

The Integration of Resequencing, Genotyping and Population Structure Data Reveals Patterns of Adaptive Genetic Diversity in Loblolly Pine (*Pinus taeda L.*)

Andrew J. Eckert¹, Jill L. Wegrzyn¹, Jennifer M. Lee¹, John D. Liechty¹, Kristian Stevens², Kathleen D. Jermstad³, C. Dana Nelson³, Santiago C. Gonzalez-Martinez⁴, Charles H. Langley¹, David B. Neale¹

Patterns of nucleotide diversity and divergence were investigated across a set of 6,200 genes distributed throughout the functional gene space of loblolly pine (*Pinus taeda L.*). In total, we discovered 21,000 SNPs located within these 6,200 genes (mean = 4 SNPs/gene \pm 4 SNPs/gene). Several thousand of these genes were integrated into existing linkage maps using SNPs genotyped within two three-generation outbred loblolly pine pedigrees. Of the 6,200 alignments, 4,400 and 1,109 had high quality orthologous sequence data from single samples of *Pinus radiata* D. Don and *P. lambertiana* Dougl., respectively. Distributions of diversity and divergence indices were similar to those expected for a conifer, with those metrics varying several-fold in magnitude across different classes of genes. Joint patterns of diversity and divergence revealed strong signals of adaptive evolution, with several instances indicative of positive selection being identified across multiple linkage groups. Not fully explored in those analyses, however, are the roles of population structure and demography in shaping the observed patterns of genetic diversity. Paleoclimate, microfossil, and metabolite data indicate that loblolly pine was restricted to two southern refugia located in present day Florida and Mexico during the last glacial maximum. Multilocus genotypes at 23 nuclear microsatellite markers were consistent with expansion from two Pleistocene refugia followed by division into six genetic clusters. Using those clusters along with genotype data for 3,080 single nucleotide polymorphisms (SNPs), we examined the distribution of extreme levels of population differentiation, as measured by F_{ST} , across the functional gene space of loblolly pine. These results are integrated with those based on population genetic analysis of the resequencing data. By doing so, we highlight several robust examples of positive selection, as well as illustrate the need to further understand large-scale patterns of linkage disequilibrium across the genome of loblolly pine.

¹ University of California - Davis

² Section of Evolution and Ecology

³ USDA National Forest Service

⁴ CIFOR-INIA