Data Management and Access Plan

Data management will be coordinated by the Principal Investigators and carried out by all project participants. The project will follow the Division of Ocean Sciences (OCE) Sample and Data Policy. We will use the Biological and Chemical Oceanography Data Management Office (BCO-DMO) as the primary data management archive for the project. For data that cannot be served by BCO-DMO, or where they are the more appropriately served by other community data repositories, metadata will still be deposited in BCO-DMO with links to the other data repositories.

1. Types of Data- This project will produce and organize the following types of data:

Digital- High-speed videos will be converted into AVI (Audio Video Interleave) files, given that this format is easily readable for both PC and Mac users. 3D renderings from microCT scans will be shared in popular formats (e.g., STL, OBJ) that are widely used in 3D printing, academia, engineering, etc. When necessary, metadata will be appended in "readme files" to explain filming/scanning parameters, variables, units, etc.

Protocols- We will continually refine and update experimental protocols.

Experimental-, In our laboratory flume and respirometry experiments as well as in field manipulations, we will collect data on gaping behavior, byssus production, growth, survival and community composition.

Physiological- We will collect data on mRNA abundance, single nucleotide polymorphisms, oxidative stress, enzyme activity, thermal tolerance, metabolism (e.g., through respirometry), lipid stores, and biomaterial properties.

Environmental data- We will collect seawater temperature, pH, dissolved oxygen, water velocity, and salinity data. We will collect bottle samples for carbonate chemistry analysis. We will also collect air temperature, solar radiation, wind speed, relative humidity,
Model code- We will produce code for our models, primarily in R and Python (open source).

2. Data standards- We will use a shared Google Drive to initially share our modeling, laboratory and field data (as csv files with associated metadata) and protocols (text documents). We will develop our model code in public repositories version controlled at github.com. For hydrodynamic instrument data, original data files will be collected in binary format that is proprietary to the instrument manufacturer or computer code. This data will be converted after download to a computer to ASCII format and saved with a *".txt"* extension. This format is easily read across multiple computer platforms and is easiest to analyze in multiple statistical programs (including R, Matlab, and Excel). For respiration, gape behavior, sequencing, and qPCR experiments, data collected using proprietary instrumentation will be similarly converted into ".txt" file formats.

3. Responsibilities- PI Carrington, in collaboration with Co-PIs Reidenbach and Nishizaki, will be responsible for ensuring implementation of the data management plan and will train all project participants in order to ensure adherence to the plan. They will review all data products, ensure that they are technically sound, and that they are publicly archived upon project completion.

4. Dissemination- Upon completion of the project (or portions thereof), data will be published in peer-reviewed journals and made publicly available through BCO-DMO. Some data may be more suitable for other online databases, such as DRYAD for experimental data or github.com for statistical and model code, and these will be linked to BCO-DMO.

5. Data sharing and reuse policies- Data will be available to all participants through a shared Google Drive. Data from the projects will be made available for any scientific and professional uses. Upon publication, all data will become "open access" (e.g., on a web page or in an online repository). We make this policy explicit by stating in each publication resulting from this grant that the data are available

according to the guidelines of the Joint Data Archiving Policy (JDAP: http://www.datadryad.org/jdap). Sequence data, including any SNPs identified will be submitted to a public sequencing repository (e.g., Genbank). We have not and will not ask for co-authorship as a condition for sharing data, only that when our data are used, the paper in which they originally appeared should be cited as the source of the data. Data used in our publications will be made available immediately upon publication of the paper in final form (in print, or official online publication). The only exceptions will be when we have already planned additional analyses of those data for use in a subsequent publication. In that case, data used in a publication will be made available unconditionally within two years of its first journal appearance. Any data not published will become available within two years of completion of the grant. There are no ethical or privacy issues with regard to these data and none of the datasets fall under IRB use or protocols. None of the datasets collected will be covered by copyright or licensed.

6. Data archiving- We will take several steps to ensure long-term preservation, oversight, and access to data after the PIs are no longer working on this project. As discussed above, data and code will be submitted in a timely manner to online databases intended to be maintained in perpetuity. In the field we will back up data using several external hard drives (kept in different places) to prevent data loss. We will archive large files (e.g., raw high-speed videos and microCT scanner files) using both physical and cyber resources (e.g. Google Drive). Additional physical copies will be kept at the PIs laboratories; these facilities are well safeguarded against fire and water damage.