112. Transcriptome sequencing and RNA-seq mRNA expression profiling in the facultative CAM model species *Mesembryanthemum crystallinum*.

Won C. Yim[1], Bernard W.M. Wone[1], Richard L. Tillett[1], Bahay G. Bilgi[1], Rebecca L. Albion[1], Karen A. Schlauch[1], HengfuYin[2], Gerald A. Tuskan[2], Xiaohan Yang[2] and **John C. Cushman[1]*** (jcushman@unr.edu).

[1]Department of Biochemistry and Molecular Biology, University of Nevada, Reno, NV;
[2]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN.

http://cambiodesign.org

Project goals: The 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). Two major objectives will be pursued to achieve this goal: 1) Define the genetic basis of key CAM modules in the facultative CAM species *via* network modeling using 'omics (transcriptome and metabolome) technologies; 2) characterize the transcriptional regulation of 'carboxylation', 'decarboxylation', and 'inverse stomatal control' modules of CAM. The use of the facultative CAM model species, the common ice plant (*Mesembryanthemum crystallinum*), facilitates the identification of genes recruited into CAM function and likely those needed to successfully engineer CAM into C_3 host species.

The salt-tolerant succulent common or crystalline ice plant (*Mesembryanthemum crystallinum*; Caryophylalles, Aizoaceae), can switch from C_3 photosynthesis to CAM under saline or water- deficit stress conditions (Cushman et al., 2008). To help define the transcriptome of this facultative CAM model, next-generation sequencing was performed. Roche 454 pyrosequencing of RNA from salt- and drought-stressed leaves, roots, mature flowers, seed pods, and seeds produced >4 M sequences. Illumina HiSeq sequencing of well-watered and water-deficit stress treated leaf tissues generated an additional 241.8 Gbp of sequence. Reads from each platform were assembled separately, using newbler for 454 reads, and Trinity for Illumina reads. The assemblies were combined using GICL for a final assembly of 31,238 contigs, with a mean length of 1,601 bp, and N50 value of 2,278 bp. Of these, 28,576 contigs passed contamination screening. 22,971 contigs showed sequence homology to known plant genes in the Phytozome database. 15,751 contigs had complete ORF. 37,427 terms were assigned in our contigs with putative Gene Ontology *via* protein motifs identified by InterProScan. 804 potential transcription factors were classified. Ice plant contigs with significant homology were found for >90% the Greencut2, eUCO, and AVPO reference sets of conserved genes. Broader gene family clustering identified 9,188 families of orthologous genes in OrthoMCL-DB into which 21,343 ice plant contigs clustered. 7,275 of these families included orthologs identified in all angiosperms at OrthoMCL. This assembled reference transcriptome will facilitate RNA-seq gene expression analysis of C_3/CAM photosynthesis and circadian regulation and will serve as an essential resource for genome annotation for in-progress genome sequencing.

In order to explore the circadian clock-regulated mechanisms that control the expression of CAM in the common ice plant, Illumina-based RNA-seq digital gene expression profiling was performed on well-watered and water-deficit stress treated leaf tissue. Samples were collected in parallel every 4 h over a 72 h time course under both 24 h light/dark (diel entrainment) and 48 h light/light (zeitgeber or free running) conditions in order to capture the full repertoire of circadian clock-controlled transcriptional outputs. cDNA libraries were constructed for 114 ice plant samples and Illumina HiSeq2000 was performed using a total of 19 flow-cell channels, with 6 samples in each. Approximately 2,241,015,571 bp of raw data were

generated yielding a total of 2,415,294,564 trimmed sequence reads. All trimmed, cleansed reads were assembled *de novo* using the Trinity program. The complete read dataset assembled into 109,902 contigs, which ranged in size from 201 bp to 15,797 bp with a mean contig length of 1228 bp. RSEM and Bowtie were then used to assign reads to multiple genes and count relative transcript abundances.

Read counts were mapped, normalized and differentially expressed genes were detected using the DESeq package. Circadian clock output differences between the C_3 photosynthesis and CAM states will be discussed in detail. These results will facilitate the identification of the molecular genetic machinery required for CAM using comparative genomics by comparing gene expression patterns of known CAM components across C_3, C_4, and CAM species. These results will also aid in future ice plant genome annotation efforts.

References

Cushman JC, Tillett RL, Wood JA, Branco JA, Schlauch KA (2008) Large-scale mRNA expression profiling in the common ice plant, *Mesembryanthemum crystallinum,* performing C_3 photosynthesis and Crassulacean acid metabolism (CAM). J. Exp. Bot. 59:1875-1894.

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Genomic Science Program under Award Number DE-SC0008834. Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under Contract Number DE–AC05–00OR22725.