WHOLE EXOME GENOTYPING IN LOBLOLLY PINE IDENTIFIED 2.8 MILLION SNPS USED FOR ASSOCIATION ANALYSES

*Mengmeng Lu[1], **Carol Loopstra[2]**, Konstantin Krutovsky[2,3,4,5], C. Dana Nelson[6], Thomas Byram[7], Tomasz Koralewski[2], and Candace Seeve[8]*

[1]University of Calgary, Calgary, Alberta, Canada
[2]Texas A&M University, College Station, TX, USA
[3]Georg-August-University of Göttingen, Göttingen, Germany
[4]N.I. Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, Russia
[5]Genome Research and Education Center, Siberian Federal University, Krasnoyarsk, Russia
[6]USDA Forest Service, Southern Research Station, Saucier, MS, USA
[7]Texas A&M Forest Service, 2585 TAMU, College Station, TX, USA
[8]USDA-ARS, Columbia, MO, USA

Until recently, the number of markers available for genomic studies of loblolly pine (*Pinus taeda* L.) was limited. A much larger number is needed to cover the enormous genome in order to facilitate the use of genomics to accelerate breeding. Sequencing of the loblolly pine genome combined with improved methods for whole exome capture and sequencing allowed us to identify and genotype much larger numbers of single nucleotide polymorphisms (SNPs) in 375 trees from the ADEPT2 population. Exome sequence capture oligonucleotide probes (Roche NimbleGene) were designed using 199,723 exons in 48,391 high quality tentative genes. The probes covered 46,206,684 bp of the target regions. Trees were multiplexed for sequencing and between 25.25 and 60.55 million sequence reads were obtained per tree. The reads were mapped to loblolly genome assembly v 1.01. When the SNPs were filtered using the criteria of at least 10X sequencing depth in at least 90 % of the individuals and a minor allele frequency ≥0.05, 972,720 SNPs were acquired for downstream analyses. When the criteria were relaxed to include sequences with at least 5X depth in 100 % of trees with a MAF ≥ 0.01, 2,822,609 SNPs were identified.

The SNP markers were used to test for single locus associations, SNP-SNP interactions and correlation of individual heterozygosity with phenotypic traits. A total of 36 SNP-trait associations were found after false discovery rate (FDR) correction for specific leaf area (5 SNPs), branch angle (2), crown width (3), stem diameter (4), total height (9), carbon isotope discrimination (4), nitrogen concentration (2), and pitch canker resistance traits (7). Eleven SNP-SNP interactions were found to be associated with branch angle (1), crown width (2), total height (2), carbon isotope discrimination (2), nitrogen concentration (1), and pitch canker resistance (3). Non-additive effects imposed by dominance and epistasis account for a large fraction of the genetic variance for the quantitative traits. Genes that contain the identified SNPs have a wide spectrum of functions. Individual heterozygosity positively correlated with water use efficiency and nitrogen concentration. Association analyses done using also expression data for 198 genes with putative roles in wood development or response to drought or disease identified 2,581 significant associations. Analyses done using concentrations of 82 metabolites identified 536 associations with 53 metabolite concentrations. In conclusion, multiple effects identified in this study influence the performance of loblolly pines. This study has provided resources for understanding the genetic control of complex traits that have potential value for assisting with breeding through marker assisted and genomic selection.

Contact Information: Carol Loopstra, Dept. of Ecosystem Science and Management, TAMU 2138, College Station, USA TX 77843-2138, Phone: 979-862-2200, Email: c-loopstra@tamu.edu