SNPeffect: Identifying Functional Roles for SNPs using Metabolic Networks

Debolina Sarkar[1]* (debolina@psu.edu), Jin Zhang[2], Kai Feng[2], David Kainer[2], Timothy Tschaplinski[2], Dan Jacobson[2], Wellington Muchero[2], Costas D. Maranas[1] and **Gerald A. Tuskan[2]**

[1]Department of Chemical Engineering, Pennsylvania State University, University Park; and [2]Center for Bioenergy Innovation, Oak Ridge National Laboratory, Oak Ridge, Tennessee

https://cbi.ornl.gov

Project Goals: The Center for Bioenergy Innovation (CBI) vision is *to accelerate domestication of bioenergy-relevant, non-model plants and microbes to enable high-impact innovations at multiple points in the bioenergy supply chain.* CBI addresses strategic barriers to the current bioeconomy in the areas of 1) high-yielding, robust feedstocks, 2) lower capital and processing costs via consolidated bioprocessing (CBP) to specialty biofuels, and 3) methods to create valuable byproducts from the lignin. CBI will identify and utilize key plant genes for growth, composition and sustainability phenotypes as a means of achieving lower feedstock costs, focusing on poplar and switchgrass. We will convert these feedstocks to specialty biofuels (C4 alcohols, C6 esters and hydrocarbons) using CBP at high rates, titers and yield in combination with cotreatment, pretreatment or catalytic upgrading. CBI will maximize product value by *in planta* modifications and biological funneling of lignin to value-added chemicals.

Genetic sources of phenotypic variation have been a focus of plant studies aimed at improving agricultural yield and understanding adaptive processes. Genome-wide association studies (GWAS) identify the genetic background behind a trait by examining associations between phenotypes and single-nucleotide polymorphisms (SNPs). Although such studies are common, biological interpretation of the results remains a challenge; especially due to the confounding nature of GWAS population structure and the systematic biases thus introduced.

Here, we propose a complementary analysis ('**SNPeffect**') that offers putative genotype-to-phenotype mechanistic interpretations by integrating biochemical knowledge encoded in metabolic models. SNPeffect was used to explain differential growth rate and metabolite accumulation in 914 *P. trichocarpa* accessions as the outcome of SNPs in enzyme-coding genes. To this end, we also constructed a genome-scale metabolic model for *Populus trichocarpa*, the first for a perennial woody tree and further extracted a leaf-specific subnetwork using transcriptomics data collected from 390 poplar trees. As expected, our results indicate that growth is a complex polygenic trait governed by carbon and energy partitioning. The predicted set of functional SNPs are associated with experimentally-characterized growth-determining genes and also suggest functions for putative genes. Functional SNPs were found in pathways such as amino-acid metabolism, nucleotide biosynthesis, and cellulose and lignin biosynthesis, in line with breeding strategies that target pathways governing carbon and energy partition. A follow-up optimization procedure 'SNP-rank' was used to rank order the functional SNP set to provide specific targets for functional characterization and genomic selection priors. This work contributes to the understanding of the molecular mechanisms governing growth and plant productivity, in order to further exploit them and develop a sustainable bioenergy feedstock.