DATA MANAGEMENT PLAN

The results of the proposed research will be disseminated via publication in peer-reviewed journals, presentations at scientific meetings, press interviews, and public presentations. Once data are published (or within 2 years of the completion of the project) data will be made broadly available via the appropriate mechanisms detailed below for each type of data we plan to collect. The PIs will work with Biological and Chemical Oceanography – Data Management Office (BCO-DMO) staff to manage and archive the data with appropriate standards.

(1) TYPES OF DATA, SAMPLES, AND OTHER PRODUCTS.

MORPHOLOGICAL DATA including diagrams and digital photographs will be retained in perpetuity in the original notebooks residing in the PIs laboratory and on computer hard drives specifically for archived data. In addition, data products as they are generated will be immediately stored in files in the cloud on Dropbox. JPEGs of all photographic data will be posted to the PI's lab websites.

GENETIC DATA on the gene and genotype frequencies from *Nucella* adults and juveniles are the main data of this type. We will primarily use microsatellite markers to provide this genetic information (with the possible addition of SNP data). We will submit all primer sequences used in these studies to GenBank (www.ncbi.nlm.nih.gov/Genbank/), along with each unique microsatellite repeat motif at each locus. Methods for developing and applying these microsatellite markers will be posted to the PI's websites, along with references to, and applications of, the statistical methods used to estimate relatedness among individuals. Population-specific allelic and genotypic frequencies will also be posted to the PI's website in Excel format, along with metadata that completely describe sampling methods and organization of the database. Full data will be supplied to the Biological and Chemical Oceanography Data Management Office (BCO-DMO) facility when they are finalized. Voucher specimens from all of the collecting sites will be stored in EtOH in the collections at Bodega Marine Lab.

(2) STANDARDS FOR DATA AND METADATA FORMAT AND CONTENT

Datasets collected will include a consistent set of metadata to be agreed upon and finalized at the end of **YEAR 1**. Minimally each dataset will include a title, description, location (latitude & longitude via GPS), date, time (UTC), PI name and contact information. Each data set will also include, as appropriate, standards used for measurements, instrumentation used, calibration information, analytical methods used, data processing information, sampling procedures, and access restrictions. Data will be entered into a standardized digital format and in consultation with BCO-DMO staff (most likely Excel spreadsheets and then converted into comma delimited files for submission to BCO-DMO for final archiving). For example, currently our metadata summary resides as a separate data sheet for each dated entry to the data workbook, with the responsible party and explanation. Primary metadata for each experiment will include, for each experiment or data sheet in the workbook, a log of comments and explanation of the instruments, operator, formulas, and units. Any changes made to the log must be saved as a separate sheet with date. A standard file-naming procedure incorporating date as a file extension will be used. 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; POLICIES AND PROVISIONS FOR RE-USE, RE-DISTRIBUTION, AND THE PRODUCTION OF DERIVATIVES

Project Participants. All data collected will be backed up and made available to all PIs, graduate students and research technicians (and additional project staff as appropriate) via shared, limited access *Dropbox* (www.dropbox.com) folders after initial quality control and assurance processing by individual investigators. The primary responsibility for uploading data in a timely fashion will reside with the SRA and PIs supported by this project.

Broader Scientific Community. 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, as appendices or supplementary material in publications. When the project is complete, full data will be posted to a public access scientific database via the BCO-DMO within two years of collection, as required by NSF, as appropriate. We will consult with BCO-DMO staff to identify

appropriate data for submission, and appropriate standards. 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.

(4) PLANS FOR ARCHIVING DATA AND FOR PRESERVATION OF ACCESS

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. The primary program through which PI Grosberg accepts graduate students, the Population Biology Graduate Group, also requires training in the responsible conduct of research (http://research.ucdavis.edu/c/cs/rcr). This program has a specific module on Data Acquisition Management, Sharing and Ownership. Grosberg has been a regular instructor in this program. In addition to archiving molecular 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.