DATA MANAGEMENT PLAN - CORNELL

We are committed to broad dissemination of our data and metadata during the timeframe of this project (within 2 years of data collection). All project personnel will be trained in data management procedures in accordance with NSF requirements (http://www.nsf.gov/pubs/2011/nsf11060/nsf11060.pdf). Our primary repository will be Cornell eCommons and we will adhere to their data management policy (https://ecommons.cornell.edu/page/policy).

PRODUCTS OF THE RESEARCH: This project will generate physical specimens including DNA samples and archived tissue from field-collected and experimental animals, and shell voucher specimens for a subset of analyzed individuals. Data will consist of genomic sequences, phenotypes (including digital images), environmental measurements, quantitative genetics model code and output, and subsequent analytical products and archives of all data. The effective transmission of data from field, to laboratory, analyses, and archive will be handled through data management pipelines specific to each data type managed by the relevant PI/co-PI.

1) Physical specimens and DNA (Hare): When sufficient tissue is available a sample will be preserved in ethanol and archived in the Hare Lab in addition to remainder genomic DNAs archived at -80°C. Voucher shells of analyzed *C. virginica* individuals will be accessioned into bivalve collections at the Paleantological Research Institute in Ithaca, NY.

2) Phenotype data/ metadata (Munroe & Hare): For each specimen collected and phenotyped, and/or genomically analyzed we will record collection date, source coordinates, source temperature and salinity, and shell height (to the nearest mm). Spat sizes during challenge experiments will be monitored with digital photography. Photos will be archived in JPG and TIFF formats and associated with field collection metadata. All data will be backed up on external hard drives. The final spreadsheets georeferencing samples with phenotype, genotype and environmental data will be submitted with appropriate metadata to the Biological and Chemical Oceanography Data Management Office (BCO-DMO) within one year after data analysis is complete.

3) Genetic data (Hare): We expect to generate variant call data for 1350 individual oyster samples and 39 pools based on 2400 Gigabases of Illumina paired-end sequence data. Raw fastQ sequence data will be backed up on external hard drives. We will make raw sequence data and BAM mapping files publically available through the National Center for Biotechnology Information (NCBI) Bioproject, Biosample and Short Read Archive system.

4) Geographical data (Munroe): For each specimen collection trip, all geographical (GPS coordinates) and associated metadata (weather conditions, vessel used, crew, environmental data) will be recorded using an on-board data logger (with backup waterproof paper copies). We will use open-source OpenCPN for plotting locations of all collections (tow start and end points for any dredge and plankton collections, and waypoints for all deployed collectors). All geographical data will be archived in co-PI Munroe's lab and backed up on Haskin Lab servers as well as external backup devices.

5) Physical measurements (North/Munroe): PI North will be responsible for collecting data on the physical and biological characteristics of the water column on 8 cruises per year during the first two years of the grant. An SBE25 SEALOGGER CTD equipped with a fluorometer (WET Labs WETstar), an Optical Backscatter Sensor (OBS, D & A), dissolved oxygen (SBE 43) will be used. Fifteen water samples will be collected on each cruise to calibrate the flourometer and OBS to chlorophyll-a and Total Suspended Solids,

respectively. All sensors will be calibrated before each field season and calibration constants will be recorded in the metadata description of data collection and processing. CTD data will be processed and QAQCed, and submitted with appropriate metadata to the Biological and Chemical Oceanography Data Management Office (BCO-DMO) within one year of collection and processing.

6) Larval salinity tolerance experiments (North): PI North's graduate student will assist with the salinity tolerance experiments and analyze the data. Physical variables, larval abundances, and sizes will be tracked through time, entered into spreadsheets, and QAQCed. The final spreadsheets will be submitted with appropriate metadata to the Biological and Chemical Oceanography Data Management Office (BCO-DMO) within one year after data analysis is complete.

STANDARDS TO BE USED AND METADATA FORMAT AND CONTENT: Specimen data will conform to the Darwin Core (DwC) metadata standard. This standard has been internationally adopted and is used by the Ocean Biogeographic Information System. Evidence for selection on SNPs based on various statistical tests will be treated as a continuous variable with the top 30% of SNP loci listed in a data archive with test statistics and significance values as recommended by Lotterhos and Whitlock (2014).

DATA ACCESS, SHARING, AND PRESERVATION: All data produced during this research will be freely available to the public; we anticipate no sensitive or confidential data. Cornell eCommons will be a permanent repository for all project data.