analytical data will be stored as their appropriate file types, with DNA sequence data stored as raw output files (abi and sff) and as final files (txt in fasta format) following standard quality control methods. Data will be hosted on an internal server at the Center for Marine Science, UNCW, with management by UNCW's IT department. All digitized data will be backed up on external hard drives both on-site (PIs' computers) and off-site to ensure data preservation and availability.

(3) Policies for access and sharing.

In addition to using BCO-DMO, publications resulting from this project will be readily available through library databases, and will be published open-access when possible. Raw datasets will be posted online to international databases (see below), as appendices or supplementary material in publications, on the PIs websites or upon request, and within 2 yrs of data collection. DNA sequence data with relevant metadata will be deposited in international databases to provide global, online access to molecular data, including the National Center for Biotechnology Information's (NCBI) GenBank sequence data archive (www.ncbi.nlm.nih.gov/genbank/) and NCBI's Sequence Read Archive (SRA) for next generation sequencing projects (www.ncbi.nlm.nih.gov/sra).

(4) Policies for re-use and re-distribution.

While we believe that the results of this research program will best be accessed as published journal articles, we will maintain datasets for provision to other researchers for comparative studies or meta-analyses upon request. DNA sequence data in particular are commonly re-used from international databases and we will encourage such meta-analyses by archiving sequence data at NCBI. Further, as Next Generation Sequencing datasets are deposited at NCBI in raw output format (sff), selected sequence files will be posted on the PIs' websites to allow rapid, online access to final datasets (e.g., screened, filtered and clustered pyrosequencing reads) and to complement the raw data deposited at NCBI.

(5) Plans for archiving data, samples and other products.

All persons associated with this research program, from PIs to undergraduate students, are responsible for preserving data quality and integrity by following standardized practices for scientific research. UNCW has a **required** institutional plan to provide training and oversight of the responsible and ethical conduct of research, with RCR online training through the Collaborative Institutional Training Initiative at the University of Miami. In addition to archiving molecular sequence data in international databases along with relevant metadata to provide global availability, the PIs will have a joint repository for project data, as well as individual archives, all available upon request.