PROGRESS ON PINE SNP DISCOVERY AND SNP ARRAY DESIGN

Fikret Isik[1], Juan Acosta[1], Trevor Walker[1], Andrew Eckert[3], Richard Sniezko[4], and Jill Wegrzyn[2]
[1]North Carolina State University, Raleigh, NC, USA
[2]University of Connecticut, Storrs, CT, USA
[3]Virginia Commonwealth University, Richmond, VA, USA
[4]USDA Forest Service, Dorena Genetic Resource Center, OR, USA

Genomic selection has the potential to revolutionize pine breeding. Our efforts funded by the USDA-NIFA aim to bring genomics to pine breeding, with the goal of greatly improving genetic gain per unit time. Our group is coordinating SNP discovery efforts to bring genetic resources together for multiple pine species and use consistent bioinformatics tools for selection of informative SNPs. The genomic resources will be organized in the TreeGenes database for public access. These genetic resources (informative SNP markers, annotation, and an improved loblolly pine genome assembly) will stimulate a new wave of novel research endeavors in pine breeding and conifer genomics. During the first year of our four-year project, we have improved the genome assembly of loblolly pine by aligning deep transcriptomic sequence (Illumina short reads and PacBio Iso-Seq) for scaffolding purposes. This resulted in a total of 11,951 linked scaffolds to create 4,545 new super scaffolds. This improved assembly (v2.01) has been released to the PineRefSeq project page and TreeGenes database. It reports an N50 of 111 Kbp across a total of 1.76 million scaffolds (~1 million of 1Kbp and longer). This release serves as the primary reference for an updated gene (29,213 models) and repeat (84%) annotation. The final gene selection totaled 29,213 full-length high-quality gene models with an additional 14,045 partial models. Genomic resources generated from the PineMAP project are currently being aligned against v2.01 and SNPs are being assessed via an in-house pipeline, which leverages Freebayes, Varscan2, and applicable filtering. Preliminary SNP calls will be validated based on sequence and mapping qualities and will be used to describe genome-wide patterns of diversity and linkage. Over the next two years of the project we hope to genotype large number of samples for genomic predictions and other research applications.

Contact information: Fikret Isik, Department of Forestry and Environmental Resources, Campus Box 8002, Raleigh, NC 27695. Phone: 919-449-6891, Email: fisik@ncsu.edu