Multivariant RNAseq reveal novel players in algae nitrogen metabolism

Stefan Schmollinger[1*](schmolli@chem.ucla.edu) and **Sabeeha S. Merchant[1,2]**

[1]University of California, Los Angeles, CA; [2]University of California, Berkeley, CA

URL: http://www.chem.ucla.edu/dept/Faculty/merchant/#research

Project goals: The single-celled, eukaryotic green alga *Chlamydomonas reinhardtii* is an excellent model to study plant metabolism while offering fast generation times and all the advantages of microbial systems. All the genomes are sequenced and well annotated, allowing for large scale transcriptomic or system biology analyses. In order to discover novel genes involved in essential metabolic processes and further the understanding of the alga we generated a multivariant RNASeq dataset within the JGI Gene Atlas project, with the goal to compare different carbon and nitrogen sources, growth regimes, light intensities and cell densities in a single dataset.

Abstract: *Chlamydomonas reinhardtii* is a unicellular green alga that has been widely used as a plant reference system for six decades, it has a quick generation time (~ 6h), can be synchronized and grown to high densities and its three genomes are sequenced and well-annotated. We have utilized Chlamydomonas as a reference organism to understand the principles underlying trace metal utilization and economy in a photosynthetic cell, and have identified a repertoire of assimilatory and distributive transporters, discovered mechanisms for reducing the metal quota and recycling metal cofactors from non-essential to essential proteins in situations of sustained elemental deficiency.

We used a multi-variant RNAseq approach to identify novel genes in nitrogen metabolism. A total of 22 different conditions were analyzed in triplicate, allowing for 300 individual comparisons, including 69 comparisons where only a single variable is changing. Among the utilized perturbations where variations of the carbon source (acetate vs CO_2), nitrogen source (NH_4, NO_3, urea), light intensity (100 vs 500 PAR), growth regime (16/8h day/night cycle vs continuously grown cultures) and culture density (10^5 vs 10^6 cells/ml). Variation of the growth regime was the most influential parameter altering gene expression, followed by the carbon source and cell density, while the variation of the nitrogen source and light intensity only affected a select group of transcripts. Consistently, growth regime and carbon source were responsible for the first two principal components in the dataset. To maximize discovery and improve confidence we used a consensus strategy that utilized multiple algorithms for differential gene expression analysis. Known genes involved in the individual comparisons allowed to verify the validity of the approach. In total we identified ~650 novel genes in the dataset, with various degrees of annotation, that we were able to categorize for their regulation to one of the studied perturbations.

This material is based upon work supported by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research program under Award Number DE-FC02-02ER63421. The work conducted at the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.