Recombination and Segregation Genomic Patterns in *Populus* to Facilitate Genomic Selection

Chanaka Roshan Abeyratne[1], David Macaya-Sanz[1]* (david.macayasanz@mail.wvu.edu),
Gancho Slavov[2], Lee Gunter[3], Kathleen Haiby[4], Richard Shuren[4], Jay Chen[3], Daniel Jacobson[3],
Brian Stanton[4], Stephen DiFazio[1], and **Gerald A. Tuskan[3]**

[1]West Virginia University, Morgantown; [2]Rothamsted Research, Harpenden, UK; [3]Center for
Bioenergy Innovation, Oak Ridge National Laboratory, Oak Ridge, TN; and [4]GreenWood
Resources Inc., Portland, OR

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.**

Black cottonwood (*Populus trichocarpa*) is a pioneer tree species identified as a promising
renewable feedstock for bioenergy and bioproducts. *Populus* can be used to study many aspects
of perennial development related to phenology, wood formation, vegetative propagation, and
dioecy that cannot be studied using conventional plant model systems such as *Arabidopsis*.
Breeding techniques to identify elite parent trees face considerable challenges: lengthy sexual
reproduction cycles (of at least 4 years), difficulty to design Recombinant Inbred Lines (RIL),
highly heterozygous genomes, and poor juvenile-mature correlation for many traits, necessitating
expensive long-term field trials. Genomic selection incorporates genomic information into
breeding strategies to alleviate these issues.

Prediction models can be improved by detailed understanding of how recombination and
segregation vary across the genome, among individuals and between sexes. Linkage information
also allows the creation of genetic maps that can be used to identify Quantitative Trait Loci
(QTL) for phenotypes of breeding relevance (height, diameter, bud set, and disease resistance).
Those QTL can also be used to improve prediction model training.

Here, we have re-sequenced the genomes of 49 families (*N* = 821 offspring), corresponding to a
full factorial cross of seven females and seven males. Using benchmark software GATK4, we
have called biallelic SNPs. Using the pedigree information, we have revealed patterns of
segregation in the genome and selected high-quality markers under Mendelian segregation. We
have phased and imputed the progeny genomes and recovered the gametic haplotypes of the

fourteen parents, allowing us to estimate fine-scale genomic patterns of recombination. By means of the R package Onemap, fourteen genetic maps have been created, and specific patterns of variation have been revealed.