Pleiotropy Decomposition of 609 *Populus Trichocarpa* Genotypes

Deborah Weighill[1,2], Carissa Bleker[1,2], Priya Ranjan[1], Nan Zhao[1], Madhavi Martin[1], Gerald Tuskan[1], Wellington Muchero[1], Tim Tschaplinski[1], Daniel Jacobson[1,2*] (jacobsonda@ornl.gov), and **Paul Gilna**[1]

[1]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee; [2]Bredesen Center for Interdisciplinary Research and Graduate Education, University of Tennessee, Knoxville

http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) focuses on fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves (1) improved plant cell walls for rapid deconstruction and (2) multi-talented microbes for converting plant biomass into biofuels in a single step [consolidated bioprocessing (CBP)]. Biomass research works with two potential bioenergy crops (switchgrass and *Populus)* to develop improved varieties and to understand cell wall biosynthesis pathways. We test large numbers of natural variants and generate specific modified plants samples. BESC's research in deconstruction and conversion targets CBP manipulating thermophilic anaerobes and their cellulolytic enzymes for improved conversion, yields, and titer. Enabling technologies in biomass characterization, 'omics, and modeling are used to understand chemical and structural changes within biomass and to provide insights into mechanisms.

Pleiotropy is the phenomenon in which a gene affects multiple phenotypes [1]. One can also have a SNP-centric view of pleiotropy as a single SNP affecting multiple phenotypes [2]. While pleiotropy used to be considered an exception to the rules of Mendelian genetics, it has since been proposed to be a common, central property inherent to biological systems [1]. Pleiotropic patterns can be detected in the results of Genome Wide Association Studies (GWASs) as SNPs within genes having multiple significant phenotypic associations. Two main pleiotropic patterns exist within GWAS results. Firstly, Type 1 pleiotropy occurs when a single SNP within a gene is associated with more than one phenotype [2]. Type 2 pleiotropy occurs when two different SNPs within a single gene have different phenotype associations [2].

In this study, we present a method called Pleiotropy Decomposition for the investigation of pleiotropic patterns from GWAS analysis of *Populus trichocarpa*. The method aims to distinguish between different pleiotropic patterns (Type 1 vs Type 2) while also providing intuitive network representations for the exploration of these pleiotropic patterns in the GWAS results.

GWAS analysis using EMMAX [3] was performed on a carefully constructed, non-related population of ~1,000 *Populus trichocarpa* genotypes and phenotype information in the form of untargeted metabolomic profiles for 609 of these genotypes, followed by False Discovery Rate correction. From the resulting set of SNP-phenotype associations, we constructed a gene-phenotype GWAS network (matrix) consisting of genes connected to their associated phenotypes. Our method involved the decomposition of the gene-phenotype network to two bipartite networks through the use

of pleiotropic modules as an intermediate, latent variable. These networks together unravel the underlying pleiotropic structure of genes.

The main intermediate step in pleiotropy decomposition involved the construction of pleiotropic modules, namely, groups of SNPs that are associated with the same sets of phenotypes. This was performed by constructing a GWAS profile for each SNP as a vector of its phenotypic associations, calculating the Proportional Similarity index between the GWAS profiles of all pairs of SNPs, and clustering using MCL [4]. Pleiotropy Decomposition thus involves the separation of the gene-phenotype matrix into a gene-module matrix and a module-phenotype matrix, which can be visualized as a gene-module bipartite network and a module-phenotype bipartite network, respectively. The gene-module matrix gives information regarding the type of pleiotropy exhibited by genes, (pleiotropic structure of genes), and the module-phenotype gives information regarding the specific phenotype associations of the pleiotropic modules that genes are comprised of.

The use of these decomposition matrices as bipartite networks provides a convenient structure for the exploration of GWAS results, while also unraveling the different pleiotropic signatures of genes. These networks allowed for interesting patterns of enrichment to be found within the set of pleiotropic genes, and highlighted the pleiotropic interactions which link functions such as plant defense responses and lignin biosynthesis. Identification of potential pleiotropic genes, as well as the functions those pleiotropic genes might affect should prove to be a useful tool in planning targeted experiments involving genetic modification, as it could give an idea as to the potential other side effects the modification/deletion of a gene of interest might cause.

References

1. Tyler, Anna L., et al. "Shadows of complexity: what biological networks reveal about epistasis and pleiotropy." *Bioessays* 31.2 (2009): 220-227.
2. Solovieff, Nadia, et al. "Pleiotropy in complex traits: challenges and strategies." *Nature Reviews Genetics* 14.7 (2013): 483-495.
3. Kang, Hyun Min, et al. "Variance component model to account for sample structure in genome-wide association studies." *Nature genetics* 42.4 (2010): 348-354.
4. Van Dongen, Stijn. "Graph clustering via a discrete uncoupling process." *SIAM Journal on Matrix Analysis and Applications* 30.1 (2008): 121-141.

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.