



Association analyses of quantitative traits and environmental variables using 2.8 million exome-wide SNPs in a clonally tested loblolly pine population (*Pinus taeda* L.)

Carol Loopstra, Mengmeng Lu, Konstantin Krutovsky

Pine Plantation Research and Decision Support Tool Rollout

May 16-17, 2017 Athens, GA



United States
Department of
Agriculture

National Institute
of Food and
Agriculture



1175

Genome genotyping and association genetics of environmental adaptation and stress mitigation in a clonally tested loblolly pine (*Pinus taeda* L.) population

Abstract: The loblolly pine (*Pinus taeda* L.) is a major commercial pine species in the southeastern United States. It is a clonally propagated species, and its genetic diversity is limited. This study aims to identify genetic markers associated with environmental adaptation and stress mitigation in a clonally tested loblolly pine population. We performed genome-wide association studies (GWAS) using a high-density SNP array. The results show that several SNPs are significantly associated with growth and stress tolerance traits. These findings provide valuable insights into the genetic architecture of these traits and can be used for marker-assisted selection in loblolly pine breeding programs.



1174

Construction of a high-density genetic linkage map using Axiom genotyping and identification of QTLs for wood property, growth and male flower formation in Japanese cedar (*Cryptomeria japonica*)

Abstract: Japanese cedar (*Cryptomeria japonica*) is a widely distributed tree species in East Asia. It is an important timber species and has a long history of cultivation. This study aims to construct a high-density genetic linkage map using Axiom genotyping and identify quantitative trait loci (QTLs) for wood property, growth, and male flower formation. We performed a genome-wide association study (GWAS) using a high-density SNP array. The results show that several SNPs are significantly associated with these traits. These findings provide valuable insights into the genetic architecture of these traits and can be used for marker-assisted selection in Japanese cedar breeding programs.





Exome genotyping, linkage disequilibrium and population structure in loblolly pine (*Pinus taeda* L.)

Molecular methods can help us identify DNA markers that can assist with identification of genes and alleles that contribute to traits of interest such as growth, resistance to insects and diseases and adaptation to the environment.





Exome genotyping, linkage disequilibrium and population structure in loblolly pine (*Pinus taeda* L.)

As methodologies have improved we have expanded the work in forest trees from cloning of single genes and QTL mapping to sequencing of transcriptomes and whole genomes, association genetics and genomic selection.

Giant genomes felled by DNA sequencing advances

Complete genetic blueprints collected for several conifer species

BY PUNEET KOLLIPARA 11:38AM, MAY 22, 2013

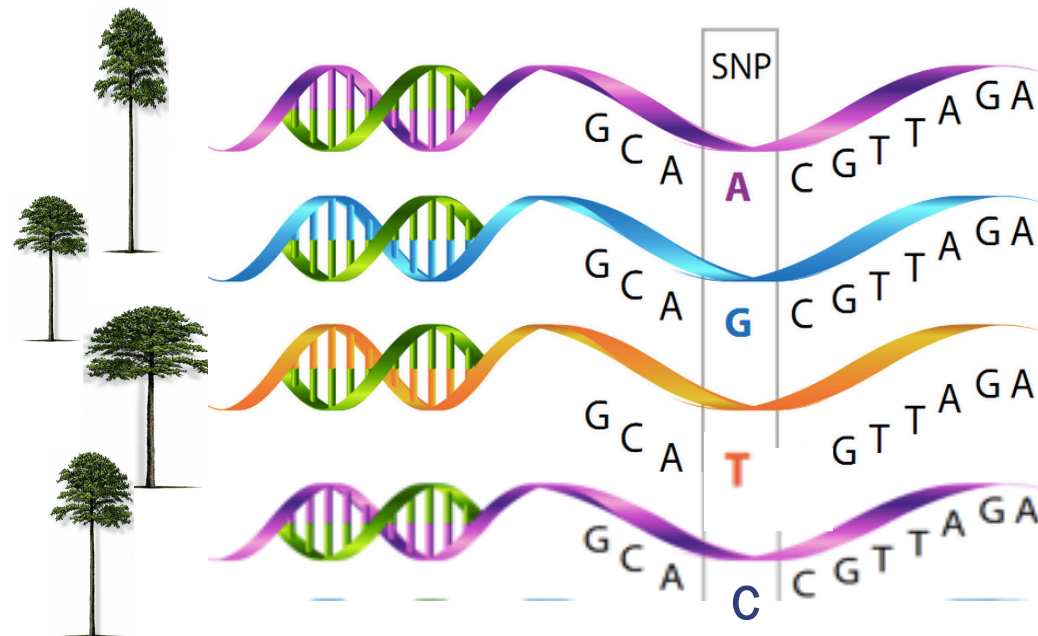


The genetic sequence of the Norway spruce (shown) could shed light on the evolution of the seed-bearing plants called gymnosperms.

Exome genotyping, linkage disequilibrium and population structure in loblolly pine (*Pinus taeda* L.)

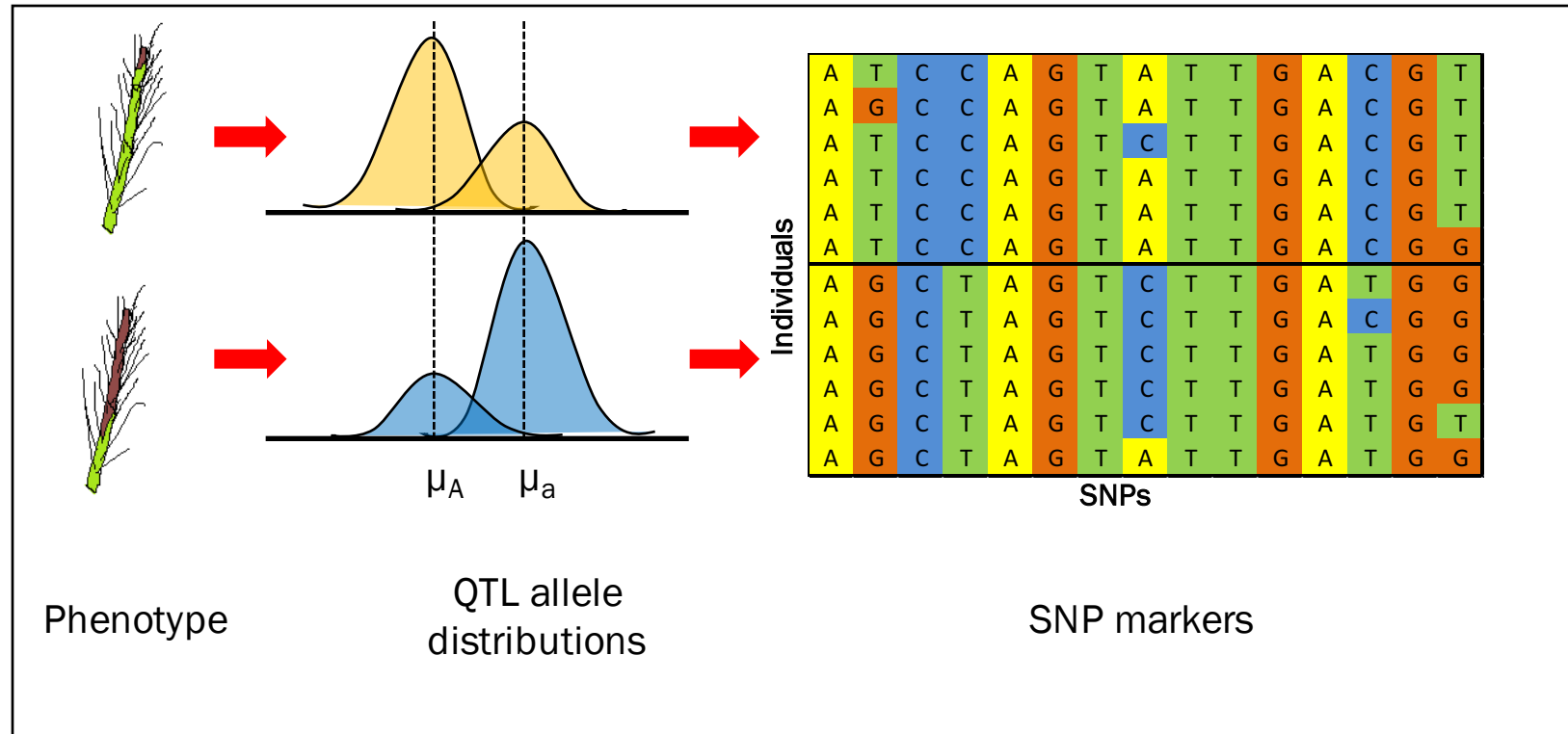
Molecular Markers are required for molecular breeding methods including QTL mapping, association genetics, and genome selection.

Single Nucleotide Polymorphisms (SNPs) are the most common type of variation found in the genome and provide abundant markers.



Using SNP genotypes for association mapping

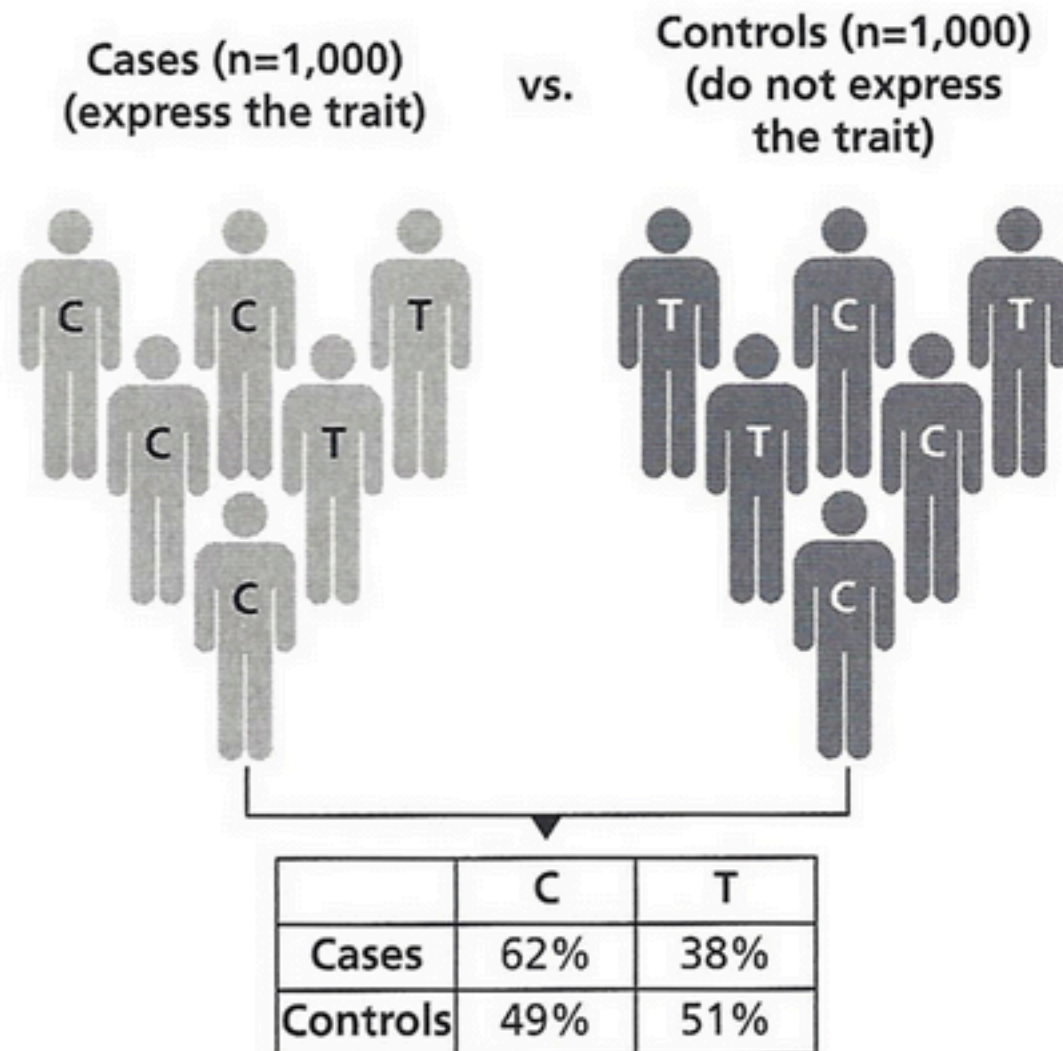
What is association mapping?



Adapted from Rafalski, 2010 *Curr Opin Plant Biol* 13: 174-180 and Jannink et al., 2001 *Trends Plant Sci* 6: 337-342

Molecular markers associated with a trait will show a bias in their allelic frequency with respect to the phenotype.

Case-control study for genetic association





To identify the genes best associated with traits of interest, we need large numbers of molecular markers.

Until recently, the numbers were insufficient.

Most association studies done as part of the ADEPT2 project had less than 1 SNP per 10 genes!

Project	Funding	Genotyping Platform	N SNPs	SNP location
ADEPT 1	NSF		43	Known genes
ADEPT 2	NSF	Illumina Infinium	7,216	ESTs
PINEMAP	USDA-NIFA	Capture-Seq	67,387	Genome-wide

CCLONES genotypic data

Exome genotyping, linkage disequilibrium and population structure in loblolly pine (*Pinus taeda* L.)

Genetic variation including SNPs can be discovered and genotyped by sequencing genomes or portions of genomes.

However – Conifer genomes are too large to sequence in large populations needed for association genetics or genomic selection.



The genome of the sugar pine, the so-called "king of conifers," is the largest ever sequenced for any organism.
(Credit: dotpolka/Flickr)

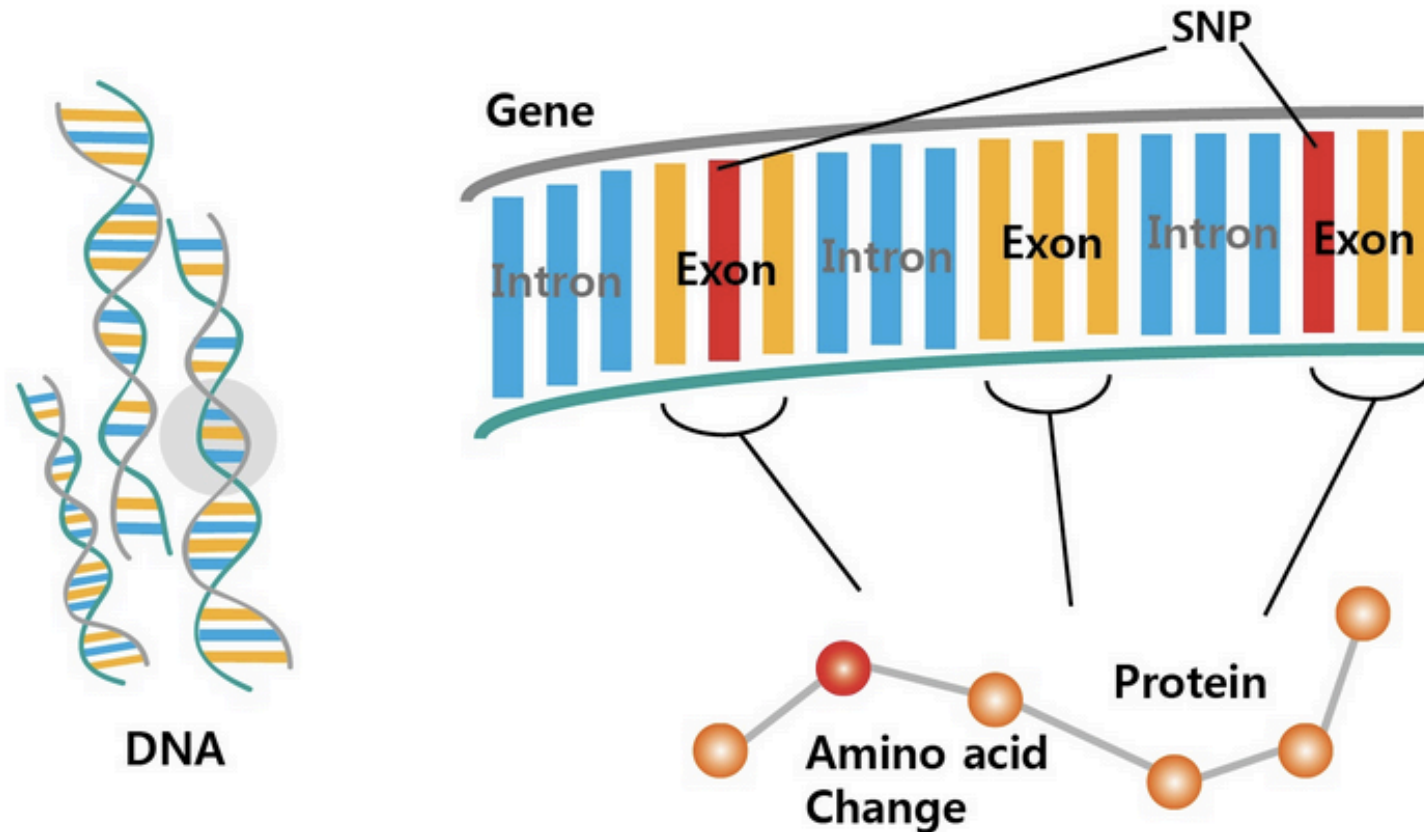
FORESTS | UNIVERSITY OF CALIFORNIA, DAVIS

Sugar pine genome is 10X the size of human genome

Posted by Pat Bailey-UC Davis | December 18th, 2015

Exome genotyping, linkage disequilibrium and population structure in loblolly pine (*Pinus taeda* L.)

Exome Capture – allows us to capture primarily coding regions from large numbers of trees and to sequence only those regions.





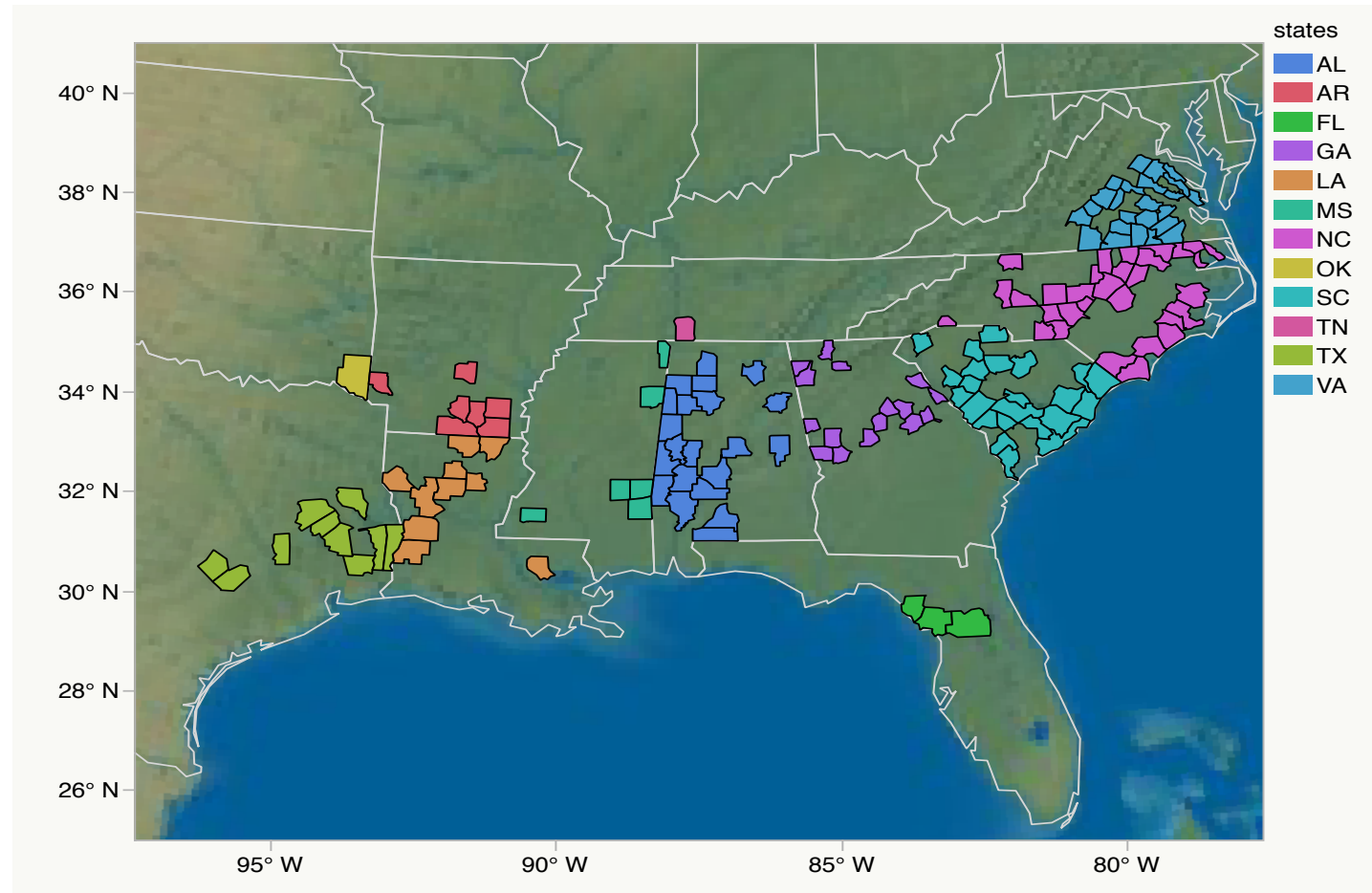
Outline

- **Use exome capture and sequencing to identify single nucleotide polymorphisms (SNPs) in loblolly pine and genotype the ADEPT2 population**
- Associate the genotyped SNPs with growth and adaptive traits to discover candidate genes
 - Dissecting the genetic basis of cellular phenotypes by association mapping and network construction
- Associate the genotyped SNPs with environmental variables and detect signals of selection which are related to local adaptation



Exome genotyping, linkage disequilibrium and population structure in loblolly pine (*Pinus taeda* L.)

The counties of origin of the maternal trees colored by state



Exome genotyping, linkage disequilibrium and population structure in loblolly pine (*Pinus taeda* L.)

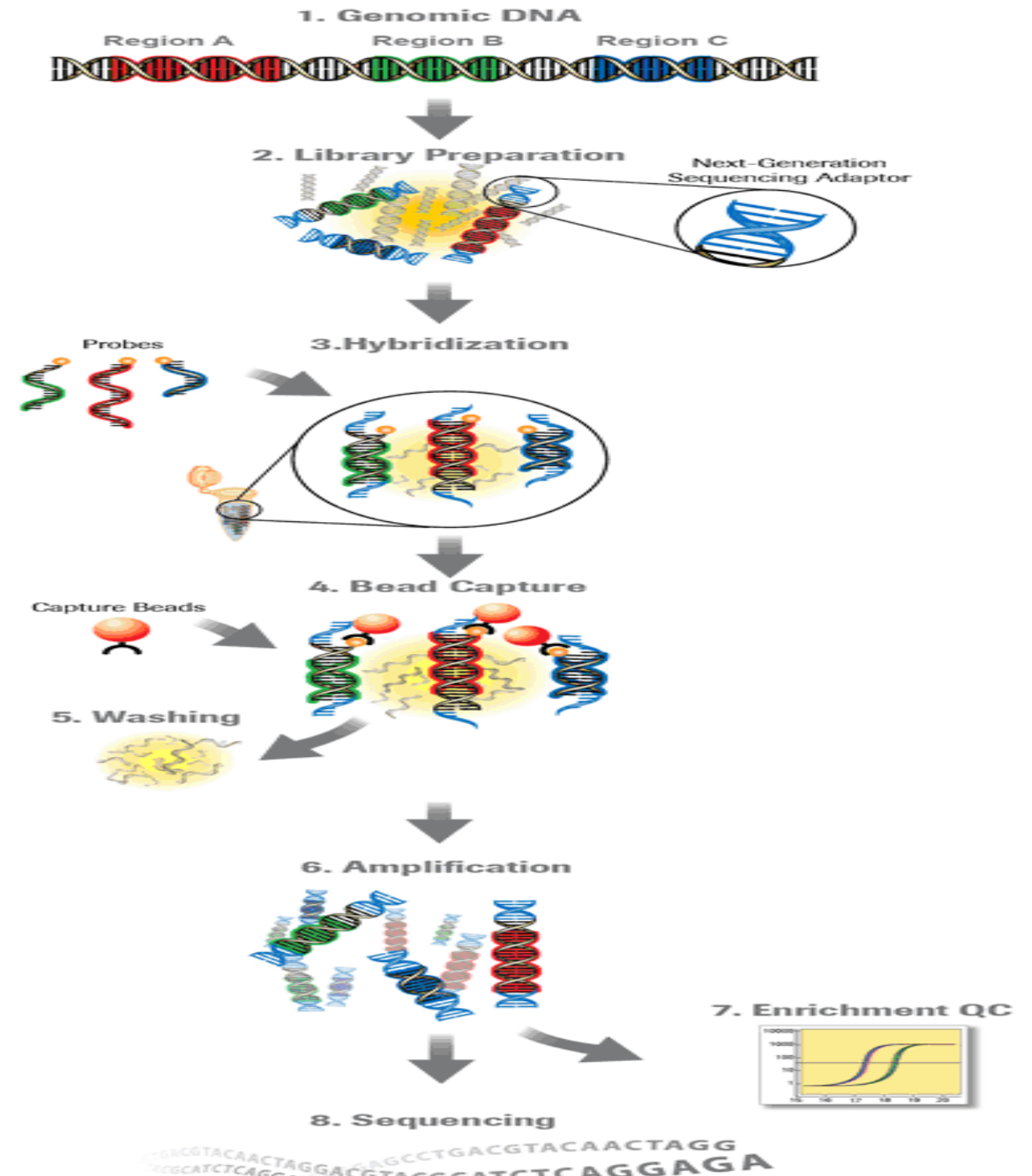
Plant material



384 × 3 unrelated trees were planted

Workflow of the NimbleGen SeqCap EZ system

<http://www.nimblegen.com/products/seqcap/ez/choice/index.html>





Exome genotyping, linkage disequilibrium and population structure in loblolly pine (*Pinus taeda* L.)

Capture probes

	Nimblegen Probes
Design source	199,723 exon regions (\approx 49 Mbp) in 48,391 high quality tentative genes
Total probes number	\approx 2.1 million
Target Region Size mapped on reference assembly v1.01	\approx 46 Mbp
Feature	55~105 bp DNA single strands



Exome genotyping, linkage disequilibrium and population structure in loblolly pine (*Pinus taeda* L.)

Sequencing output

- 10 uniquely indexed samples were multiplexed for capture and sequencing
- \approx 250 million reads were output from each of 38 lanes.

Sequenced at Texas A&M University Genomics and Bioinformatics Service with 2×125 bp paired-end format using Illumina HiSeq 2500 V4.



SNP discovery

Bi-allelic sites

10X sequencing depth in 90 % of the individuals

MAF \geq 0.05

972,720 SNPs

Bi-allelic sites

5X sequencing depth for all individuals without missing data

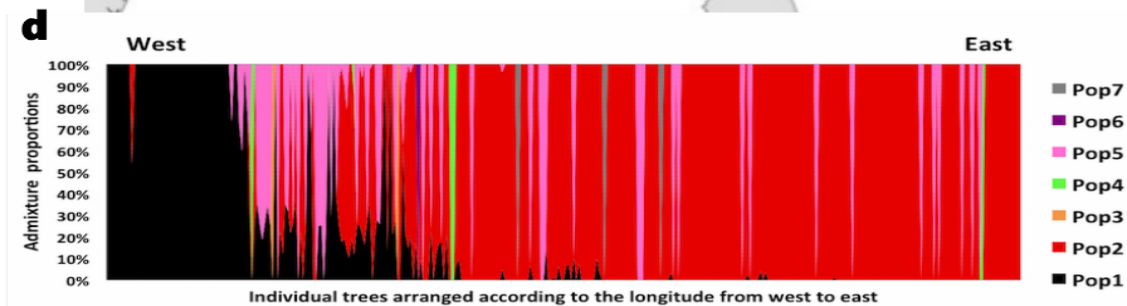
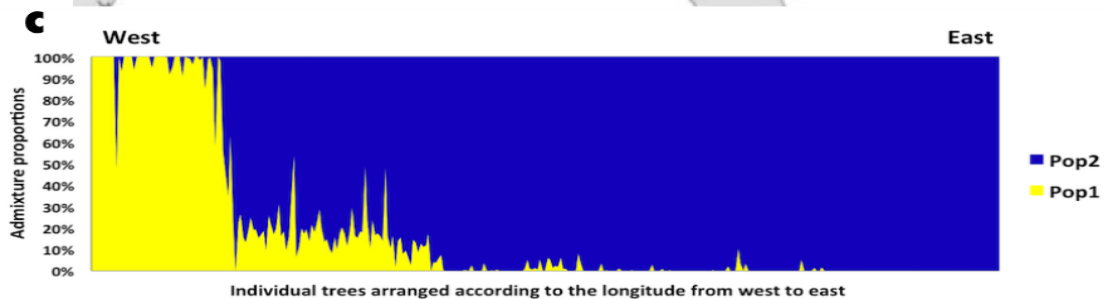
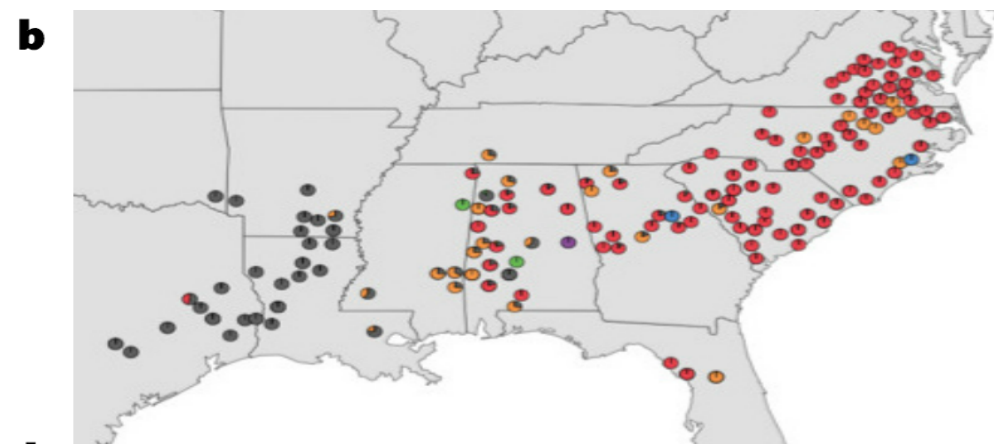
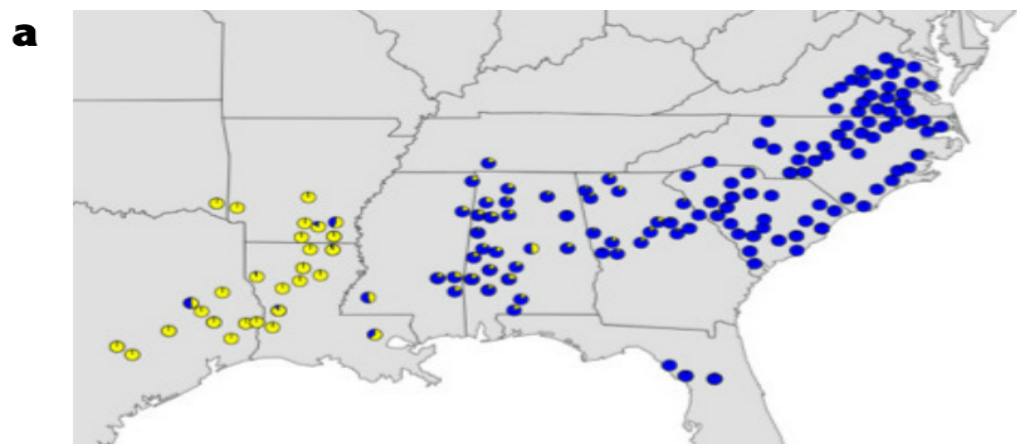
MAF \geq 0.01

2,822,609 SNPs



Genetic structure analysis

$K = 2$ and $K = 7$ were chosen to explain the population structure





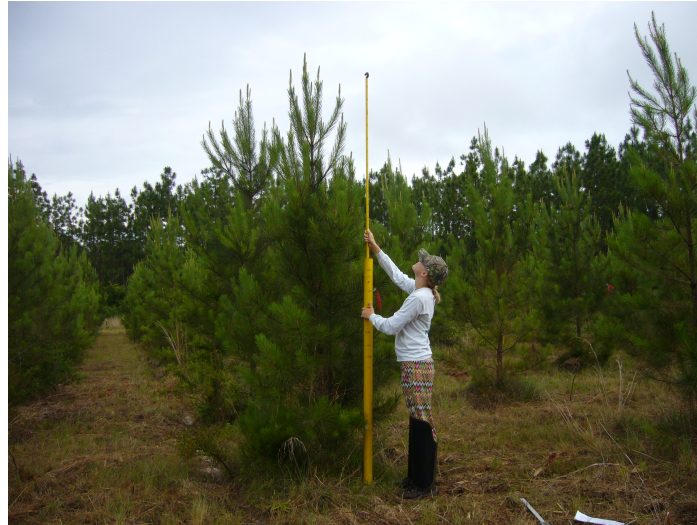
Outline

- Use exome capture and sequencing to identify single nucleotide polymorphisms (SNPs) in loblolly pine and genotype the ADEPT2 population
- **Associate the genotyped SNPs with growth and adaptive traits to discover the candidate genes**
- Dissecting the genetic basis of cellular phenotypes for by association mapping and network construction
- Associate the genotyped SNPs with environmental variables and detect signals of selection which are related to local adaptation

Association genetics of adaptive and growth traits

Phenotyping

PINEMAP Undergraduate Fellowship Program





Association genetics of adaptive and growth traits

Phenotyping

Growth traits:

- Total height (2014, 2015 Before, 2015 After)
- Diameter (DIA)

Crown structure traits:

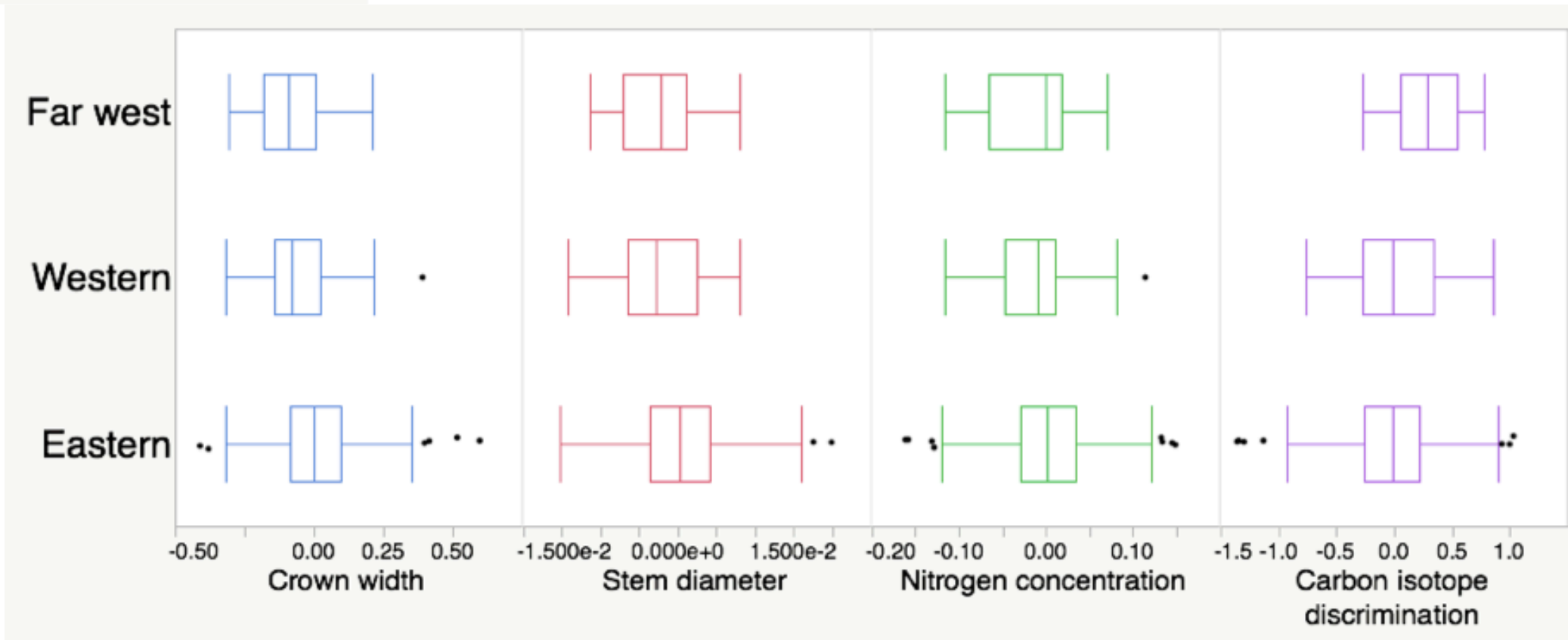
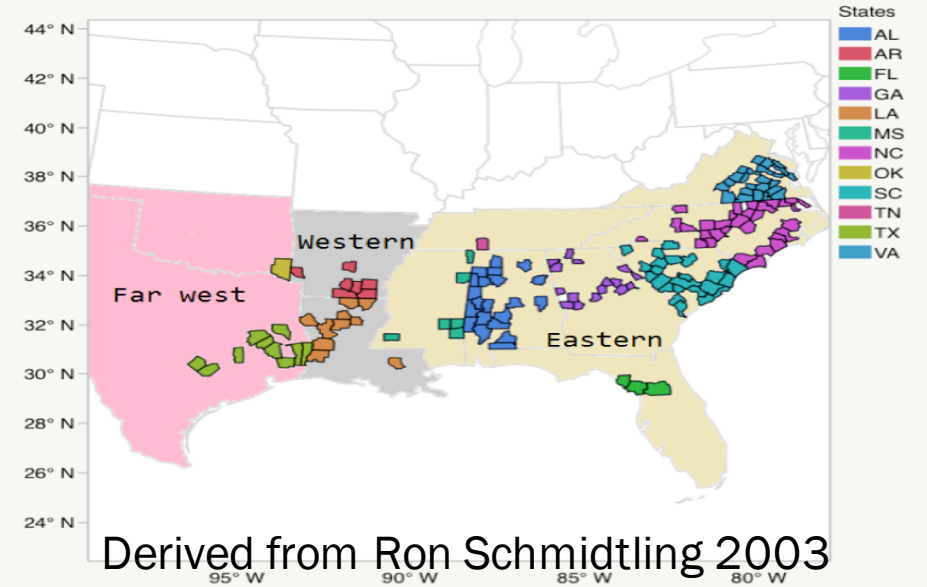
- Specific leaf area (SLA)
- Crown width (CW)
- Branch angle (BA)

Physiology and resistance traits:

- Carbon isotope discrimination ($\Delta^{13}\text{C}$)
- Nitrogen concentration (N)
- Pitch canker disease resistance (Quesada et al., 2011)

The range was divided into 3 regions

Geographical variation among 4 traits





Association genetics of adaptive and growth traits

Marker-Trait Association Analyses

Traits	SNP numbers
Specific leaf area	5
Branch angle	2
Crown width	3
Stem diameter	4
Total height	9
Carbon isotope discrimination	4
Nitrogen concentration	2
Pitch canker resistance	7

Association genetics of adaptive and growth traits

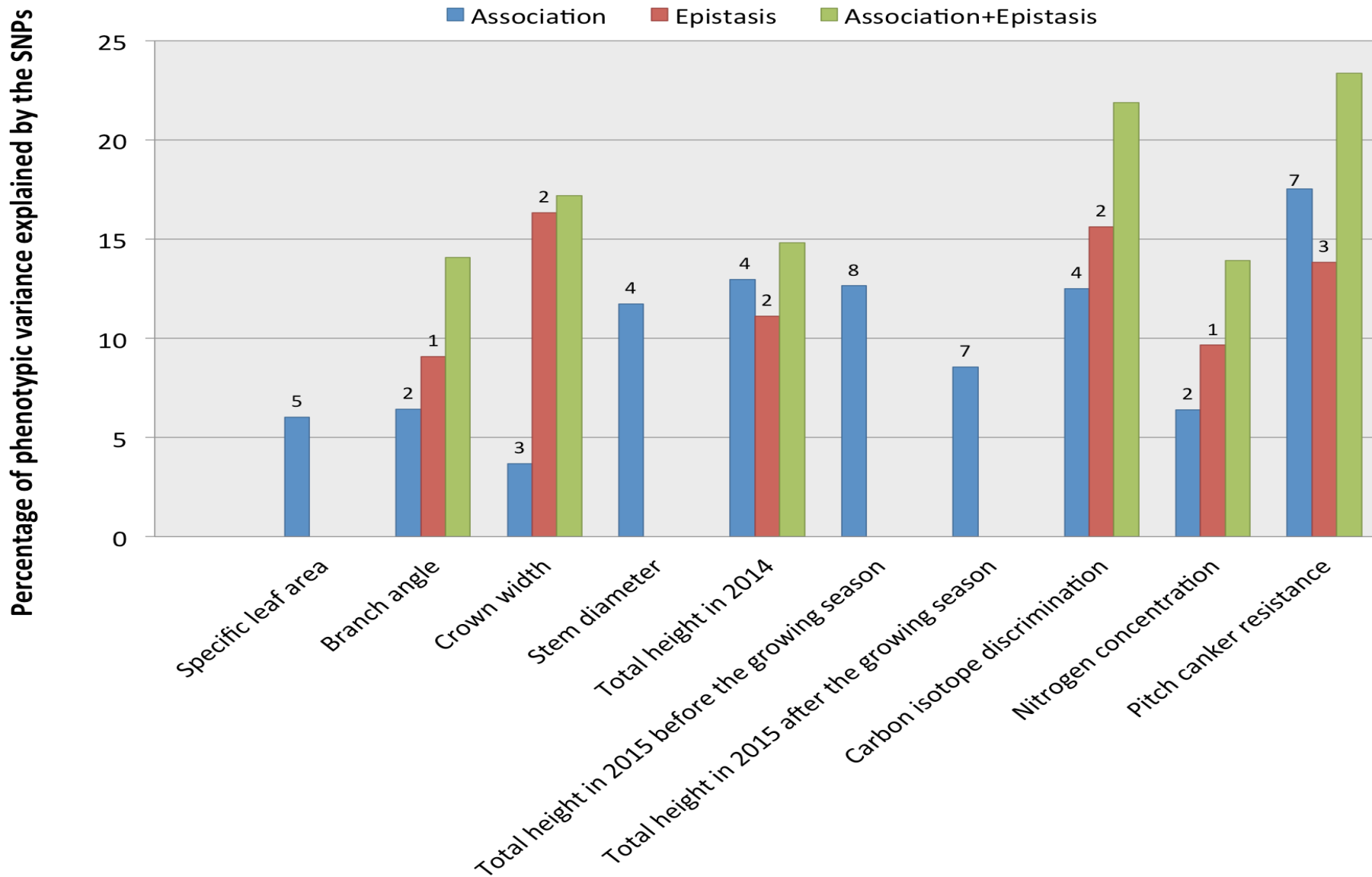


Epistasis Analyses

Traits	Numbers of SNP-SNP interaction
Branch angle	1
Crown width	2
Total height in 2014	2
Carbon isotope discrimination	2
Nitrogen concentration	1
Pitch canker resistance	3



Percentage of phenotypic variance for each trait contributed by the SNPs detected by association and epistasis





Outline

- Use exome capture and sequencing to identify single nucleotide polymorphisms (SNPs) in loblolly pine and genotype the ADEPT2 population
- Associate the genotyped SNPs with growth and adaptive traits to discover the candidate genes
- **Dissecting the genetic basis of cellular phenotypes by association mapping and network construction**
- Associate the genotyped SNPs with environmental variables and detect signals of selection which are related to local adaptation



Dissecting the genetic basis of cellular phenotypes by association mapping and network construction

Cellular phenotypes

- ✓ Expression of 111 xylem/wood development genes (Palle et al., 2011)
- ✓ Expression of 88 disease- and drought-responsive genes (Seeve 2010)

2,580 associations, including 2,434 SNPs and expression of 193 genes

Data was acquired under a grant from the National Science Foundation
(DBI-0501763; IOS-PGRP-0501763)



Dissecting the genetic basis of cellular phenotypes by association mapping and network construction

Cellular phenotypes

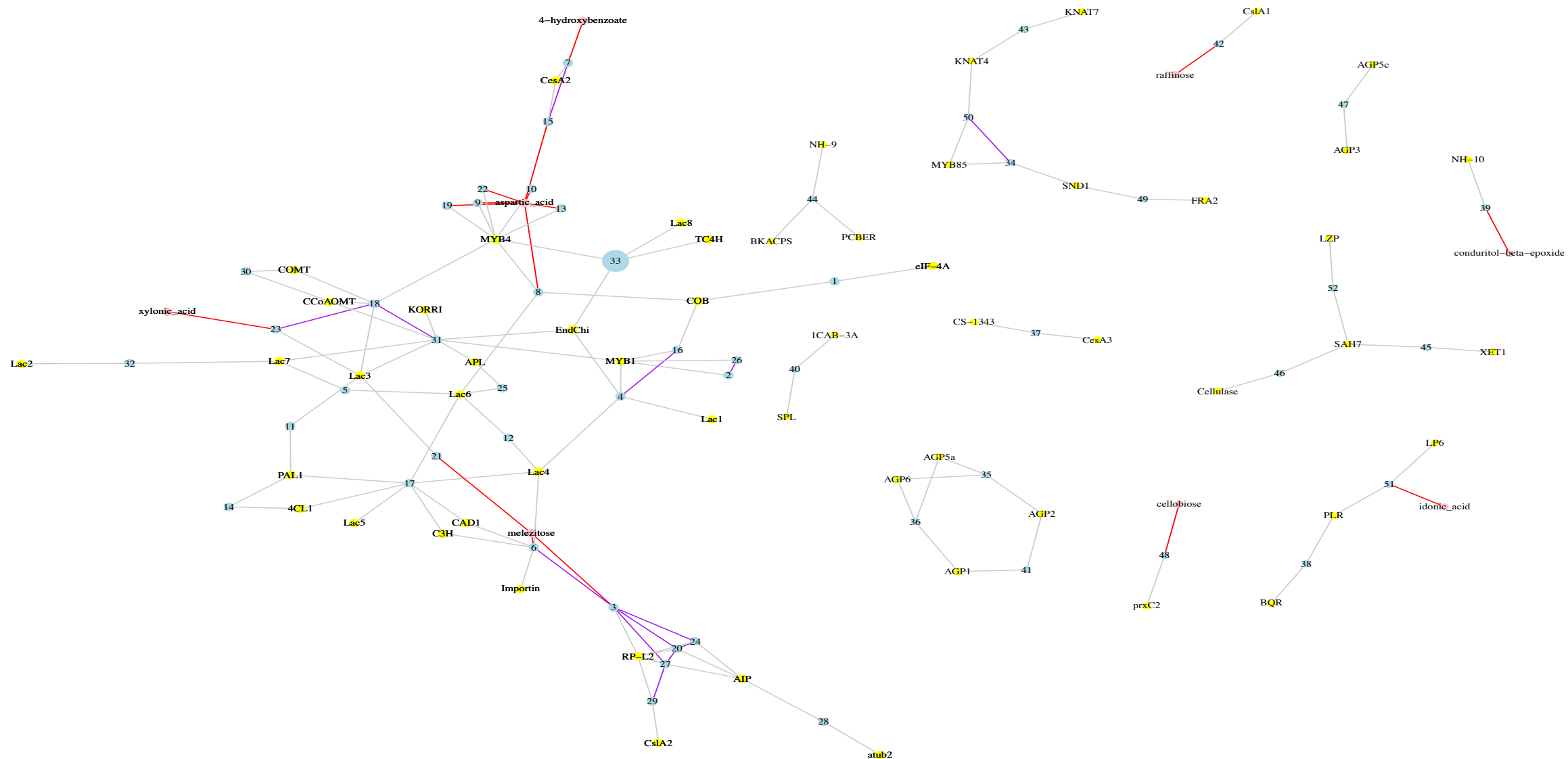
✓ Concentration of 82 metabolites (Eckert et al., 2012)

536 associations, including 536 SNPs and 53 metabolite concentrations

Data was acquired under a grant from the National Science Foundation
(DBI-0501763; IOS-PGRP-0501763)



Network plot connecting SNPs, gene expressions and metabolites





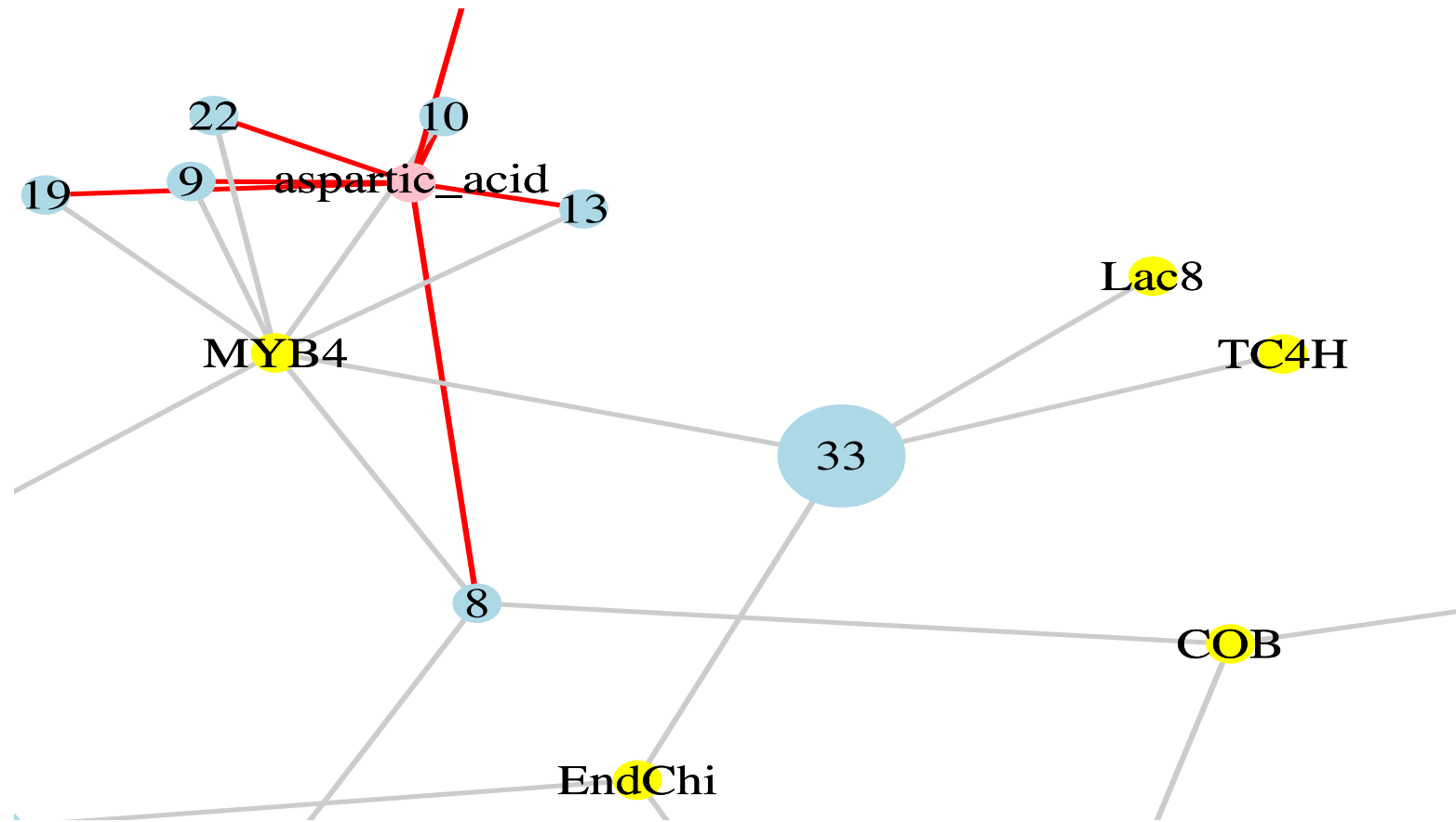
SNP#33 resides in transcription factor GAMYB

EndChi: wood development enzyme gene

Lac8: Lignin biosynthesis enzyme gene

MYB4: wood development transcription factor gene

TC4H: Lignin biosynthesis enzyme gene





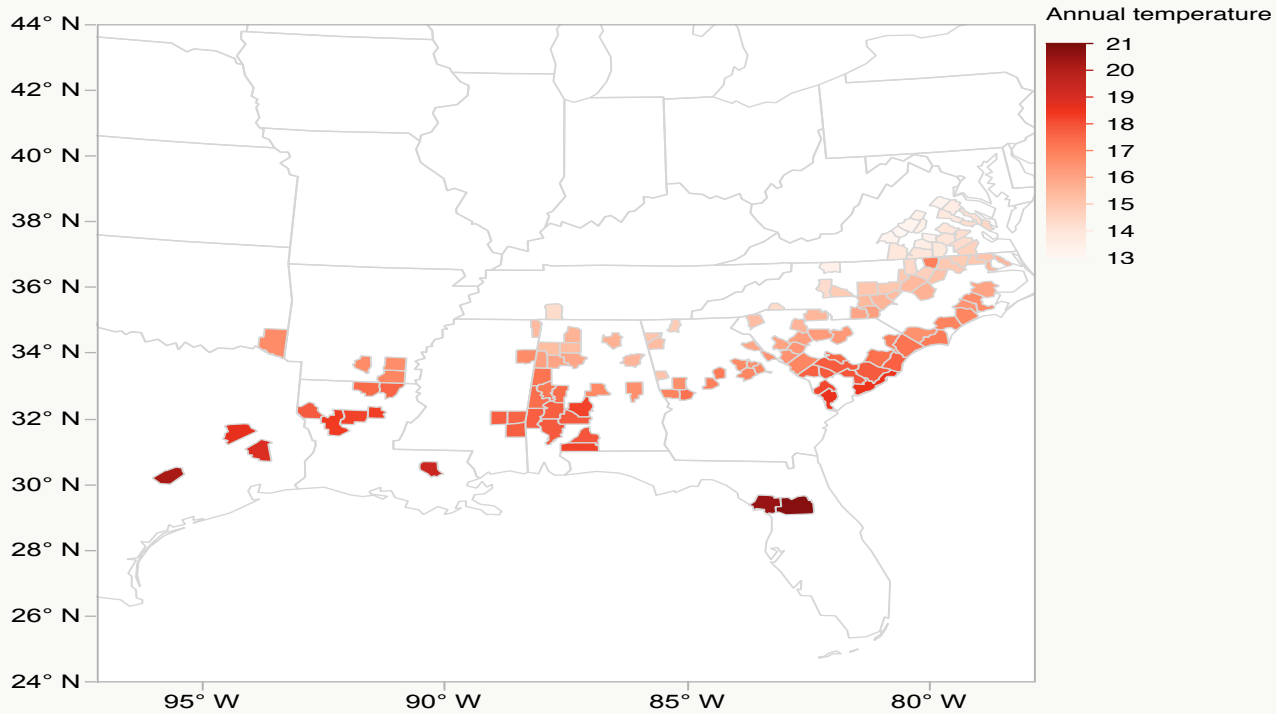
Outline

- Use exome capture and sequencing to identify single nucleotide polymorphisms (SNPs) in loblolly pine and genotype the ADEPT2 population
- Associate the genotyped SNPs with growth and adaptive traits to discover the candidate genes
- Dissecting the genetic basis of cellular phenotypes by association mapping and network construction
- Associate the genotyped SNPs with environmental variables and detect signals of selection which are related to local adaptation

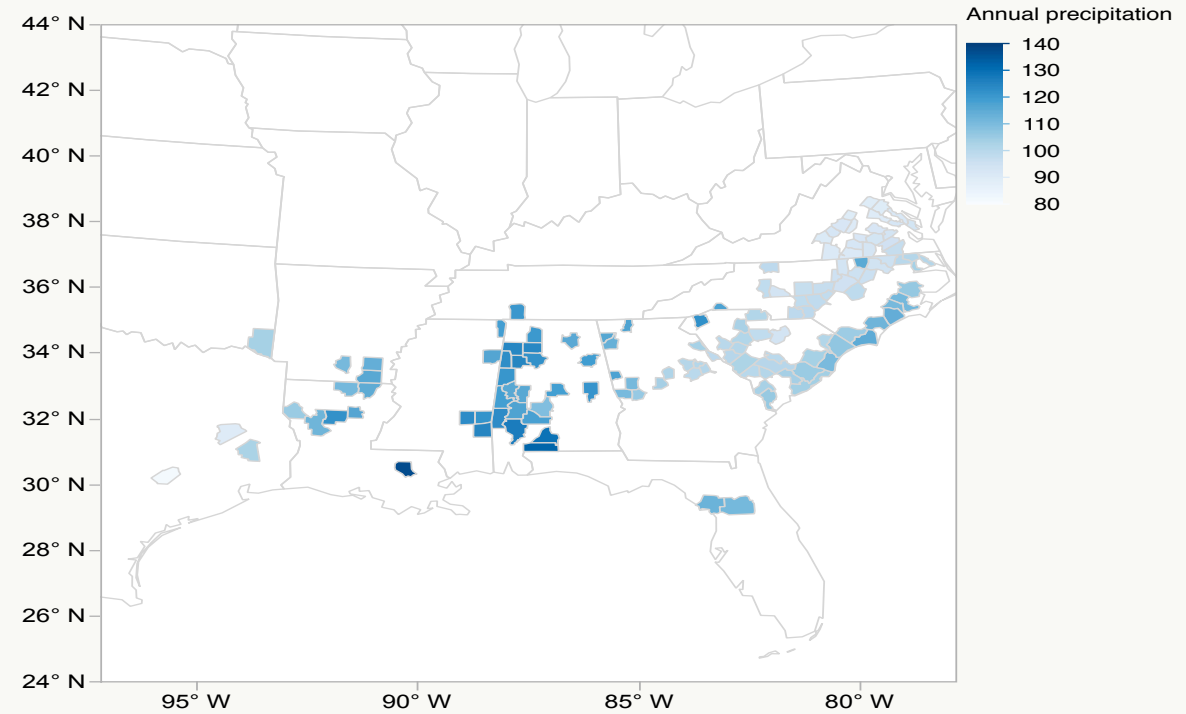


Climate in the loblolly pine natural range

- Annual temperature



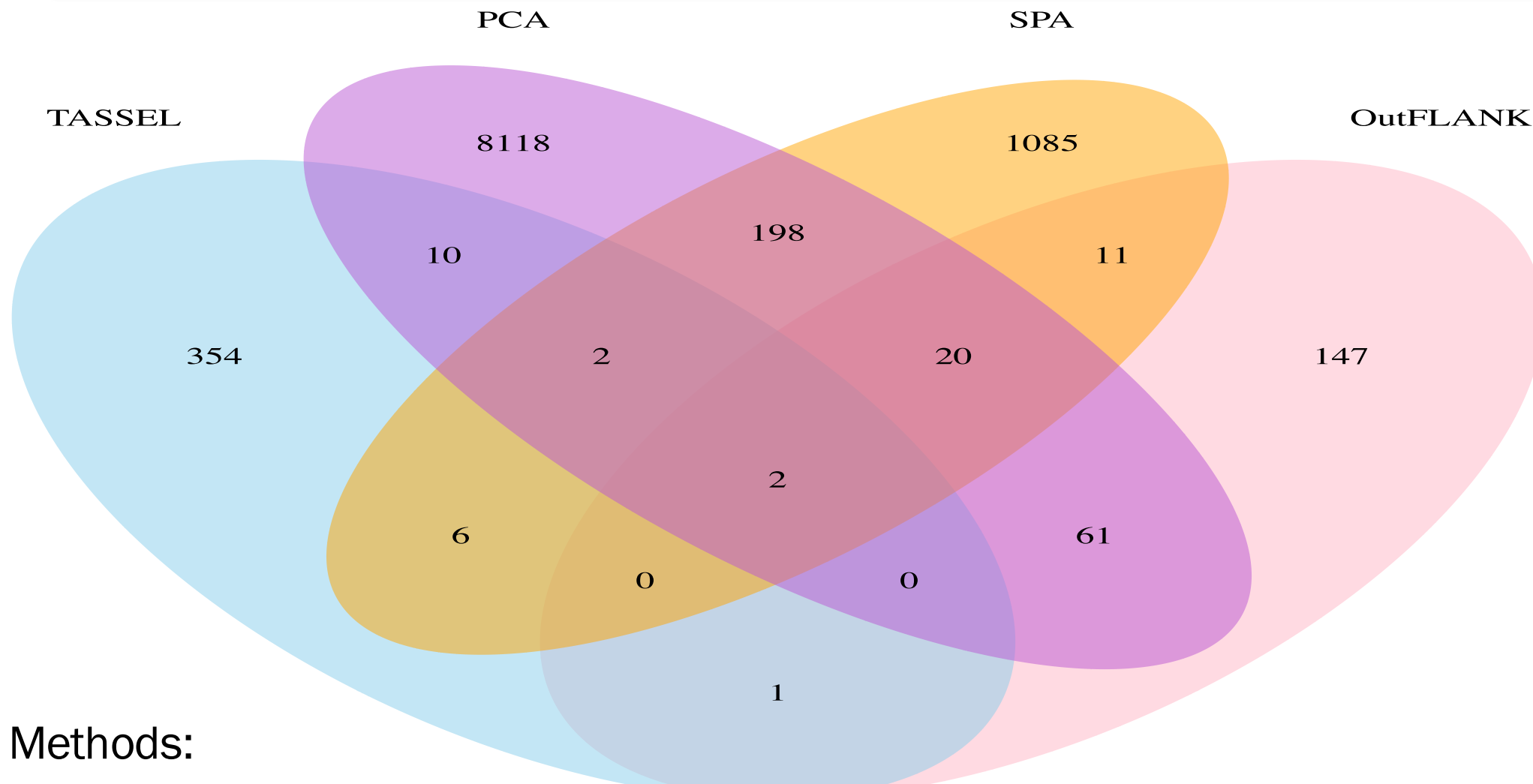
- Annual precipitation



Climate data is from Eckert et al. 2010



Identified SNPs numbers by different methods



Methods:

- ✓ Association mapping by software “TASSEL”
- ✓ SNP outlier detection by software “PCAdapt, OutFlank, SPA”



How much of the genetic variation is uniquely explained by climate?

Software “Vegan” was used to test how much of the genetic variation detected by different methods can be explained by climate, geography or both.

	TASSEL %	OutFLANK %	SPA %
Pure climate	63 %	55%	49%

63% of variation in 375 SNPs can be explained by climate alone.

These SNPs may be useful in seed transfer decisions

Geography is not important for these 375 SNPs.

Many of these SNPs are in genes related to resistance to abiotic stresses.



Conclusions

- Our results demonstrated the efficiency of exome capture for genotyping a species with a large, 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.
- Identification of candidate genes will facilitate elucidation of the genetic architecture of the loblolly pine traits and contribute molecular tools for selection of loblolly pine genotypes adapted to changing climate scenarios.



Acknowledgements

Texas A&M group: Dr. Carol Loopstra, Dr. Konstantin Krutovsky, Dr. Tom Byram, Dr. Claudio Casola, Dr. Tomasz Koralewski, Dr. Jason West, Dr. Jason Vogel, Dr. Fred Raley, Jeffery Puryear, Nathalie Reilly

PINEMAP group members: Dr. Tim Martin, Dr. Ross Whetten, Dr. Dana Nelson and many others

PINEREFSEQ project: Dr. Jill Wegryn, Dr. David Neale, Dr. Kiethanne Mockaitis, et al.

ADEPT2 project: Dr. David Neale, Dr. Barry Goldfarb, Dr. Jill Wegrzyn, Dr. Andrew Eckert, Dr. Carol Loopstra, Dr. Tom Byram, Dr. Sreenath Reddy Palle, Dr. Candace Seeve, Dr. Patrick Cumbie, Dr. John Davis, Dr. Tania Quesada, et al.

Harrison Experimental Forest at Southern Institute of Forest Genetics: Dr. Dana Nelson, Dr. Vikram Chhatre Gay Flurry, Chance Parker, Chuck Burdine

This study was 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.