

## Sequencing and Mapping the Gene Space of the Southern Pines

Craig Echt<sup>1</sup>, Richard Cronn<sup>2</sup>, Nurul Islam-Faridi<sup>1</sup>, Konstantin V. Krutovsky<sup>3</sup>, Chun Liang<sup>4</sup>, Carol A. Loopstra<sup>3</sup>, C. Dana Nelson<sup>1</sup>, Daniel G. Peterson<sup>5</sup>

Genetically improved southern pines are planted on over 600,000 hectares each year and are common or predominant species on over 25 million hectares of non-planted forest land. Their ability to efficiently convert CO<sub>2</sub> into biomass and their widespread use in tree plantations have also made them cost-effective feedstocks for lignocellulosic ethanol production and promising tools in efforts to curb greenhouse gas levels via carbon sequestration. The demand for wood and fiber from the southern pines is projected to increase 35-40% within the next 50 years to accommodate the needs of a growing populace and to offset decreased tree harvesting in western and northern states. However, the number of acres reserved for tree growth in the southern states is likely to decline due to increasing population growth and urbanization, forest health threats, and climate change impacts. Consequently the U.S. forest products industry is looking for ways to significantly increase wood yields per acre, and forest managers are needing to genetically manage forest restoration efforts in changing environments. Both pursuits are greatly limited by a paucity of information on southern pine gene space diversity.

By *de novo* DNA sequencing ~ 5000 gene-enriched BAC clones and deep resequencing ~20,000 expressed gene regions, we plan to achieve comprehensive characterization of the gene space diversity in a reference loblolly pine genome and in four major southern pine species: loblolly pine, slash pine, longleaf pine, and shortleaf pine. We will achieve this through high-throughput massively parallel sequencing of genes from the loblolly pine reference genotype 7-56 and from population samples of 100 unrelated individuals in each species. DNA sequence diversity within and among the species will be characterized and publicly distributed. We will construct a high resolution, high density reference gene space map of loblolly 7-56 which will be integrated with BAC and FISH based physical maps. Comparative gene space maps for the four southern pines will be constructed. It is our intention that public distribution of the data will lead to new cost-effective SNP genotyping assays that will be used for genome mapping, molecular breeding and genetic improvement of carbon sequestration traits and biomass feedstock traits, genetic management of breeding populations, and conservation of native forest populations. The extensive population sequence data for the four species will also provide invaluable comparative population genomics information about nucleotide diversity, population structures, phylogeographic and phylogenetic relationships.

---

<sup>1</sup> U.S. Forest Service, Southern Institute of Forest Genetics, Harrison Experimental Forest, 23332 Old Mississippi 67, Saucier, MS 39574.

<sup>2</sup> U.S. Forest Service, Dept. of Botany and Plant Pathology, Oregon State University, Corvallis, Oregon 97331-2902

<sup>3</sup> Department of Ecosystem Science and Management, Texas A&M University, College Station, TX 77843

<sup>4</sup> Department of Botany, Miami University, Oxford, OH 45056

<sup>5</sup> Department of Plant and Soil Sciences, Mississippi State University, Mississippi State, MS 39762