Exome genotyping and association genetics of adaptation and stress mitigation traits in loblolly pine (*Pinus taeda* L.)

Mengmeng Lu[1, 2], Konstantin Krutovsky[1, 3], C. Dana Nelson[4], Jason West[1, 2], Tomasz Koralewski[1], Tom Byram[1, 5], Carol Loopstra[1, 2]

[1]Department of Ecosystem Science and Management, Texas A&M University, College Station, TX; [2]Molecular and Environmental Plant Sciences Program, Texas A&M University, College Station, TX; [3]Department of Forest Genetics and Forest Tree Breeding, Georg-August-University of Göttingen, Göttingen, Germany; [4]USDA Forest Service, Southern Research Station, Southern Institute of Forest Genetics, Saucier, MS; [5]Texas A&M Forest Service, College Station, TX

Abstract

Loblolly pine, *Pinus taeda* L., is the most widely planted and commercially important tree species in the southeastern U.S. To increase the number of known single nucleotide polymorphisms (SNPs) and functional markers available for research and tree breeding, we used genotyping by sequencing for targeted exome regions. The exons were captured in a population of 375 trees using NimbleGen capture probes and then sequenced using the Illumina HiSeq 2500 v4 platform. Oligonucleotide probes were designed for 199,723 exons (≈49 Mbp) partitioned from the loblolly pine reference genome (PineRefSeq v1.01) representing 48,391 high quality genes. Bioinformatics analyses demonstrated that the probes covered 90.2% of the target regions. Capture efficiency analyses showed that an average of 67.2% of the reads from each tree could be mapped to the capture target regions and more than 70% of the captured target bases had at least 10X sequencing depth. A total of 972,720 SNPs were acquired after filtering. We found there was a high genetic diversity and fast linkage disequilibrium decay within this population. Two distinct clusters representing western and eastern parts of the loblolly pine range were shown from the structure analysis. Association and epistasis analyses were conducted for the traits related to growth, crown structure, physiology and disease resistance using over two million SNPs, which were from re-filtering the raw SNPs using a relaxed condition. Forty-four significant SNP-trait associations and eleven significant SNP-SNP interactions were identified. Our results demonstrated the efficiency of exome capture for genotyping a species with a large and complex genome such as loblolly pine. The significant associated markers can be used to accelerate the selection of better trees.

Results

Table 1 Categorization by genomic location for 972,720 SNPs

Location	Proportion
Exon	58 %
Coding regions	53 %
5’ UTR	2 %
3’ UTR	3 %
Intron	13 %
Unclassified	29 %

Table 2 Genetic correlations (r^2) among the traits investigated

	SLA	BA	CW	2014H	DIA	CID	N
SLA		NS	NS	NS	NS	NS	0.022
BA			-0.010	NS	NS	0.024	NS
CW				0.573	0.591	-0.034	0.013
2014H					0.803	-0.018	0.012
DIA						NS	NS
CID							-0.063
N							

SLA: specific leaf area; BA: branch angle; CW: crown width; 2014H: Total height at the year of 2014; DIA: diameter; CID: carbon isotope discrimination; N: nitrogen concentration

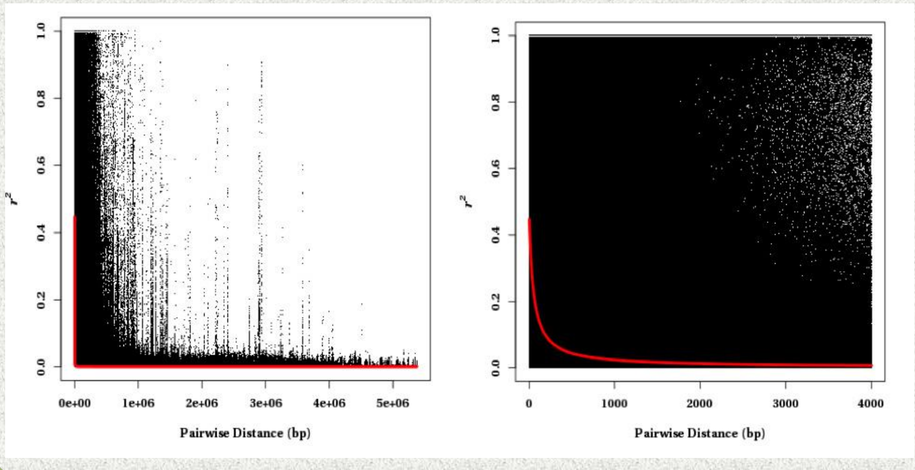

Figure 2 Linkage disequilibrium measure r^2 plotted against the physical distance between all pairs of single nucleotide polymorphism (SNP) markers from the same scaffold (left) and between all pairs of SNP markers from the same scaffold located within 4000 bp (right) calculated for all 375 trees. The trend lines of the nonlinear regression of r^2 against physical distance between the SNP markers were indicated in red.

Materials

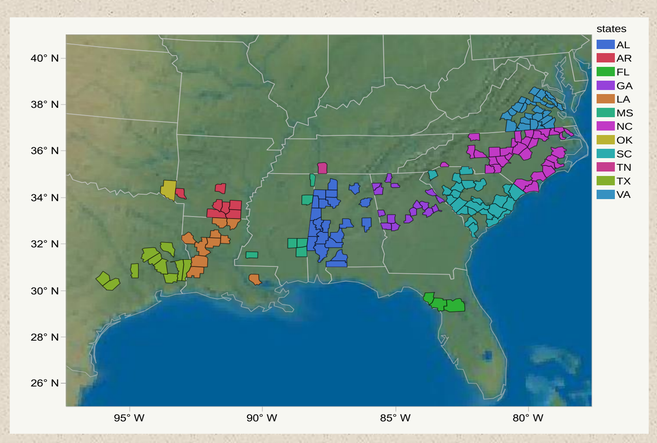

Figure 1 The counties of origin of the maternal trees colored by states. This population was established in Saucier, Mississippi during the spring of 2010.

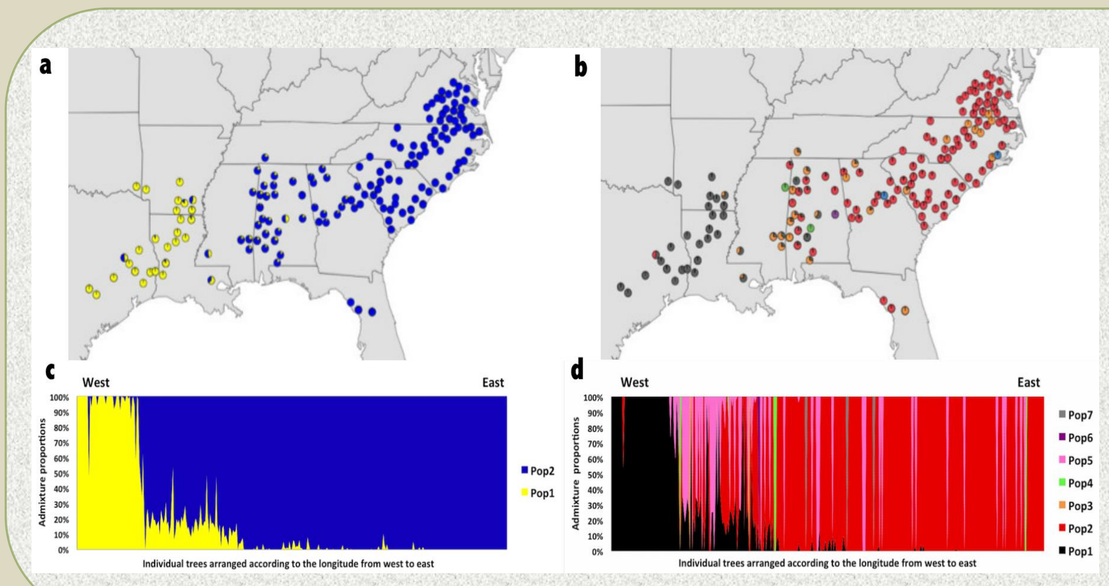

Figure 3 Illustrated are admixture proportions of each tree when *K*=2 (a, c) and *K*=7 (b, d) (K is the number of potential clusters or subpopulations). Each pie chart was divided by the population assignments inferred by fastStructure (a, b). The trees were aligned on the x-axis according to the longitude from west to east (c, d).

Methods

- Genotype the 375 trees using NimbleGen SeqCap EZ system with capture probes designed from 48,391 high quality tentative genes in Gene Annotation v.2.0 for loblolly pine genome assembly v. 1.01.
- Phenotype the traits, including growth traits such as total height and diameter; crown structure traits such as specific leaf area, crown width and branch angle; physiology and resistance traits such as carbon isotope discrimination, nitrogen concentration and pitch canker disease resistance (disease data is available from Quesada et al., 2011).
- Conduct association mapping using TASSEL 5.0 and epistasis analyses using PLINK 1.9 to link the genetic variation and traits.

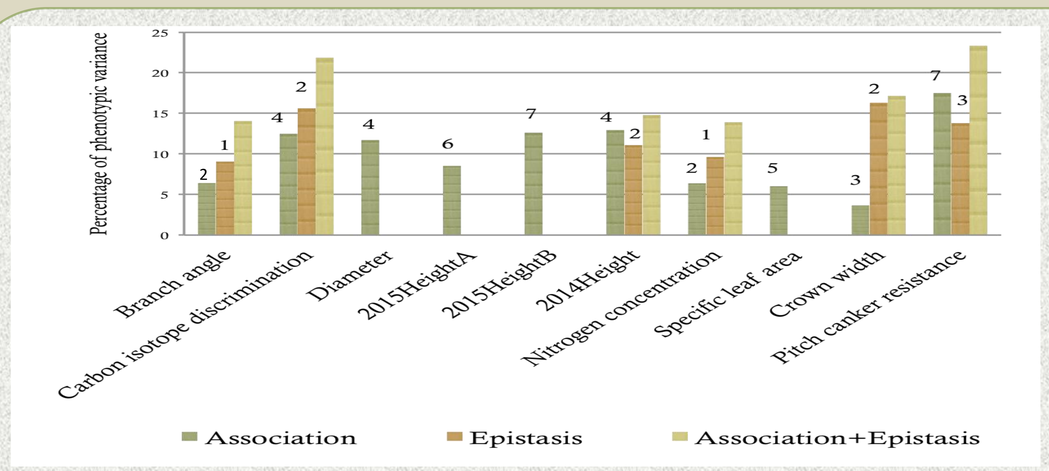

Figure 4 Percentage of phenotypic variance accounted for by SNPs identified by association and epistasis analyses. The number of significant SNPs detected for each trait were labeled on top of the bars.

Traits	Gene functions for potential related genes
Crown structure	Auxin-mediated development; Meristem and organ development; Bark storage proteins
Growth	Lignan biosynthesis; Stress response; Embryonic development
Physiology and disease resistance	Stress response; Transporter; Auxin-mediated development

Conclusion

Our results demonstrate the efficiency of exome capture for genotyping a species with a large and complex genome. The highly diverse genetic variation reported in this study provides a valuable resource for loblolly pine breeding through marker-assisted selection and genomic selection.

Acknowledgement

We thank Dr. Jill Wegrzyn and PineRefSeq Project team for providing the draft loblolly pine reference sequences and exon annotation, Dr. John Davis and Dr. Tania Quesada for providing the pitch canker disease resistance data, Texas A&M Institute for Genome Sciences and Society (TIGSS) for providing computational resources and system administration support for the TIGSS HPC Cluster, and Texas A&M University Genomics and Bioinformatics Services for providing sequencing service. Special gratitude goes to Dr. Vikram Chhatre, Jeffery Puryear and technicians Gay Flurry, Chance Parker, Chuck Burdine of the Harrison Experimental Forest at Southern Institute of Forest Genetics for the sample collection and grinding.

The study is funded by the Pine Integrated Network: Education, Mitigation, and Adaptation project (PINEMAP), a Coordinated Agricultural Project funded by the USDA National Institute of Food and Agriculture, Award #2011-68002-30185.