

Rapid domestication of poplar using genomic selection and machine learning

David Kainer^{1*} (kainerm@ornl.gov), Jonathon Romero¹, Hari Chhetri^{1,2}, David Macaya-Sanz^{1,2}, Stephen DiFazio^{1,2}, Daniel Jacobson¹, and **Gerald A. Tuskan¹**

¹Center for Bioenergy Innovation, Oak Ridge National Laboratory, TN; ²Center for Bioenergy Innovation, West Virginia University, WV

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.

Populus trichocarpa is an undomesticated woody species with high potential as a bioenergy feedstock, however, its long breeding cycle slows our ability to develop improved varieties. To rapidly improve biomass yield per hectare we need to optimize gains per generation. Sustainable and uniform yield is a highly complex conglomerate of traits, so genetically modifying or selecting for a few key loci is unlikely to achieve our goals. Genomic Selection¹ however, models the genetic (breeding) value of genome-wide variants on target traits in a training population. Parents from a breeding population are then selected solely on their aggregate genetic value for those traits. Furthermore, the same model can be used to predict progeny phenotypes well before they reach maturity, allowing early culling and rapid cycle times based on the genotype.

To this end, we are implementing an annual genomic selection to select high performing parents based on a multi-trait genomic best linear unbiased prediction (GBLUP) index that captures the goals of improved biofuel feedstocks. As a compliment to GBLUP, the CBI computational biology group is applying explainable machine learning models in order to incorporate non-linear relationships between genetic variants (epistasis) and between genetics and environment (GxE) into predictive models. These predictions are being utilized and tested in a cross of poplar parentals. Additionally, we have developed a computational pipeline that generates virtual recombinant progeny from a given cross, thus allowing a machine-learning prediction of progeny performance as a result of prospective parental selections and crosses.

References:

1. Hayes, B. J., and M. E. Goddard. "Prediction of total genetic value using genome-wide dense marker maps." *Genetics* 157.4 (2001): 1819-1829.

The Center for Bioenergy Innovation 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.