Crassulacean Acid Metabolism (CAM) Informatics Tools

Karen A. Schlauch[1*] (schlauch@unr.edu), Robert C. Moseley[2,3], Juli Petereit[4], David Weston[2], Xiaohan Yang[2], and **John C. Cushman[1]**.

[1]Department of Biochemistry and Molecular Biology, University of Nevada, Reno, NV; [2]Biosciences Division, Oak Ridge National Laboratories, Oak Ridge, TN; [3]Bredesen Center, University of Tennessee, Knoxville, TN; [4]Department of Biomedical Engineering, University of Nevada, Reno, NV.

http://cambiodesign.org/

Project Goals: The long-term goal of this research is to enhance the water-use efficiency (WUE) and adaptability to hotter, drier climates of species that normally perform C_3 photosynthesis by introducing crassulacean acid metabolism (CAM). Four major objectives will be pursued to enhance photosynthetic performance and WUE in *Arabidopsis* and *Populus*: 1) Define the genetic basis of key CAM modules in eudicot and monocot CAM species *via* co-expression network models; 2) characterize the regulation of 'carboxylation', 'decarboxylation', and 'inverse stomatal control' modules of CAM using comparative genomics, network models, and loss-of-function testing; 3) deploy advanced genome engineering technologies to enable stacking of a large number of transgenes into a single genomic locus to improve transgene persistence and transfer fully functional 'carboxylation' and 'decarboxylation' modules from CAM species to C_3 species that can accommodate overnight malic acid storage in the vacuole; and 4) analyze the effects of these transgenic modules on 'stomatal control', CO_2 assimilation and transpiration rates, biomass yield, and WUE in *Arabidopsis* and *Populus*.

CAM is a specialized mode of photosynthesis that provides increased plant water-use efficiency (WUE). This adaptation has prompted research to examine the genetic and metabolic networks governing CAM physiology as part of a biodesign strategy to produce bioenergy feedstocks with enhanced WUE. To assist this effort, two informatics tools were developed: (1) **petal** is a co-expression network modeling and analysis program offering novel analytical approaches for CAM-specific whole-omics data; (2) **CAMbase** is an integrative resource platform where tools are deployed for intuitive analysis and visualization of genomic data.

petal is a whole-omics co-expression network analysis algorithm that offers two approaches not currently provided by most other available methods. Firstly, as most high-throughput expression data are not normally distributed, network models generated by **petal** can be based on either parametric and non-parametric similarity measures to provide more statistically sound network construction. Secondly, as most biological networks are known to have approximate scale-free and small-world structures, **petal** generates network models following these two properties. By allowing the network construction to be governed by these

properties, our networks are guaranteed to be both scale-free and small-world without user intervention (e.g., specifying a threshold/cut-off value). Connectivity structures in the co-expression network models such as cliques, modules, clusters, and paths represent gene or protein interactions. Cliques are of particular interest, as their structure represents more robust connectivity based upon their unique mathematical properties. **petal** offers downstream network analysis by identifying and examining the cliques in the networks, which provide putative functional groupings within the network. The petal system is implemented in **R** and easily processes whole-omics data.

CAMbase is an integrative platform in which tools are deployed in an infrastructure for intuitive analysis and visualization of genomic data. Currently, the database includes high-resolution, time-course RNA-seq data for *Kalanchoe laxiflora*, *Arabidopsis thaliana*, and *Agave americana*, as well as reference genomes for *Kalanchoe laxiflora*, orchid, and pineapple. Additionally, CAMbase includes gene ontology and homologs identified from other plant genomes for available CAM plants. Jbrowse and Apollo are available to explore genome features and annotation updating, respectively, of available CAM genomes. CAMbase also includes an eFP browser for comparative analysis of omics time-course data. Current progress of CAMbase development will be presented with a focus on the implementation of tools (e.g., eFP browser, Jbrowse, Apollo) for newly sequenced CAM model species. CAMbase data are being formatted in a KBase compliant manner, with the goal of pushing novel RNA-seq and genome data into the DOE-Kbase cyberinfrastructure for downstream modeling applications related to feedstock biodesign.

References

1. Borland AM, Hartwell J, Weston D, Schlauch KA, Tschaplinski TJ, Tuskan GA, Yang X, Cushman JC. (2014) Engineering crassulacean acid metabolism to improve water-use efficiency. Trends Plant Science 19: 327-338.
2. Petereit J, Harris, FC Jr, and Schlauch K, **petal**: A Novel Co-Expression Network Modeling System, Proceedings of the 2015 IEEE International Conference on Bioinformatics and Biomedicine (IEEE BIBM 2015) pp 234-241, November 9-12, 2015, Washington, DC.