

Data Management Plan

ProPortal database: enabling data access, integration, and analysis

Our work on *Prochlorococcus* over the past 20 years has produced a lot of data on this organism, including genomic, transcriptomic, proteomic, and physiological data, along with information about global ocean distributions. Our long-term goal is to build a customized data management system to organize and integrate *Prochlorococcus* related data. This system is designed to allow our framework to evolve and adapt as new data types, such as information gathered in this proposal about *Prochlorococcus* vesicle content and abundance, become available. We have begun this effort with the construction of a web site, *Prochlorococcus* Portal

(<http://proportal.mit.edu/>) [6] that

currently houses access to completely sequenced genomes, microarray data, environmental cell distributions, metagenomic data from the Global Ocean Survey (GOS:

<http://www.jcvi.org/cms/research/projects/gos/overview/>), and publications

(including supplementary files) from our group. We will continue to build the Portal with the data emerging from the proposed work, and integrate it into the database (Fig. 1). The

ProPortal website is currently divided into four main sections for data access: Genomes, Environmental Cell Distribution, Microarrays, and Metagenome. All of these data are publicly available.

ProPortal website is currently divided into four main sections for data access: Genomes, Environmental Cell Distribution, Microarrays, and Metagenome. All of these data are publicly available.

ProPortal, an integrated resource for the *Prochlorococcus* system

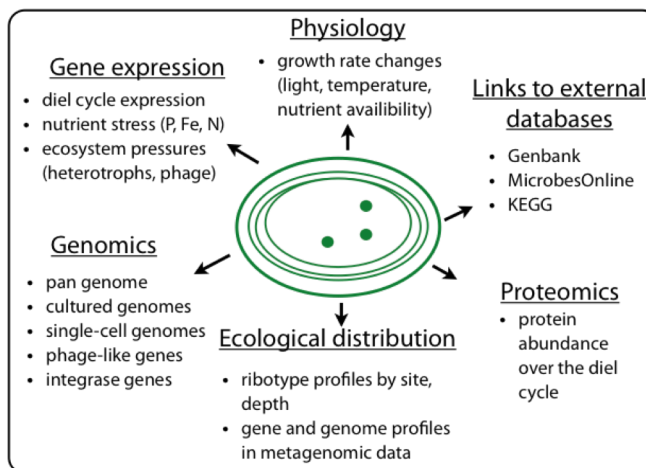


Figure 1, Data sources in the ProPortal database. ProPortal integrates genomic, taxonomic, and functional data into a public, searchable, and flexible resource.

Database development and public access to database schema

ProPortal is built and maintained with Django (<http://www.djangoproject.com/>), a high-level, Python-based web framework. While the data presented in ProPortal is specific to *Prochlorococcus*, *Synechococcus*, and cyanophage, the database framework is generalizable for any group working on microbial and phage genomes for which extensive metadata is available. We therefore provide access to our database schema (<http://proportal.mit.edu/schema/>) should other groups want to replicate the whole database or portions of it to maintain their own data.

Vesicle field and content data submission

Links to data collected from field surveys of membrane vesicles will be made accessible through the Biological and Chemical Oceanography Data Management Office (BCO-DMO) database. We have registered the project at the following URL: <http://www.bco-dmo.org/project/472071>. Sampling logs and other relevant field site metadata will be provided to BCO-DMO.

Along with making the vesicle data available through BCO-DMO and ProPortal, we will also submit information gathered on vesicle proteomics and nucleic acid content to the Vesiclepedia database (<http://www.microvesicles.org/>) [7], a new, open resource developed by the extracellular vesicle research community. No bacterial datasets available in Vesiclepedia as of January 2013, thus the information we produce on *Prochlorococcus* vesicles will significantly diversify this resource.

Sequence data submission

All sequence data produced will be submitted to Genbank (<http://www.ncbi.nlm.nih.gov/genbank/>). Furthermore, we provide links *via* ProPortal to directly submit NCBI BLAST jobs for genes of interest, and to search for genes specific to particular NCBI-defined gene clusters. Gene identifiers are also linked to the MicrobesOnline (<http://www.microbesonline.org/>) database, which provides extensive annotation related to gene structure, function, and phylogeny.

Distribution of Cultures and DNA, and other information

We donate our cultures to the CCMP at the Bigelow Labs (<https://ccmp.bigelow.org/>) for distribution, and also supply them directly to researchers. We also supply DNA from our cultures upon request. Our public website (<http://chisholmlab.mit.edu/>) provides contact information for interested groups, as well as protocols, a list of cultures, publications, and genomes available.