135. A Systems Biology, Whole-Genome Association Analysis of the Molecular Regulation of Biomass Growth and Composition in *Populus deltoides*

Annette M. Fahrenkrog,[1] Leandro G. Neves,[1] Juan J. Acosta,[2] William "Brad" Barbazuk,[1,3,4]
Kirst M[1,2,3]* (mkirst@ufl.edu)

[1]Plant Molecular and Cellular Biology Program, University of Florida, Gainesville; [2]School of Forest Resources and Conservation, University of Florida, Gainesville; [3]Genetics Institute, University of Florida, Gainesville; [4]Department of Biology, University of Florida, Gainesville

Project Goals: Poplars trees are well suited for biofuel production due to their fast growing habit, favorable wood composition and adaptation to a broad range of environments. The availability of a reference genome sequence, ease of vegetative propagation and availability of transformation methods also make poplar an ideal model for the study of wood formation and biomass growth in woody, perennial plants. The objective of this project is to conduct a genome-wide association genetics study to identify genes that regulate bioenergy traits in *Populus deltoides* (eastern cottonwood). Association mapping is being pursued by combining sequence-capture followed by high-throughput sequencing to genotype coding and regulatory sequences in the whole-genome of *P. deltoides*. To identify genetic polymorphisms that regulate biomass productivity and carbon partitioning we are pursuing the following goals: (1) optimizing sequence-capture for unbiased, high- throughput and low cost recovery of target coding and regulatory sequences in P. deltoides; (2) carrying out "whole-genome" genotyping of a *P. deltoides* unstructured population for association mapping; and (3) identifying significant SNP-trait associations with biomass growth and carbon partitioning to define genes and alleles that regulate trait variation.

To develop an optimal platform for sequence capture in *Populus* we initially measured the efficiency of ultra-long oligonucleotides in retrieving specific sequences of the genome of *P. deltoides* for detection of SNP polymorphisms. Oligonucleotides that tile regulatory and coding sequences of all poplar genes previously identified as expressed in the main vegetative tissues were designed. Next, genomic DNA from three *P. deltoides* individuals previously sequenced by DOE's Joint Genome Institute, were hybridized to oligonucleotides in solution, captured and sequenced. Captured fragments were aligned to the reference sequence, and the degree of sequence enrichment and the power to detect known SNPs was evaluated. Oligonucleotide probes that were effective for capture and genotyping included exons and parts of the 5' and 3' untranslated regions (UTRs) of 18,153 genes.

Using the sequence-capture oligonucleotides developed previously, we analyzed a *P. deltoides* association population composed of 579 unrelated individuals. This is a subset of a larger population composed of 815 unrelated genotypes collected from 13 states in Central, South and Eastern US, covering 35 river systems. In total, we identified approximately one million single nucleotide polymorphism (SNP) markers in the whole population, distributed along the genes captured. In addition, we also captured and genotyped regions distributed every 15 Kbp in the genome, to provide a genome-wide view of the genetic diversity of the species, to be used for population genetic studies. In parallel we propagated the association population and established it in greenhouse and field test sites for biomass and lignocellulosics composition measurements.

Currently the field trials are completing one year of age. Wood samples collected from the greenhouse trials have been collected and phenotyped using pyrolysis MBMS. Detailed analysis of these phenotypic measurements is currently in progress.

With the near completion of the experiments planned for the first two goals of this project, we are now initiating the last aim. Specifically, the analysis will begin with the identification of any family and population structure in the population. Analysis of SNP-trait associations that account for covariance due to relatedness among individuals will be carried out using analysis of variance, followed by a combined analysis of the most highly associated SNPs in one single Bayesian model to estimate joint, epistatic effects of multiple loci.