Data Management Plan

1. Types of Data

This project will generate experimental data for Rhodopsin from cultured and natural occurring (Long Island Sound) dinoflagellates.
There are two types of data.
A) Rhodopsin cDNA (for cultures) and gene (for LIS assemblages) sequence dataset.
B) Rhodopsin gene diversity and abundance over spatial and temporal scales in Long Island Sound.

2. Standards and Formats

Preparation of cDNA and genomic sequence data will follow standards and formats required by the major data repository, National Center for Biotechnology Information (NCBI). Data submission to NCBI's GenBank will follow procedure described in the following,

```
For GenBank submission information:
```
http://www.ncbi.nlm.nih.gov/genbank/submit.html;

```
For The source modifiers:
```
http://www.ncbi.nlm.nih.gov/WebSub/html/help/genbank-source-table.html

```
For features annotation:
```
http://www.ncbi.nlm.nih.gov/projects/collab/FT/index.html#7.4.5.

```
Submission to CAMERA will follow instructions in the following,
```
https://portal.camera.calit2.net/gridsphere/gridsphere?cid=Data+Ingestion+Portlets.

Spatial and temporal distribution pattern of Rhodopsin gene will be evaluated for suitability and worthiness to be deposited to NCBI Gene Expression Omnibus (a public functional genomics data repository supporting MIAME-compliant data submissions (http://www.ncbi.nlm.nih.gov/geo/).

The following resources will be consulted when preparing the data:

- Standard Reference Data (NIST) http://www.nist.gov/ts/msd/srd/
- Environmental Data Standards (EPA) http://www.epa.gov/fem/data_standards.htm
- Earth Science Data (NASA) http://science.nasa.gov/earth-science/earth-science-data/
- Formats for Long Term Access (MIT)http://libraries.mit.edu/guides/subjects/data-management/formats.html
- For information about using metadata: Understanding Metadata (NISO) http://www.niso.org/publications/press/UnderstandingMetadata.pdf; Metadata : How to Describe Your Data (University of Minnesota) http://www2.lib.umn.edu/data/management/metadata.phtml

3. Provisions for Archiving and Preservation

All the raw data will be saved in two storage hard drives at the PIs' laboratories. The cDNAs and genomic sequences of Rhodopsin will be deposited into NCBI GenBank database for public usage. No proprietary software is required in formatting or restoring the dataset. Therefore, the stored data will remain retrievable in the future. The following resources will be consulted,

- Backups and Security (MIT) http://libraries.mit.edu/guides/subjects/data-management/backups.html
- Stewardship and Archiving (University of Minnesota) http://www2.lib.umn.edu/data/management/archiving.phtml

4. Access Policies and Provisions

All the annotated clean sequence reads will be deposited into NCBI GenBank and CAMERA databases for free access by all users. The raw data can be shared to researchers who are interested when requested.

5. Plans for Transition or Termination of the Data Collection

The annotated data will be active after the project is finished and discovery has been published in the peer review journals. NCBI will store these data for public use. PIs will store two copies of the data in storage hard drives. We do not have a schedule to destroy our data.