Computational Biology to Target Plant Pathways and Experimental Data

Ying Xu[1,2]* (xyn@bmb.uga.edu), Eberhard Voit,[1,2] Yanbin Yin,[1,2] Wen-Chi Chou,[1,2] Shan Wang,[1,2] Xizeng Mao,[1,2] Michael Udvardi,[1,2] Maor Bar-Peled,[1,2] Jerry Tuskan,[2] Brian Davison,[2] and **Paul Gilna (BESC PI)**[2]

[1]University of Georgia, Athens, Georgia; [2]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee

http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) is focused on the fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves 1) designing plant cell walls for rapid deconstruction and 2) developing multitalented microbes for converting plant biomass into biofuels in a single step (consolidated bioprocessing). Addressing the roadblock of biomass recalcitrance will require a greater understanding of plant cell walls from synthesis to deconstruction. This understanding would generate models, theories and finally processes that will be used to understand and overcome biomass recalcitrance. This grand challenge calls for an integrated research approach, which is illustrated by the project described below. BESC research is multidisciplinary by design and multi-institutional in composition.

The computational biology team of BESC is working with experimental biologists on a number of collaborative projects to study plant cell wall recalcitrance. Four examples are highlighted.

To facilitate gene discovery and genomic research in switchgrass the assembly of ~11 million switchgrass, ESTs were sequenced by 454 technologies and ~700k Sanger ESTs into a “Switchgrass gene index.” The data set contains 139,200 assembled UniGenes, based on which an Affymetrix cDNA chip was designed and manufactured. A systematic annotation of the “gene index” set was performed by searching against CDD, Panther and CAZy databases.

A novel mathematical modeling approach was developed to investigate the lignin biosynthesis pathways by analyzing different omics datasets in wild type and transgenic plants. The modeling effort led to two novel postulates regarding the control of the lignin biosynthetic pathway, which were partially validated by subsequent laboratory experiments. (Lee, et al., 2011)

A systematic bi-clustering co-expression analysis using *Arabidopsis* microarray data (with >1,300 experimental conditions) and ~800 annotated cell wall-related (CWR) proteins was completed. This analysis identified ~2,000 new proteins co-expressed with annotated CWR genes, forming 217 co-expression modules. Conserved regulatory motifs were also predicted for genes from the same co-expression modules. Based on this result, three BESC partners have suggested over 50 genes to be verified within by our coordinated plant genetic transformation.

The dbCAN database and webserver (http://csbl.bmb.uga.edu/dbCAN/index.php) is now automated

for CAZy annotation. Member proteins and literature of each CAZy family are analyzed and manually curated to extract a signature domain to represent the family. A hidden markov model (HMM) was build based on the alignment of all the signature domains in the member proteins for the HMM to specifically represent the corresponding CAZy family. With the HMMs, dbCAN provides the annotation service so users can upload their proteins for signature domain-based CAZy annotation.

References:

Lee, Y., et al., PLoS Computational Biology, 7(5), 2011.

[The BioEnergy Science Center is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.]