DATA MANAGEMENT PLAN

I. Types of data: samples, physical collections, software, derived models, curriculum materials, and other materials to be produced in the course of the project. Records will be comprised of, transcriptomic data, experimental metadata, phenotypic observations. Sequence data will be submitted to the **Biological and Chemical Oceanography Data Management Office (BCO-DMO)** and National Center for Biotechnology Information (NCBI)'s publicly available databases Gene Expression Omnibus (GEO) or Sequence Read Archive (SRA), within 2 years of the data collection. Modelling software scripts for data will be made available through open source software repositories (i.e., GitHub), publications, and at the request of other researchers.

II. Data and Metadata Standards
Experimental metadata, phenotypic observations and transcriptomic data, will be stored in different formats (such as tab-delimited, FASTA, FASTQ, BAM, xls, nexus, msf etc.) to ensure compatibility between different programs and databases. We will collect contextual details such as experimental set up, library construction and sequence analysis, alignment strategies (e.g. (e.g. paired end), log files, and search databases along with other metadata standards. We will adapt metadata standards that are compliant with MIxS (Minimum Information About any (X) Sequence) standards (http://gensc.org/projects/mixs-gsc-project.

III. Policies for access, sharing and provisions for appropriate protection/privacy
All raw and processed data in the form of web accessible databases, R packages, Java packages and applications and in other forms will be available on through indexed public websites and databases, (and our own website) and freely accessible through NCBI archive. Data privacy, confidentiality and security will be adjusted according to publication timeline.

IV. Policies and provisions for re-use, and re-distribution
There will be no permission restrictions on data or vector assemblies, they will be open for sharing and re-use upon publication.

V. Plans for archiving of data
Collected and generated data will be curated both automatically, manually and computationally before made available. Periodic data backups and submission to online repositories such as BCO-DMO and NCBI's publicly available databases GEO or SRA will be done according to suggested journal and archive guidelines. All captured metadata and protocols will be documented and available in the form of references and research papers. Data preservation beyond the life of the project will be determined on case-by-case bases and will be backed up on tapes for long-term storage.

VI. Algorithms and Interpreted Data
Any algorithms developed *for this project* will be published as part of conference proceedings or research papers. Code and data will be made available on the website and open source software repositories.