

## Exploiting Nature's Knockouts: How the Pangenome Reveals New Insights into Poplar Phenotypic Variation

Doug Hyatt<sup>1</sup>, David Kainer<sup>1\*</sup> ([kainerd@ornl.gov](mailto:kainerd@ornl.gov)), Piet Jones<sup>2</sup>, Manesh Shah<sup>1</sup>, Jared Streich<sup>1</sup>, Wellington Muchero<sup>1</sup>, Jay Chen<sup>1</sup>, Tim Tschaplinski<sup>1</sup>, Posy Busby<sup>3</sup>, Daniel Jacobson<sup>1</sup> and **Gerald A. Tuskan<sup>1</sup>**

<sup>1</sup>Center for Bioenergy Innovation, Oak Ridge National Laboratory, Oak Ridge, TN; <sup>2</sup>University of Tennessee, Knoxville; <sup>3</sup>Oregon State University, Corvallis

<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 will address 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 and C6 esters) using CBP at high rates, titers and yield in combination with cotreatment or pretreatment. CBI will maximize product value by *in planta* modifications and biological funneling of lignin to value-added chemicals.**

Most GWAS are performed using SNPs obtained from reads aligned to a reference genome of the same species, or a closely related species. However, it is becoming clear that a reference genome assembled from one individual may not adequately represent the genomic feature space of the species. A large number of genes and other genomic elements may be present or absent in any given individual, including the original reference individual. The overall species-wide genetic space is known as the Pangenome, containing core genes found in most or all individuals, and distributed genes that are found in fewer individuals and may be rare. Distributed genes that are absent in the original reference individual (e.g. Nisqually-1) may explain some of the 'missing' heritability of important quantitative traits, provide new sources of introgression material for breeding, and new explanatory power for understanding population structure and the evolution of the species. We can view the presence and absence of these genes within a population as a naturally evolved assay of knockouts.

We have assembled the pangenome of *Populus trichocarpa* using DNA resequencing and RNAseq data from almost 1000 genotypes, resulting in a new marker dataset representing the presence or absence variation of putative distributed genes across the population. Here we present results of association between these markers and phenotypes assayed in the same population.

*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.*