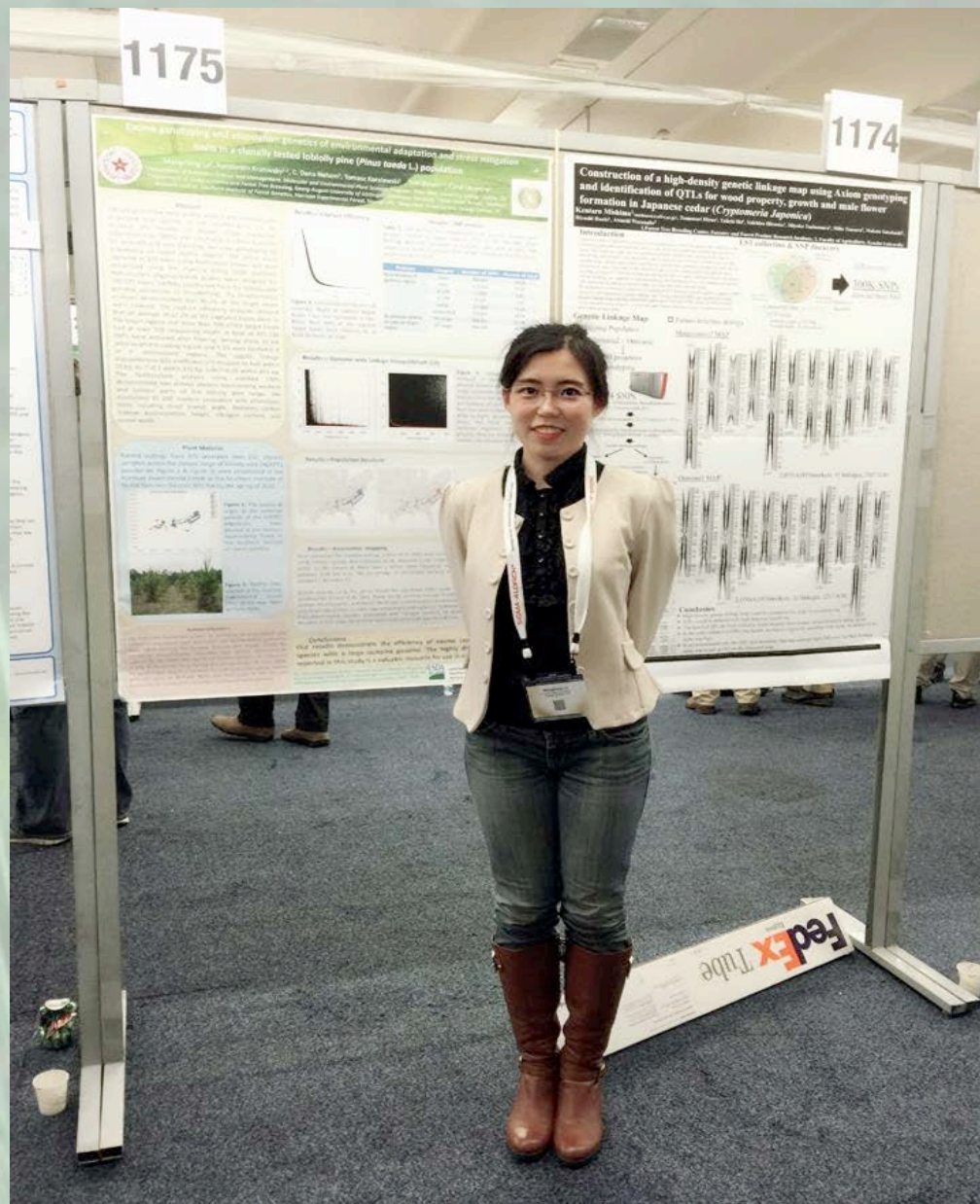


Whole Exome Genotyping in Loblolly Pine and Association Analyses

Mengmeng Lu, **Carol Loopstra**, Konstantin Krutovsky, C. Dana Nelson, Thomas Byram, Tomasz Koralewski, Candace Seeve



United States
Department of
Agriculture

National Institute
of Food and
Agriculture

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

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.

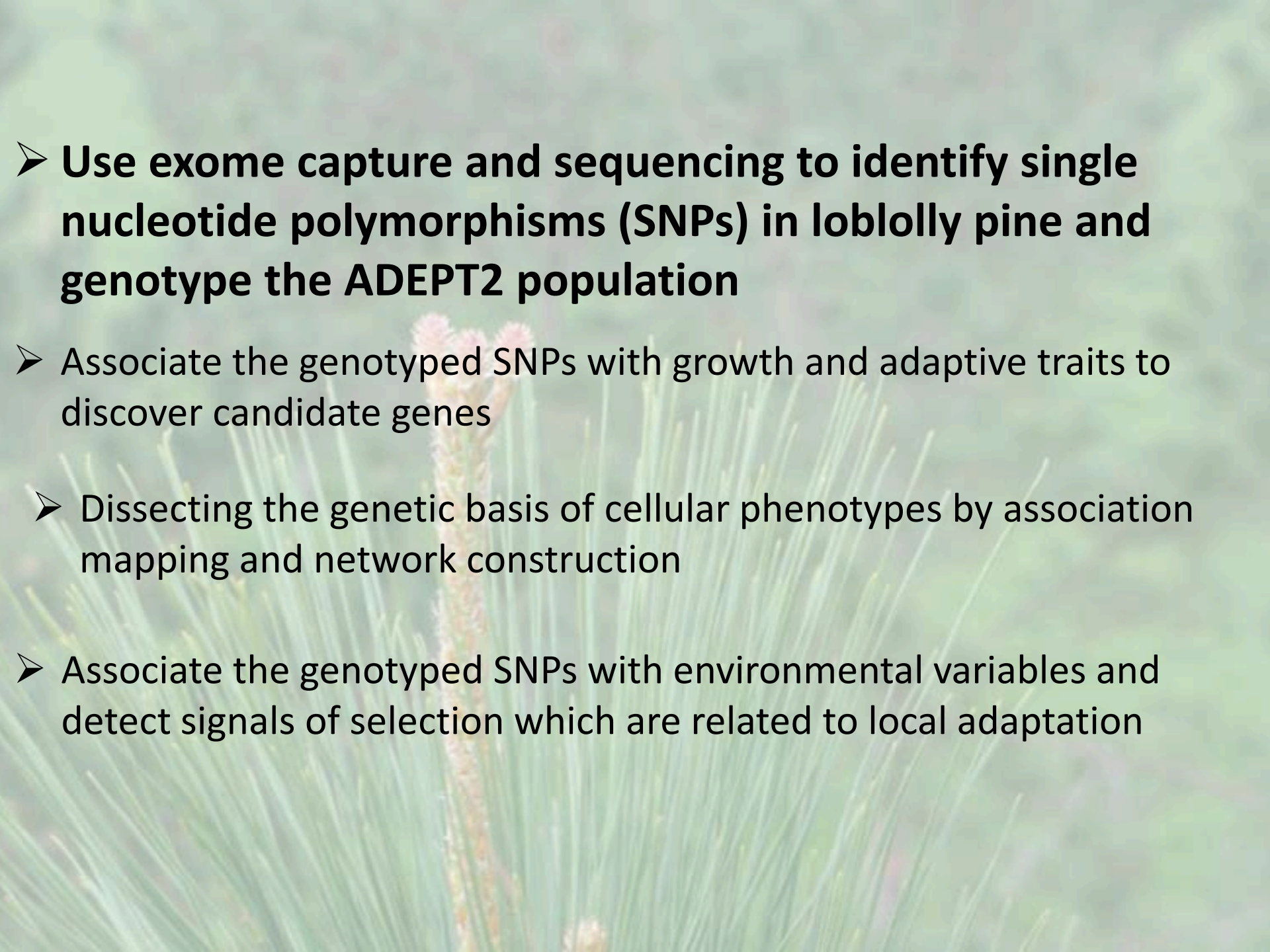


The genome of the sugar pine, the so-called "king of conifers," is the largest ever sequenced for any organism.
(Credit: [dotpolka/Flickr](#))

FORESTS | UNIVERSITY OF CALIFORNIA, DAVIS

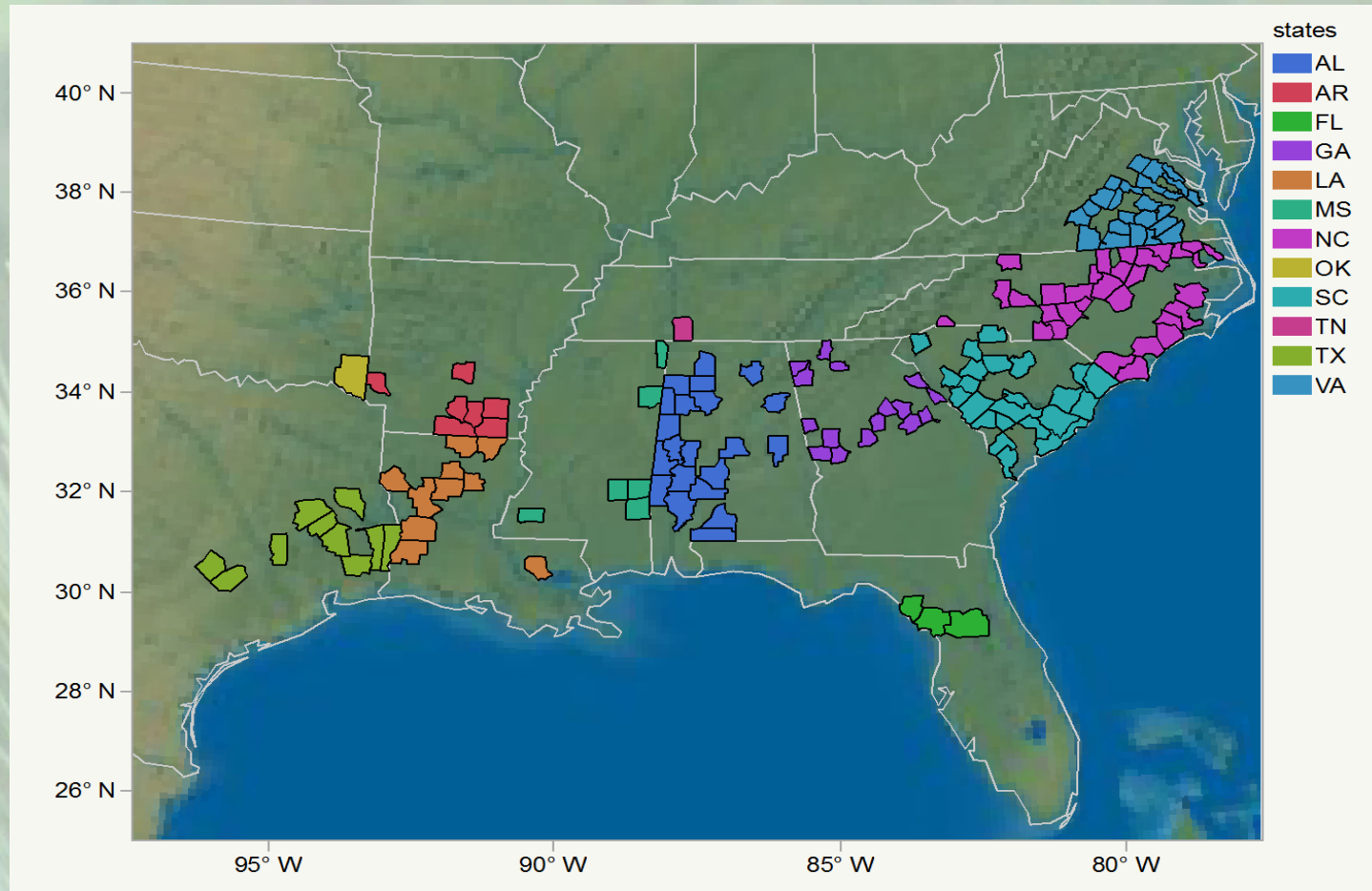
Sugar pine genome is 10X the size of human genome

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

- 
- **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.)

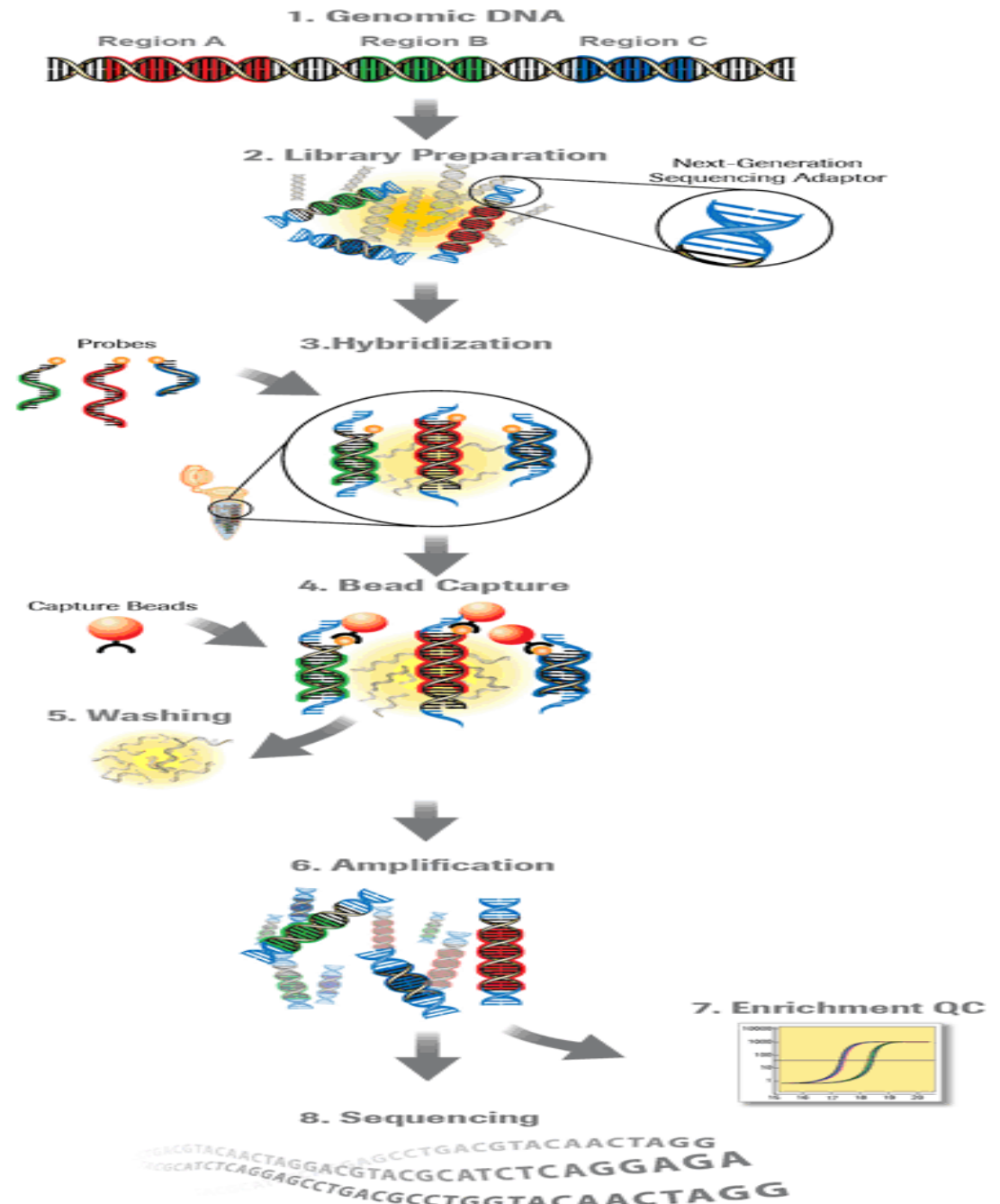


384 × 3 unrelated trees were planted

Lu et al. (2016) *BMC Genomics* 17:730.

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

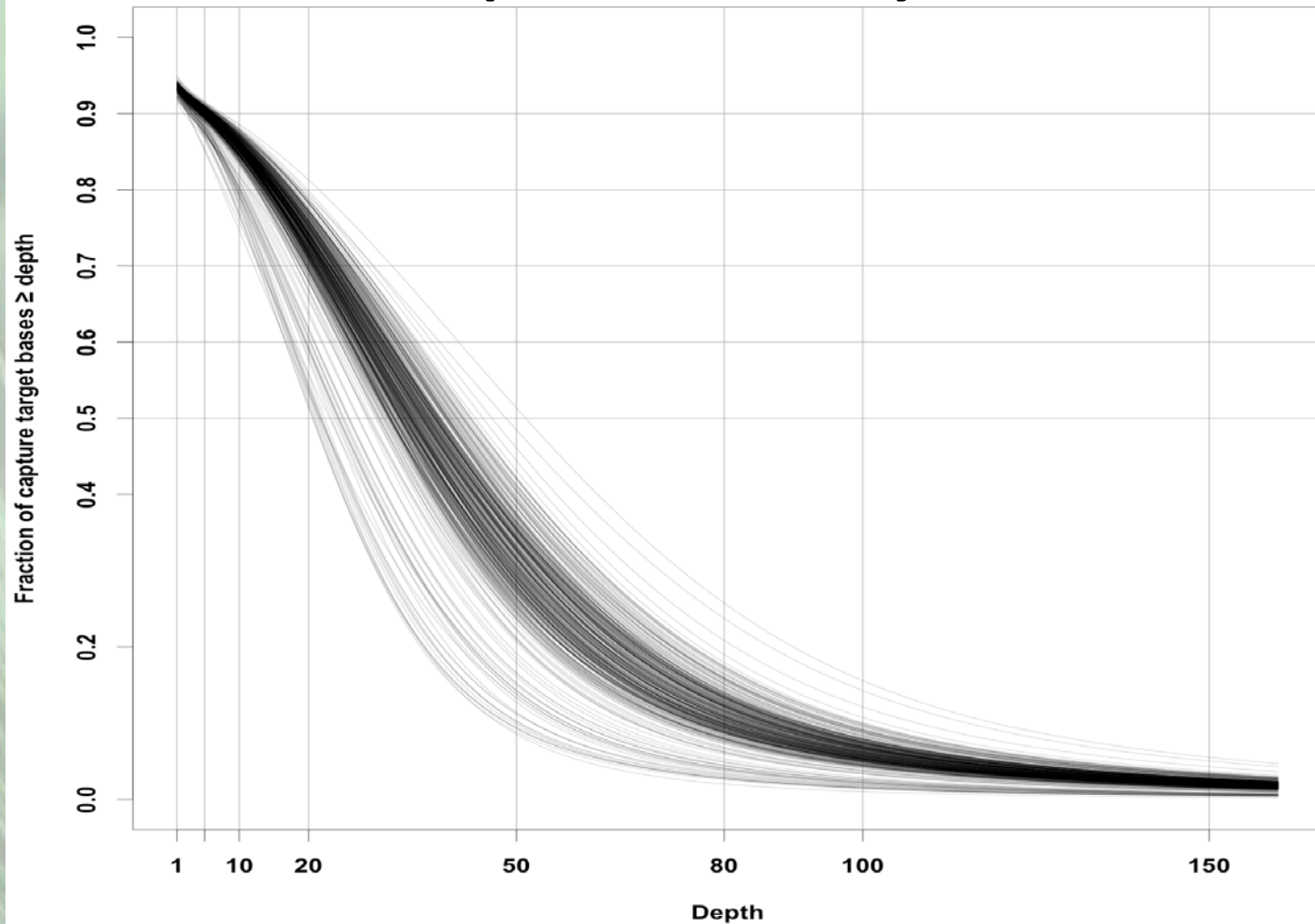
Exome genotyping, linkage disequilibrium and population structure in loblolly pine (*Pinus taeda* L.)

- 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.

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

Capture efficiency



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

Bi-allelic sites

10X sequencing depth in 90% of individuals

Minor allele frequency ≥ 0.05

972,720 SNPs

Bi-allelic sites

5X sequencing depth in all individuals (no missing data)

Minor allele frequency ≥ 0.01

2,822,609 SNPs

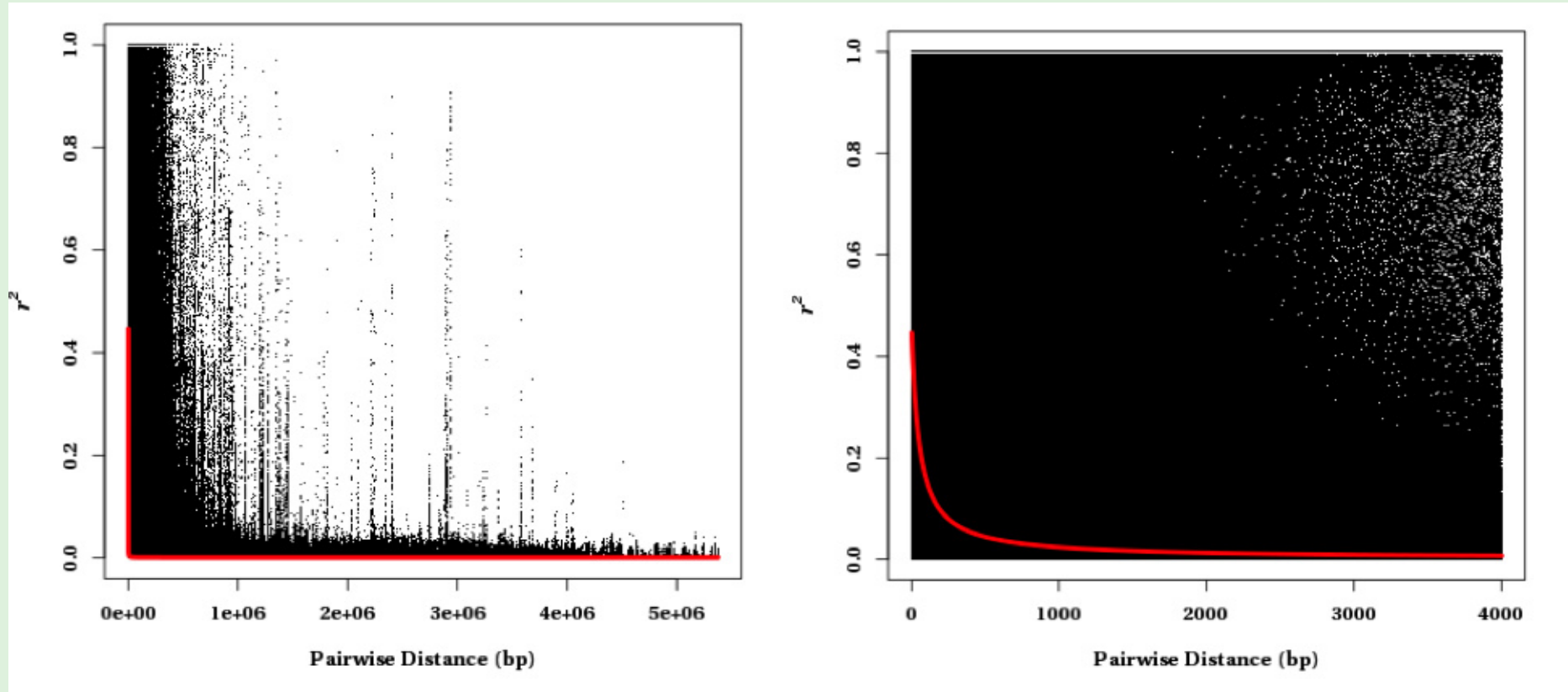
Exome genotyping, linkage disequilibrium and population structure in loblolly pine (*Pinus taeda* L.)

Classification for 972,720 SNPs

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

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

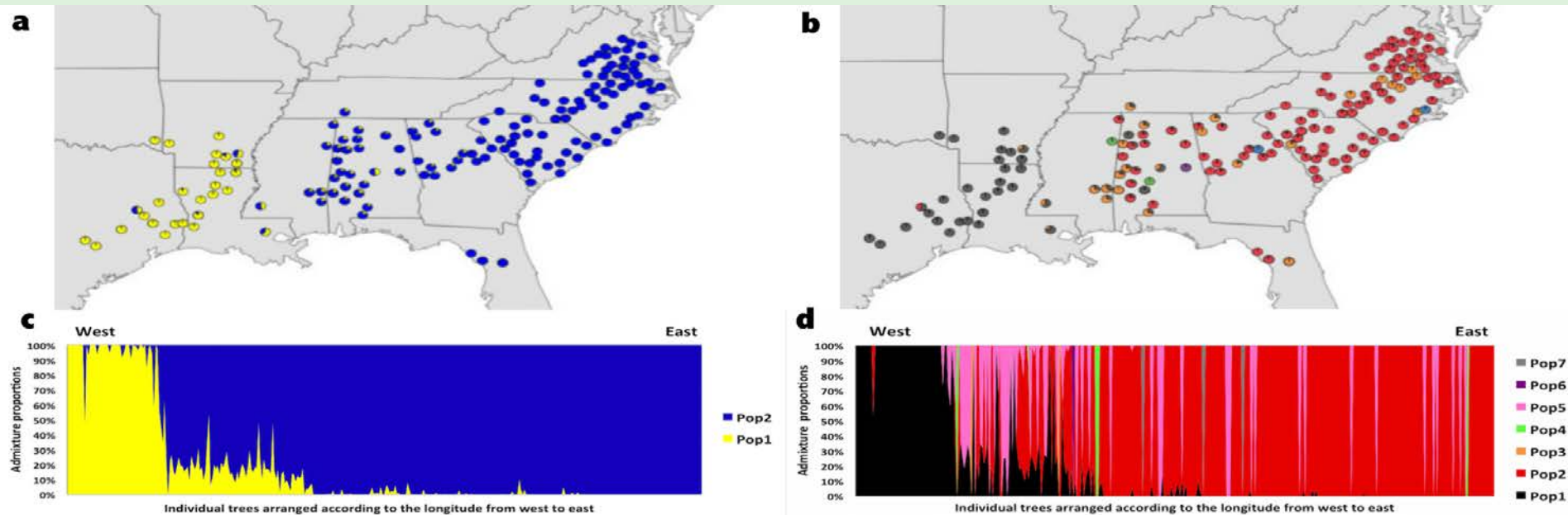
Linkage disequilibrium (LD) decay plot



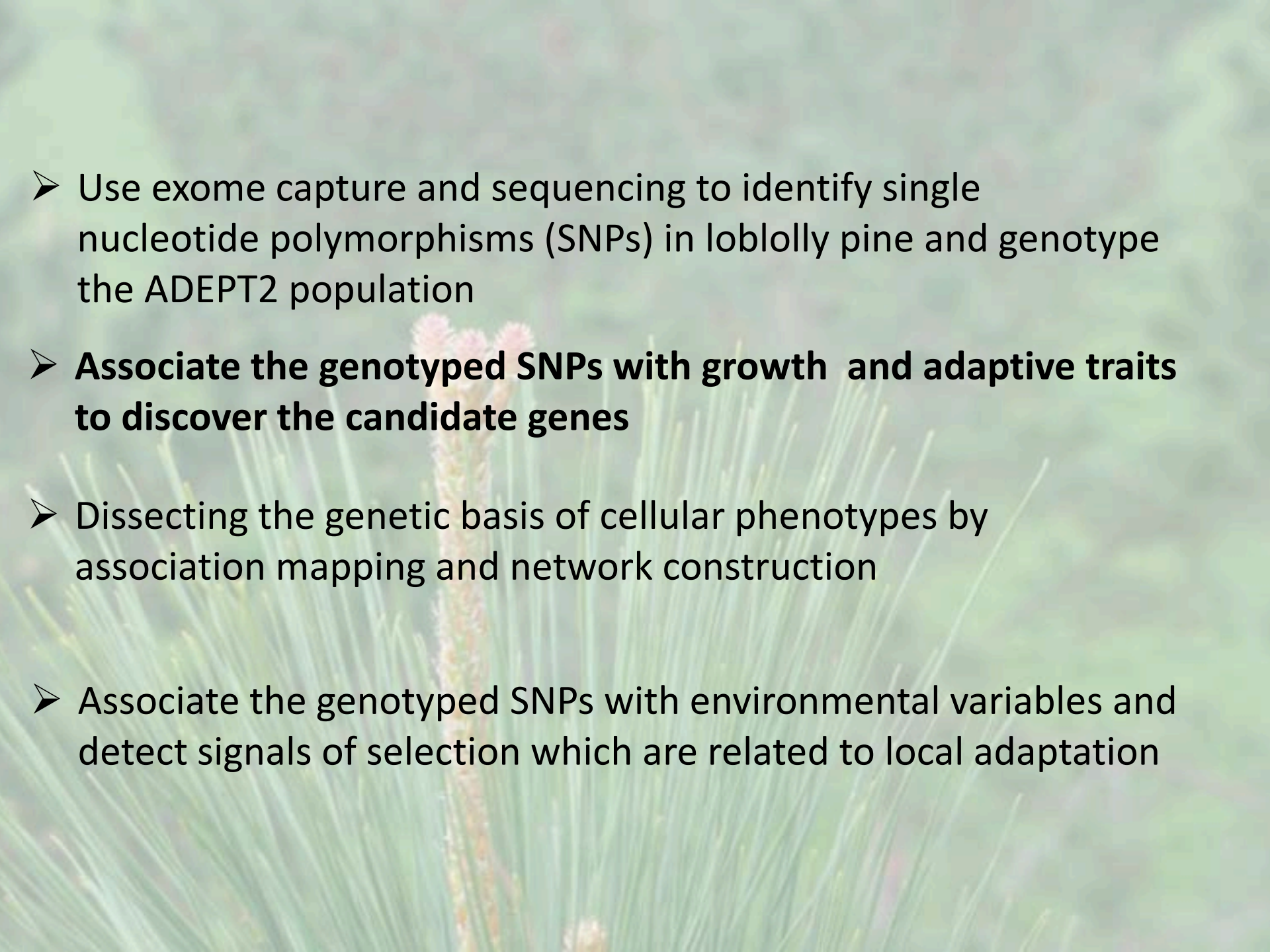
Linkage disequilibrium measure r^2 plotted against the physical distance between all pairs of SNP markers from the same scaffold (left) and between all pairs of SNPs within 400 base pairs.

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

Genetic structure analysis



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

- 
- 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

Association genetics of adaptive and growth traits

Phenotyping

PINEMAP Undergraduate Fellowship Program



Lu, M., Krutovsky, K.V., Nelson, C.D. et al. *Tree Genetics & Genomes* (2017) 13: 57.

Association genetics of adaptive and growth traits

Phenotyping

Growth traits:

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

Crown structure traits:

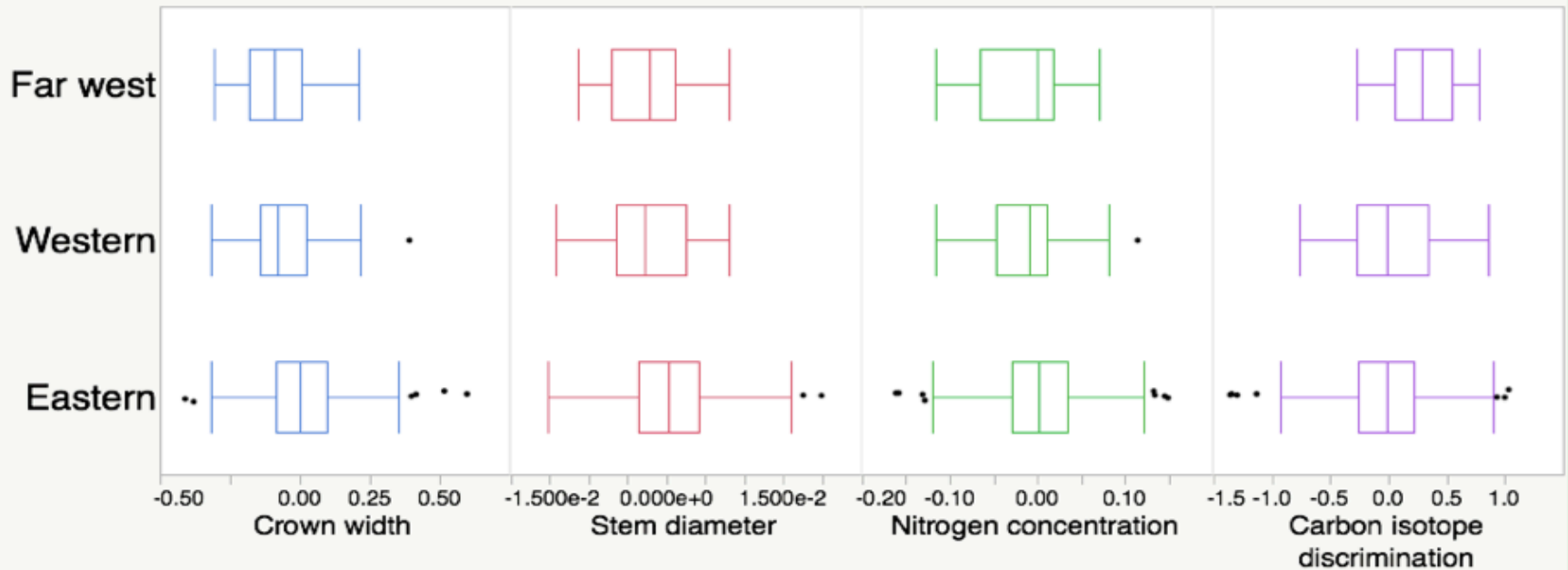
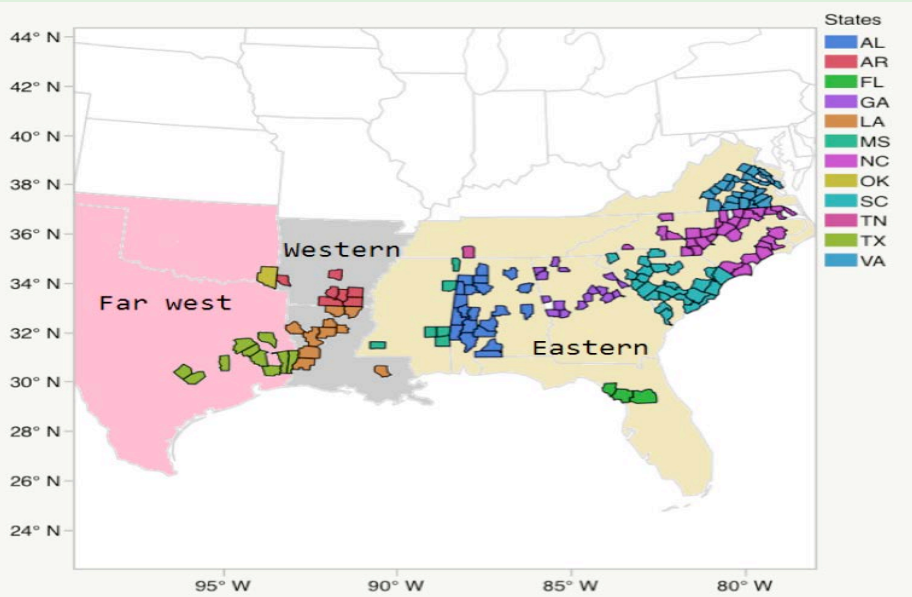
- Specific leaf area (SLA)
- Crown width (CW)
- Branch angle (BA)

Physiology and resistance traits:

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

Association genetics of adaptive and growth traits

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

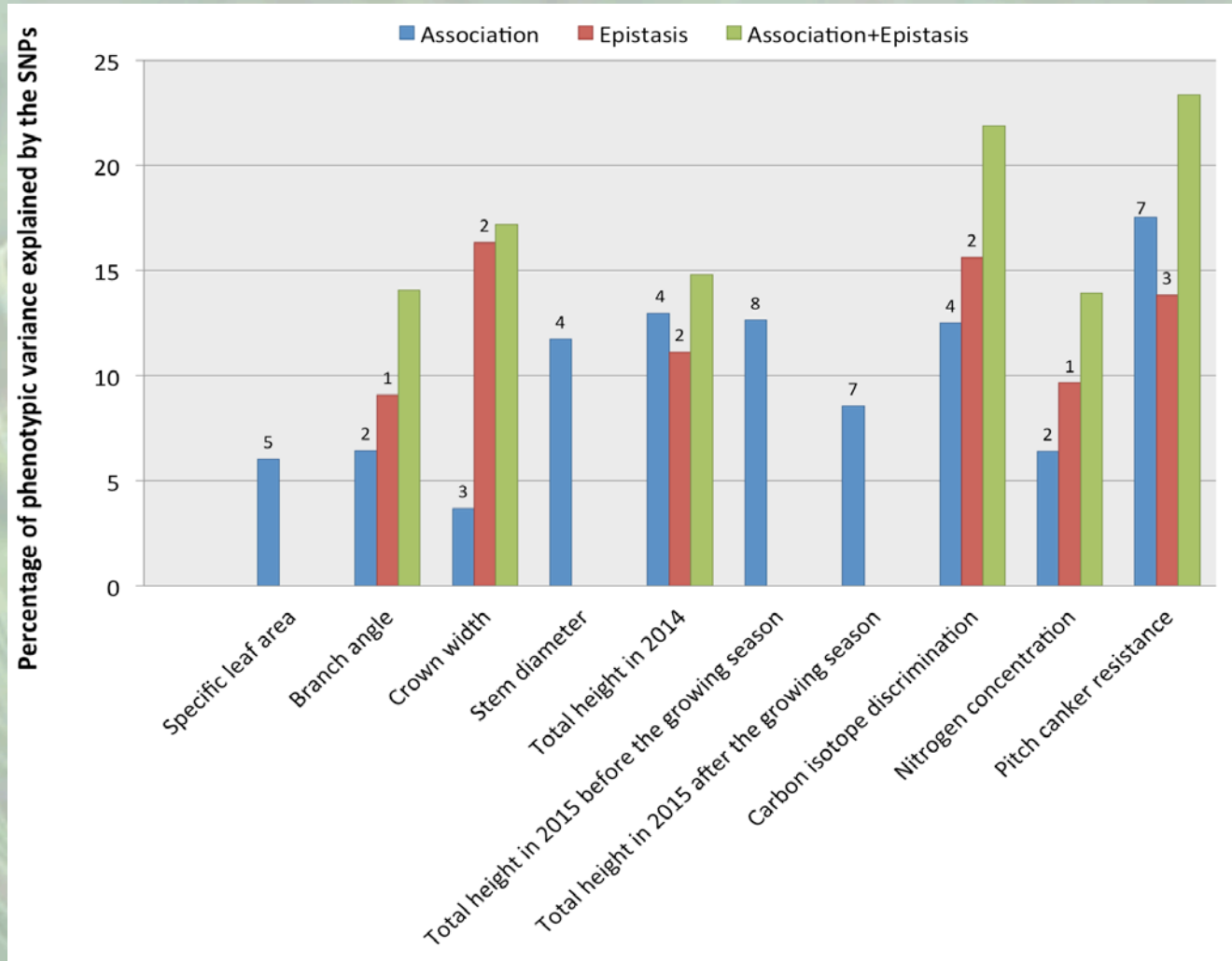
Association genetics of adaptive and growth traits

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

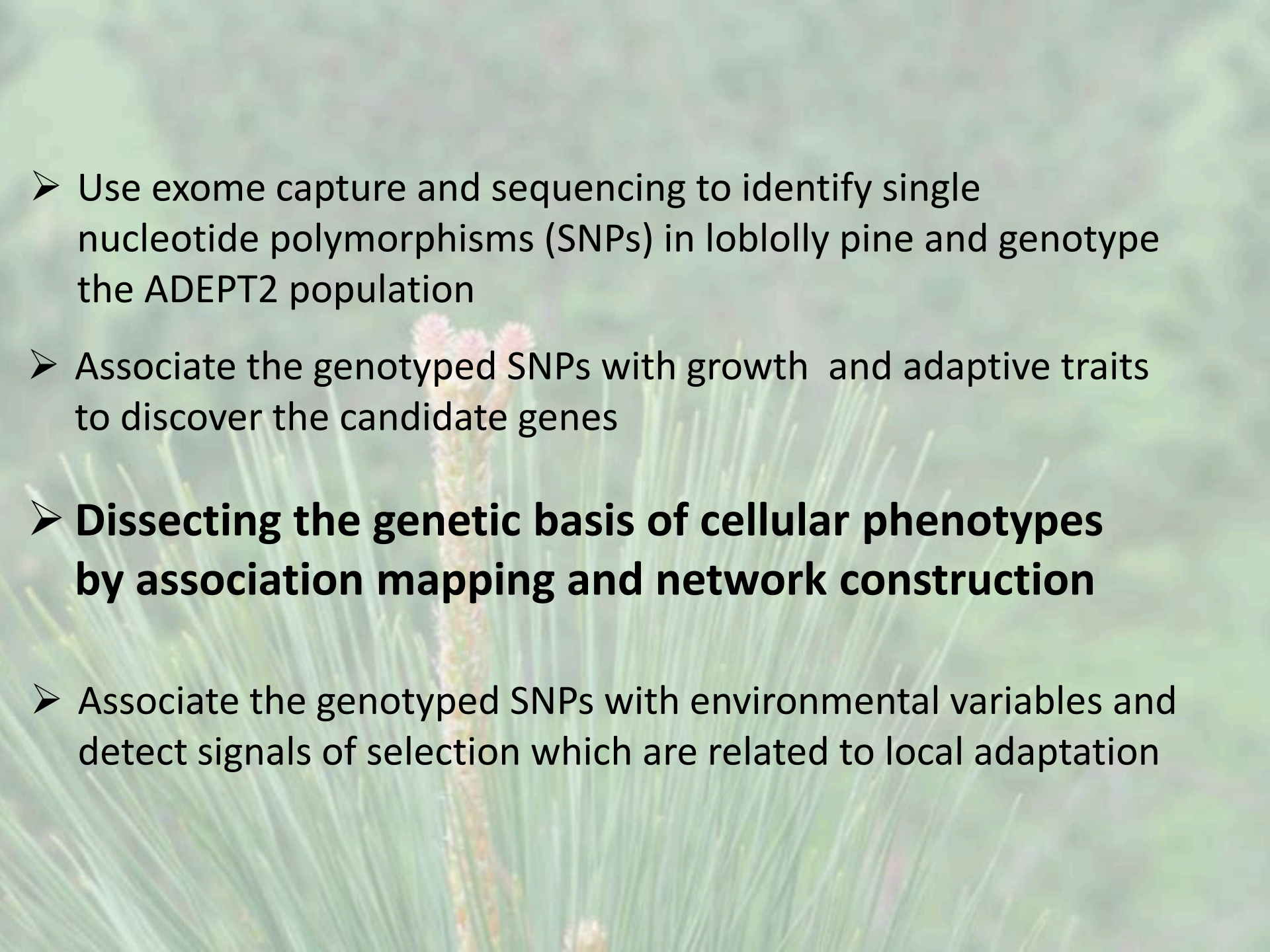
Association genetics of adaptive and growth traits

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



Association genetics of adaptive and growth traits

Location	Number
Coding region – non-synonymous	12 change in AA 2 change to stop
Coding regions -synonymous	6
3' or 5' UTR	2
3' or 5' potential regulatory regions	5
Intron	3
Unclassified	6

- 
- 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

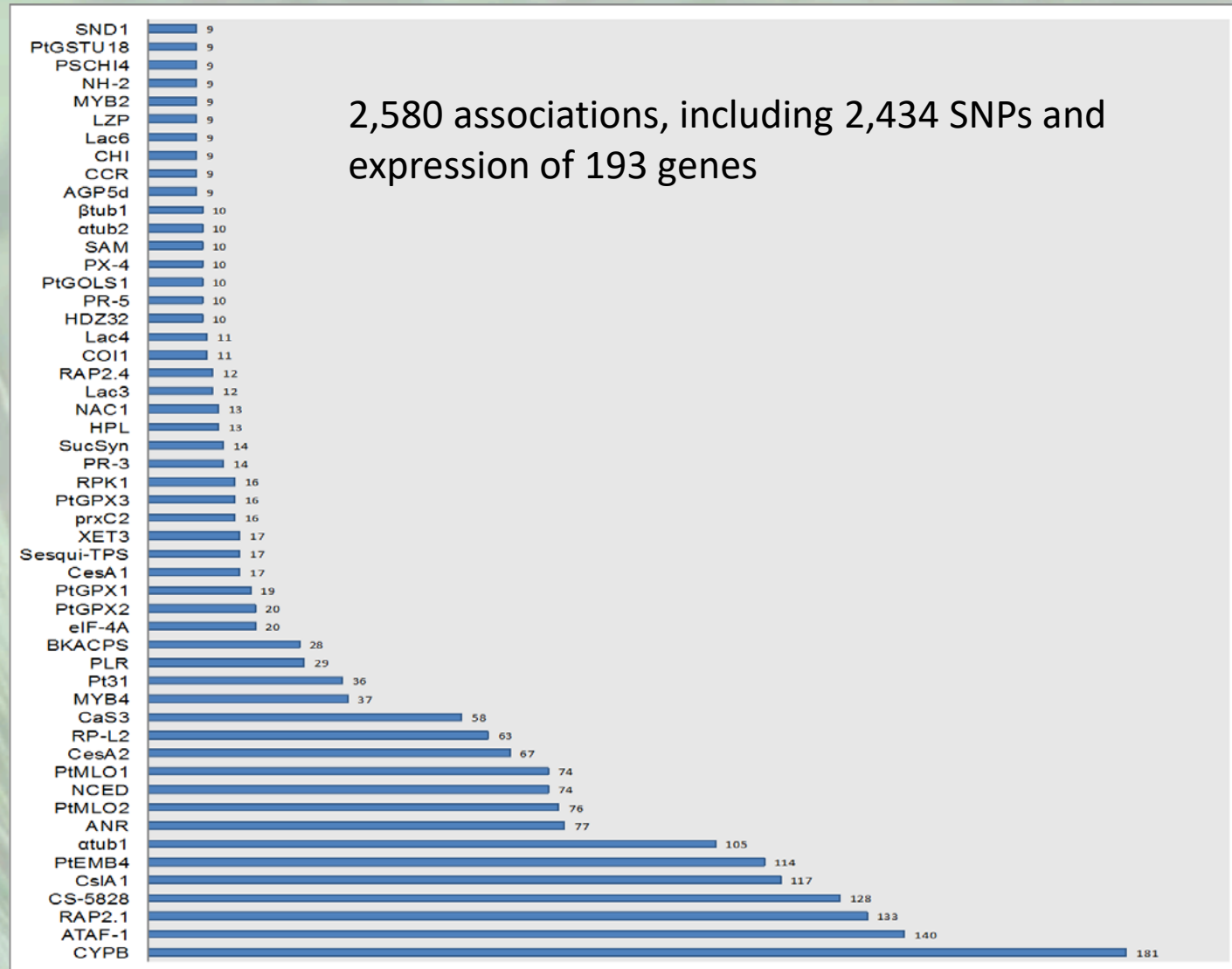
Cellular phenotypes

- ✓ 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 the ADEPT2 grant from the National Science Foundation (DBI-0501763; IOS-PGRP-0501763)

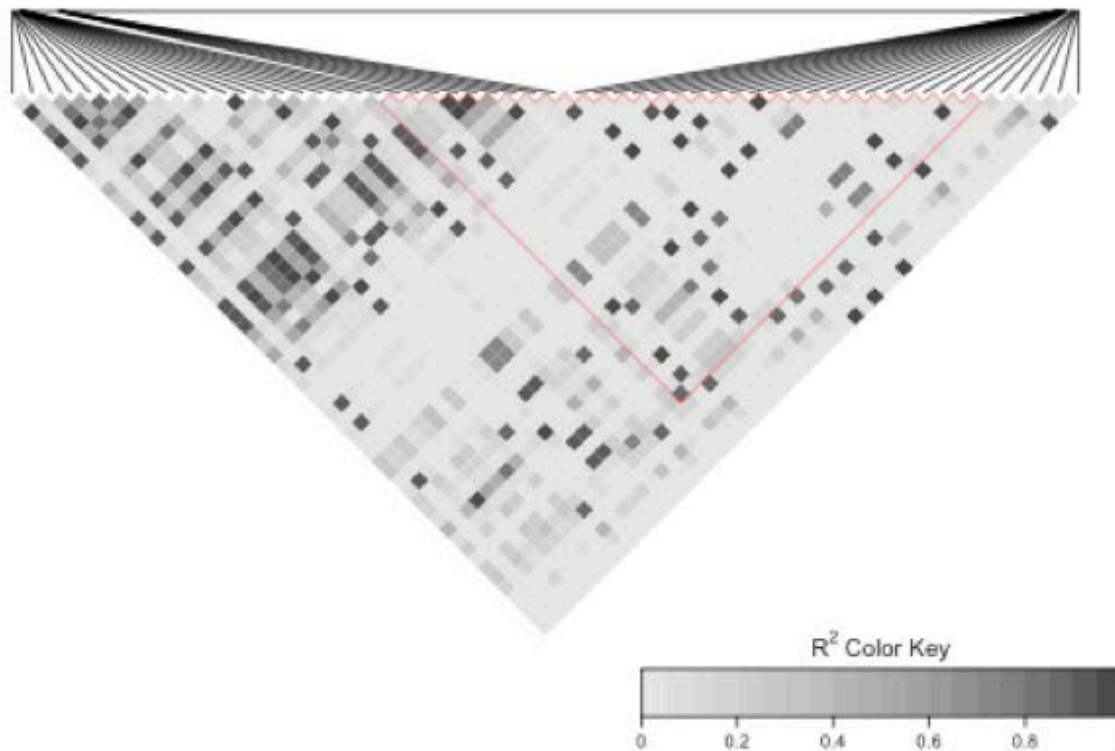
Dissecting the genetic basis of cellular phenotypes by association mapping and network construction



Are there large haplotype blocks?

Pairwise LD on tscaffold6318 from 700894 bp to 746243 bp

Physical Length:46kb



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

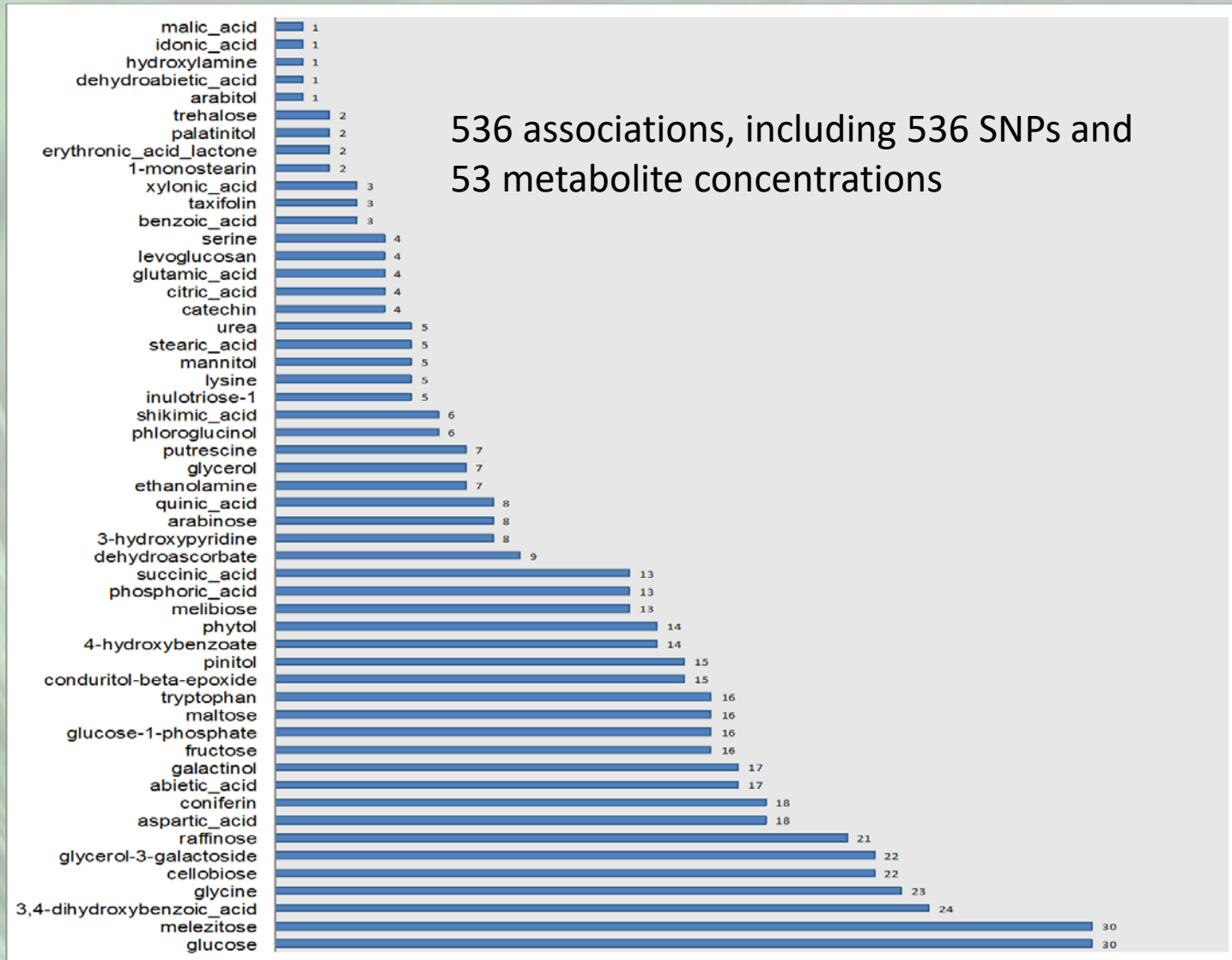
Cellular phenotypes

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

536 associations, including 536 SNPs and 53 metabolite concentrations

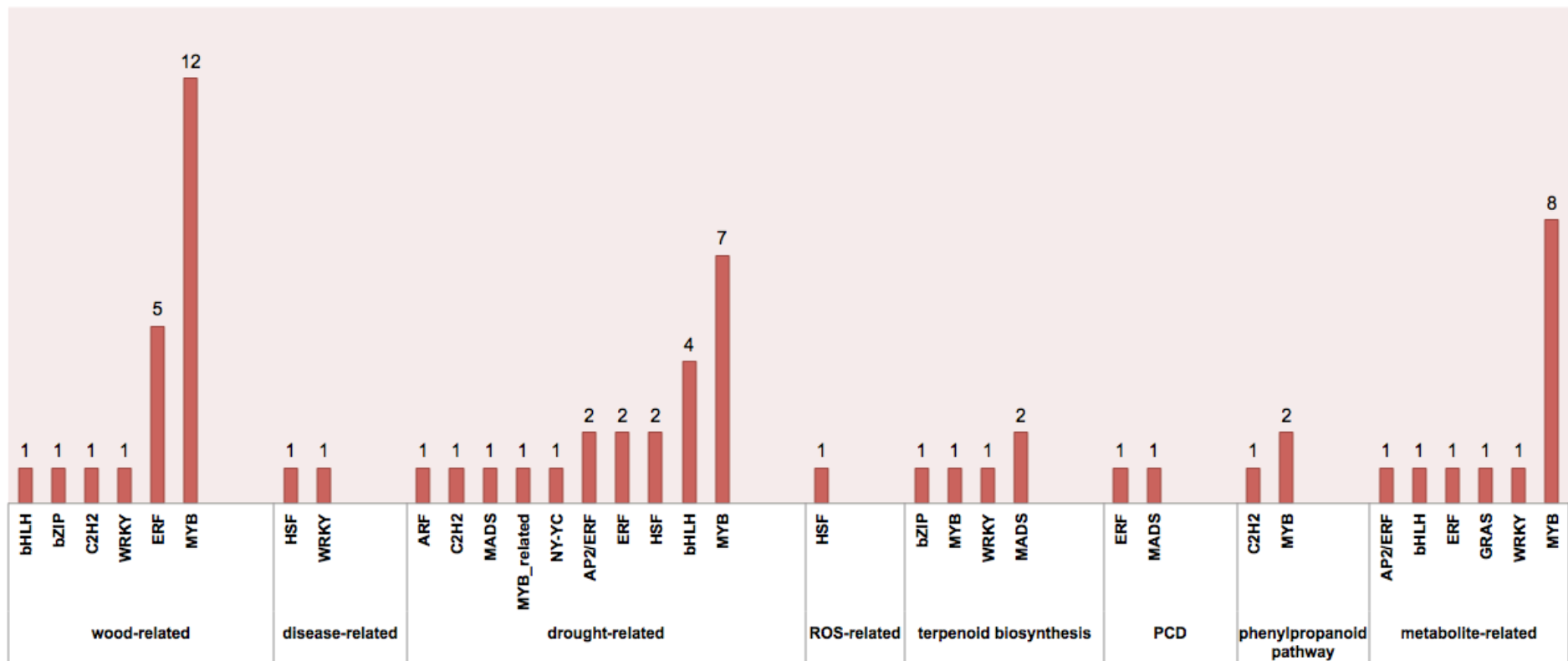
Data was acquired under the ADEPT2 grant from the National Science Foundation (DBI-0501763; IOS-PGRP-0501763)

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



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

Candidate transcription factors



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

SNP#33 resides in transcription factor GAMYB

Associated with expression of:

Endochitinase (wood development)

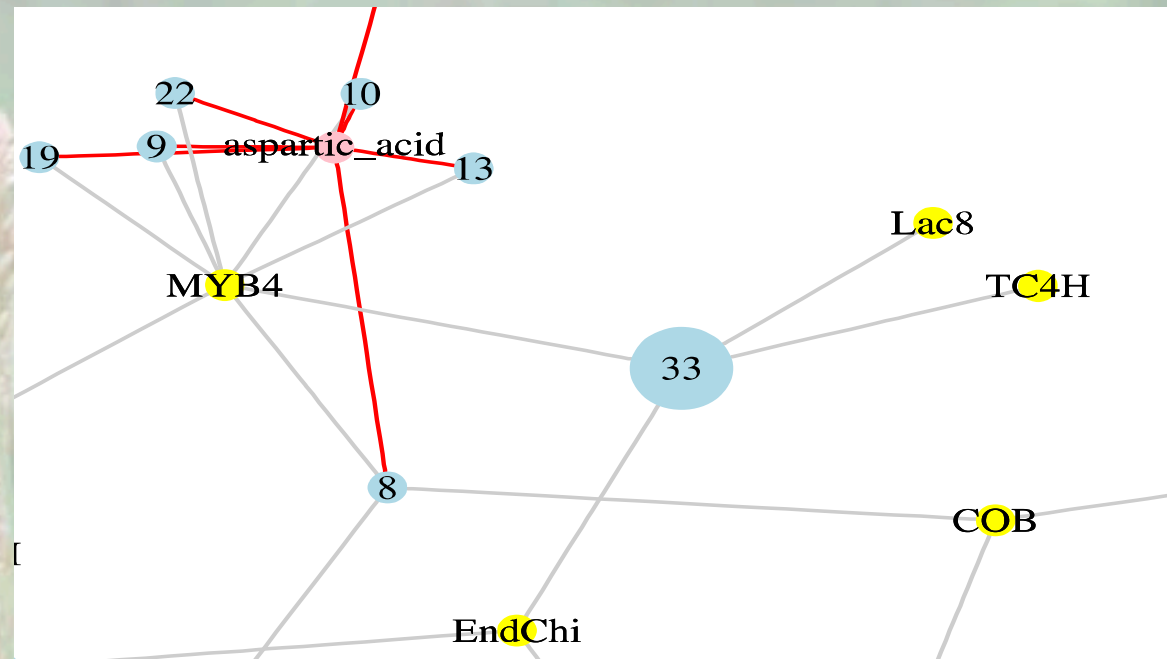
Laccase 8 (Lignin biosynthesis enzyme)

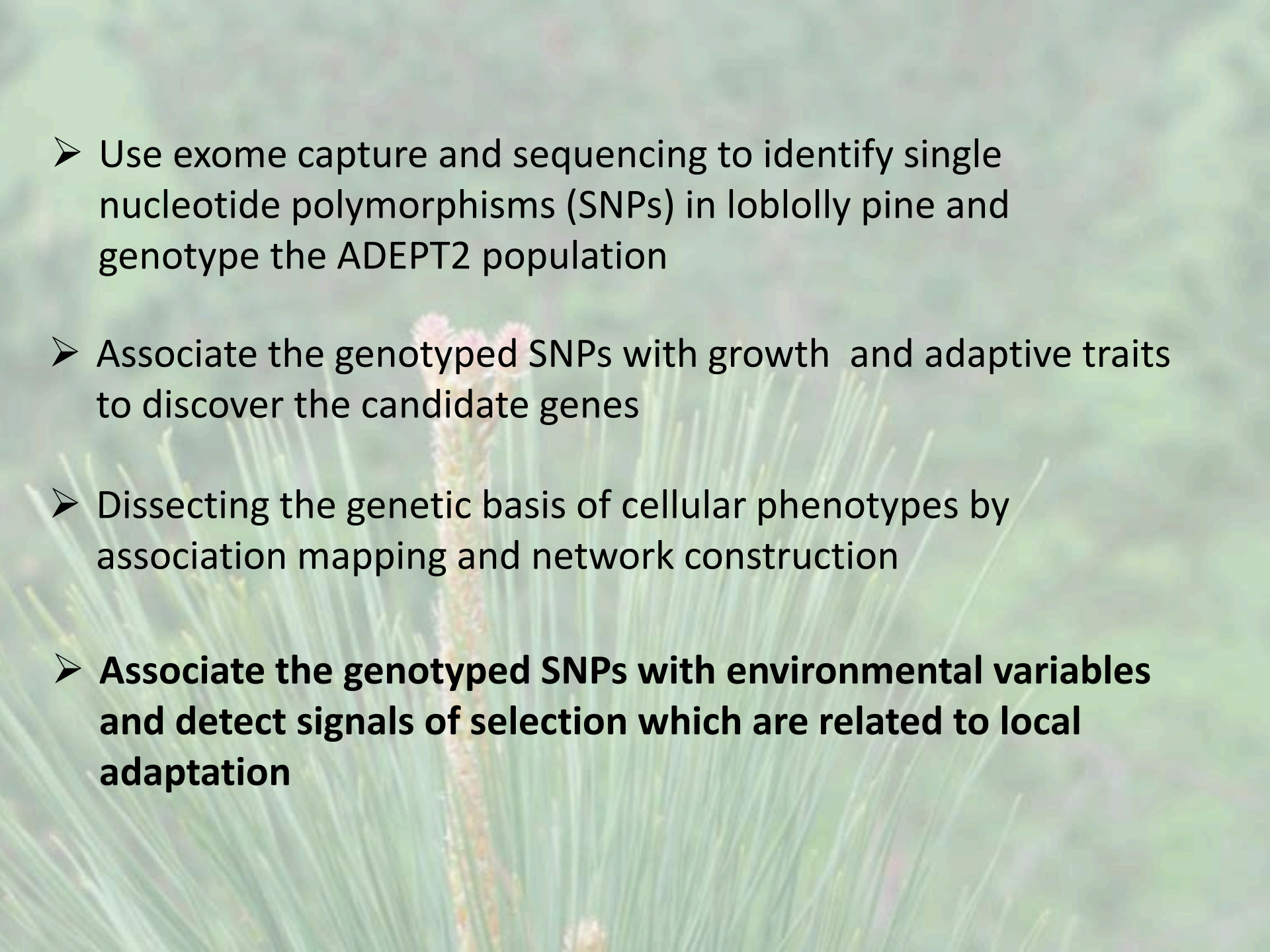
MYB4: wood development transcription factor

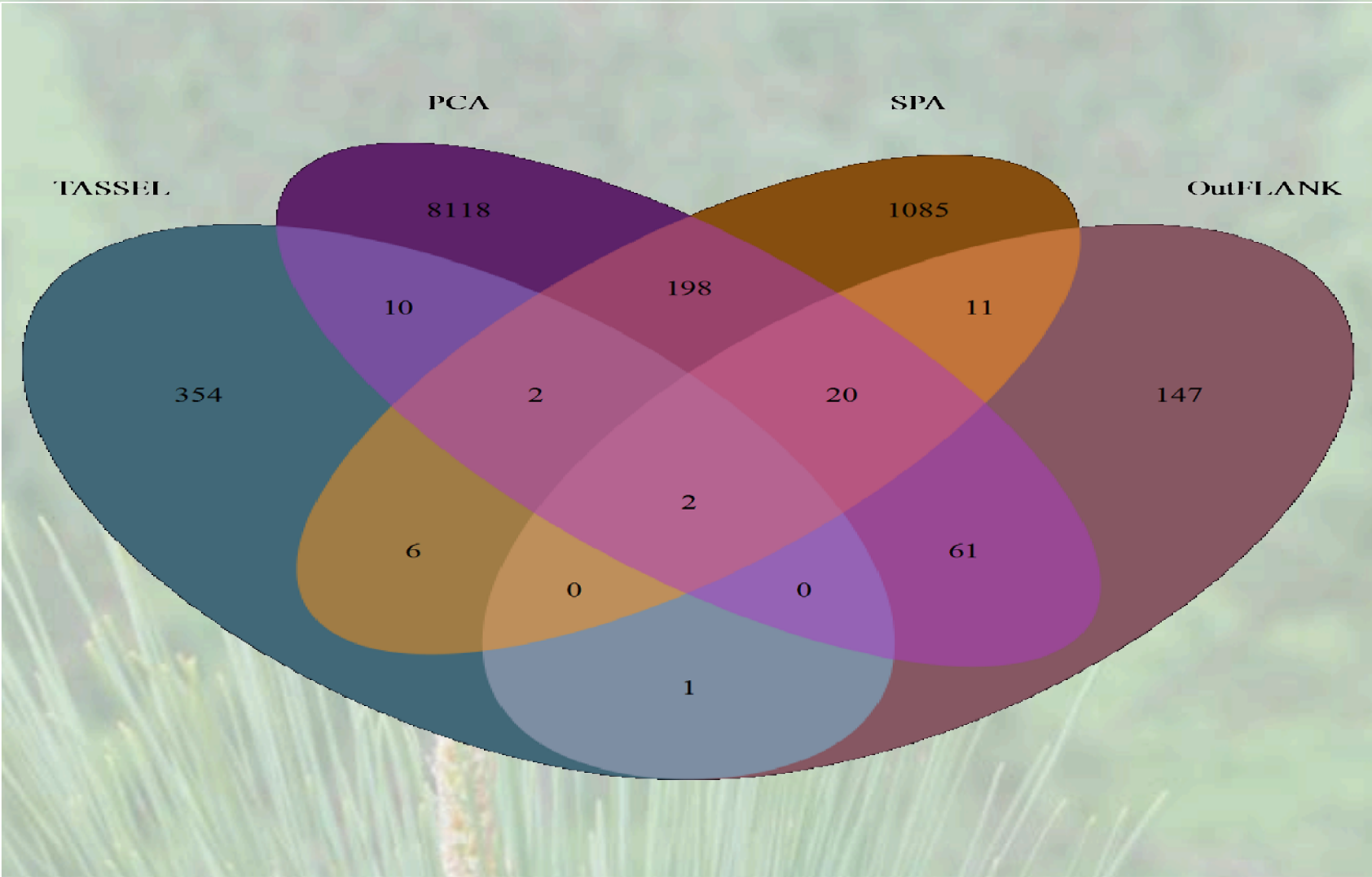
TC4H: Lignin biosynthesis enzyme

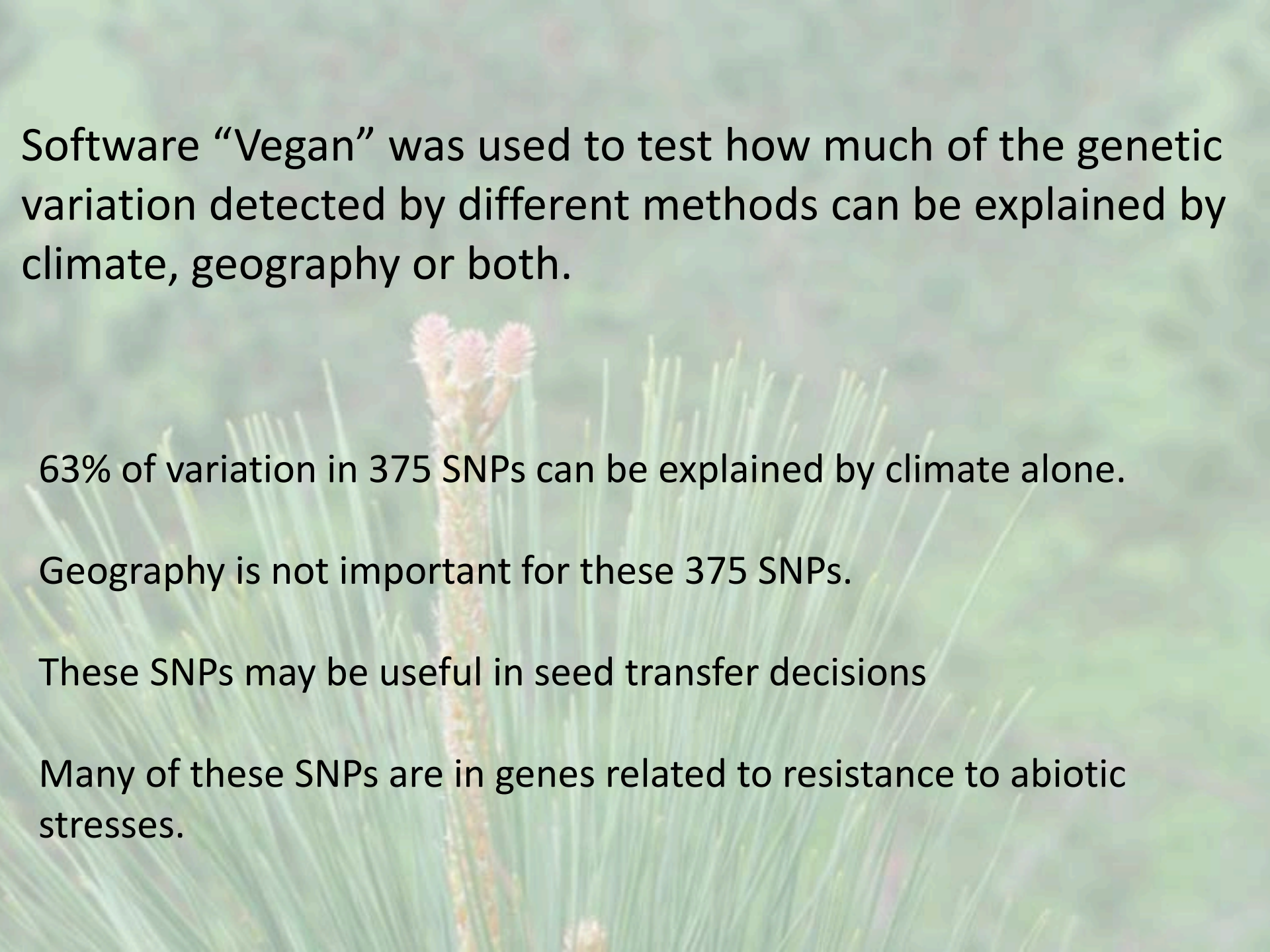
ABI1: Drought signaling

PtEMB2: Drought-related (LEA)



- 
- 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**





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

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

Geography is not important for these 375 SNPs.

These SNPs may be useful in seed transfer decisions

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. Keithanne 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.