Computational and Experimental Approaches to Systems Biology of Marine Green Algae

Marijke J van Baren[1], Peter Waltman[2], Daniel Sam[2], Emily Reistetter[1], Samuel Purvine[3], Charles Ansong[3], Stephen Callister[3], Joshua Stuart[2], and Alexandra Z. Worden[1,4]

Institutions:
[1]Monterey Bay Aquarium Research Institute, Moss Landing, CA
[2]Biomolecular Engineering Department, University of California Santa Cruz, Santa Cruz, CA
[3]Pacific Northwest National Labs, Richland, WA
[4]Ocean Sciences Department, University of California Santa Cruz, Santa Cruz, CA

Presenter: Marijke J van Baren,
PI: Alexandra Z. Worden, Monterey Bay Aquarium Research Institute & UCSC, azworden@mbari.org
Co-PIs: Stephen Callister, Richard Smith, Joshua Stuart

Project Goals: To develop a systems biology approach to the study of the widespread marine alga *Micromonas* and use it to investigate gene function, pathways and consequences of environmental perturbations on primary production.

URL:
http://www.mbari.org/staff/azworden/

Abstract:
We are developing a model system for eukaryotic marine algae. Thus far there are no model systems relevant to both marine carbon cycling and evolution of the Viridiplantae – the eukaryotic lineage containing all land plants and green algae. The primary drivers for developing the dual-*Micromonas* system are: *i*) approximately half of global photosynthetic CO_2 uptake is performed by marine algae yet there is little understanding of the physiological consequences of current global change scenarios and *ii*) green algae provide insights to eukaryotic cellular processes and the ancestor of land plants.

Prasinophytes are a group of unicellular marine green algae that are evolutionarily distinct from the model green alga *Chlamydomonas*, but are related to both the latter and land plants. *Micromonas* is a widespread prasinophyte that is exceptional in its size (<2 micrometer diameter) and having a small genome (21 Mb). The genomes of two *Micromonas* strains share 90% or less (depending on criteria used) of their protein encoding genes, have low gene redundancy and contain gene 'fusions', which join together domains typically encoded by separate genes. These features are valuable for investigating and assigning functions to genes and domains by their association with a known pathway or physiological response. Our strategy is to subject the two different strains to ecologically relevant perturbations and use whole transcriptome and proteome profiling, as well as traditional cellular measures, to understand their responses to perturbation and develop knowledge of cellular pathways. The genomes of *Arabidopsis thaliana*, *Physcomitrella patens*, *Chlamydomonas reinhardtii*, *Ostreococcus* RCC809, *Saccharomyces cerevisiae* and *Escherichia coli* are being used for comparative purposes.

To investigate genetic pathways in *Micromonas*, a mixed computational inference strategy is being utilized that integrates data from diverse sources, including the transcriptome and proteome expression. As a first step, we have implemented a scoring system to track supporting evidence for all predicted genes in the two strains. Every possible annotation is scored, in a database that holds all annotations and their evidence, and the highest scoring gene at a locus then selected automatically. We are currently expanding the database to hold all biologically relevant information for different experimental conditions. This database provides a foundation for pathway and perturbation analyses. Interaction orthologs (interologs) are included to help identify putative interaction pathways. An integrative multi-species biclustering algorithm is also being used to identify putatively conserved modules that are shared by the two *Micromonas* species, as well as species-specific differences in these modules and modules that are unique to each organism. In turn, these modules were used in conjunction with several network-inference methods to identify additional putative pathways whose activity levels were identified using PARADIGM, a state-of-the-art pathway prediction algorithm. Given its relationship to plants the dual-*Micromonas* system will enable modeling of more general primary producer responses across ecological and evolutionary

scales. Our overall goal is to develop an efficient system for gaining insights to the green lineage, including novel and conserved genetic mechanisms, with high relevance to marine carbon cycling. Development of such a system is important given the onset of climate change and limited understanding of how earth systems will move forward under current perturbations.

Funding:
This research is supported by U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Biological Systems Research on the Role of Microbial Communities in Carbon Cycling Program Award No. DE-SC0004765 made in July 2010 with input from other funding sources.