Data Management Plan

Introduction:
Data management will be coordinated by PI Marchetti and carried out by all project participants. The plan encompasses two areas: use policies, and data preservation and archival. We will leverage existing systems wherever possible. All data will be communicated in a timely fashion following the NSF policy.

Two main types of data will be generated in the proposed activities: transcriptome sequence libraries and targeted gene expression in laboratory isolates and field samples. Transcriptomic sequence libraries will be created using the Illumina HiSeq 2000 platform. Gene expression analysis will be performed using quantitative PCR. Other data generated in this project includes physiological measurements on phytoplankton growth characteristics and phytoplankton composition (e.g.18S clone libraries) of field samples.

Data access and sharing policies:
We are committed to making all data types publicly available through peer-reviewed publications and public databases with as few restrictions as possible. As sequence libraries are created and samples are analyzed, data will be processed, and raw and processed data will be uploaded to networked servers maintained at UNC-Chapel Hill. These servers are backed up weekly, and can be made available to collaborators needing access to the data. Data will be transferred to Biological and Chemical Oceanography Data Management Office (BCO-DMO) following processing, and public access will be granted to data following its publication or at most two years after its collection.

Plans for archiving and preserving data:
Sequences obtained through Sanger DNA sequencing or High-throughput Sequencing platforms will be stored in long-term storage space provided to Marchetti by UNC-CH Research Computing and deposited in public sequence databases: all individual gene sequences (e.g. ferritins, ISIPs, etc.) will be deposited in the National Center for Biotechnology Information (NCBI) GenBank and all sequences obtained through Illumina sequencing will be deposited into the Sequence Read Archive (SRA) managed by NCBI. All other data will be archived at BCO-DMO. In all our efforts we will work with the BCO-DMO (http://www.bco-dmo.org) to archive the data and to ensure our metadata conform to their standards.