Phenotyping and Exome Capture in a population of *Pinus taeda* L.

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

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

Abstract

Loblolly pine, *Pinus taeda* L., is one of the most widely planted, commercially and ecologically important tree species in North America. To develop guidelines for selecting loblolly pine trees that would be more resilient to changing climate, we are searching for alleles that are associated with environmental adaptation and stress mitigation traits in loblolly pine. We took an association genetics approach using 384 clonally replicated unrelated trees representing range wide natural populations and genome-wide array of single nucleotide polymorphisms (SNPs) to find their associations with phenotypic variation in total tree height, diameter, crown width, specific leaf area and branch angle after four growing seasons. To develop the current SNPs pool, we applied the NimbleGen SeqCap EZ method to capture exon-targeted genome sequences in loblolly pine. To design probes for exon target enrichment we used 199,723 exon regions (≈33Mb) that were annotated in the PineRefSeq project. The capture efficiency for targeted regions was 67-83%, and the mean sequencing coverage was ~58X. Based on preliminary analysis ~two million SNPs were discovered by GATK software.

Objectives

1) Genotype 384 trees in the ADEPT2 association population using an exome wide genotyping by sequencing approach.
2) Assess the productivity and adaptive traits of individual trees in the ADEPT2 population for association study.

Plant Material

Rooted cuttings from 384 unrelated trees (i.e., clones) sampled across a natural range of loblolly pine (ADEPT2 population, Fig. 1) were established at the Harrison Experimental Forest at Southern Institute of Forest Genetics (Saucier, MS) during the spring of 2010.

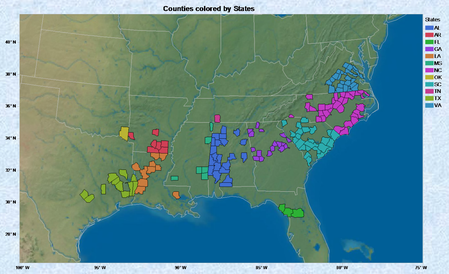

Fig.1. The counties of origin of the ADEPT2 trees planted at the Harrison Experimental Forest.

Fig.2. ADEPT2 trees planted at the Harrison Experimental Forest (This photo was taken on June,2014).

Acknowledgement

We thank Dr. David Neale, Dr. Jill Wegrzyn, Hans Vasquez-Grossfor and John Liechty providing the draft loblolly pine reference sequence, exon annotation and bioinformatics assistance; WSGI at Texas A&M University for providing computational resources and system administration support for the WSGI HPC Cluster; Dr. Charles Johnson, Dr. Joshua Hill, Dr. Richard Metz, Sabrina Moisan for providing sequencing service; Dr. Jason West, Dr. Jason Vogel, Dr. Fred Raley for help with phenotyping. Special gratitude goes to Nathalie Reilly and technicians in Harrison Experimental Forest at Southern Institute of Forest Genetics.

Phenotyping Methods and Results

The following traits were phenotyped:

① Total height (m);
② Diameter measured at the 18 inches high above the ground (m);
③ Crown width of the third major whorl from the top (m);
④ Branch angle (degree) measured as an average of three branch angles relative to level of the third major whorl from the top;
⑤ Specific leaf area (cm^2/g) calculated as twenty random needles' leaf area divided by the dry weight of these twenty needles from three bundles collected from the middle and south facing part of each tree;
⑥ Carbon isotope discrimination and nitrogen content (in progress).

Fig.3-6. Intern student Nathalie Reilly sponsored by PINEMAP Undergraduate Fellowship Program was working at the Harrison Experimental Forest to measure the phenotypic traits and collect the needle samples.

A mixed model analysis was used to assess the clonal effects for the measured traits.

Table 1. Summary of statistics for traits investigated.

Trait	Specific leaf area (cm^2/g)	Mean Branch Angle relative to level (degree)	Crown width (meter)	Total height (meter)	Diameter (meter)
Mean	28.88	35.43	1.40	3.25	0.05
Standard deviation	5.19	8.74	0.34	0.73	0.016
Minimum	14.11	9.3	0.43	0.85	0.005
Maximum	43.38	62.5	3.6	5.44	0.10
Median	29.06	35.97	1.39	3.3	0.05
N	920	915	916	922	922

Table 2. Clonal repeatability of the traits investigated.

Trait	Variance		Repeatability
	Clone	Residual	
Specific leaf area	2.71	3.02	0.47
Mean branch angle	24.79	1.36e-7	1.00
Crown width	0.0399	2.80e-10	1.00
Total height	0.16	0.07	0.70
Diameter	7.34e-5	2.58e-5	0.74

Table 3. Phenotypic (below diagonal) and Clonal (above-diagonal) correlations (r^2) among traits investigated. The cutoff *P* value is 0.05.

	Specific leaf area	Mean branch angle	Crown width	Total height	Diameter
Specific leaf area	--	Not significant	Not significant	Not significant	Not significant
Mean branch angle	Not significant	--	1.05e-2	Not significant	Not significant
Crown width	Not significant	Not significant	--	0.56	0.58
Total height	Not significant	Not significant	0.52	--	0.80
Diameter	Not significant	5.74e-3	0.53	0.82	--

Exome Capture Methods and Results

The NimbleGen SeqCap EZ system was applied to capture and enrich the selected sequences.199,723 exon regions from loblolly pine genome assembly and annotation v2 (generated in the USDA PineRefSeq Project and available on http://loblolly.ucdavis.edu/bipod/ftp/Genome_Data/genome/pinerefseq/) were used to design the probes. The size of target region mapped on reference genome v1.01(by PineRefSeq) was 33Mb. Individual genomic DNA libraries were constructed using barcoded DNA isolated from 375 trees. Then 10 trees were multiplexed and used together for exome capture and enrichment. Each multiplexed, captured and enriched library was sequenced in one lane of Illumina Hiseq2500v4.

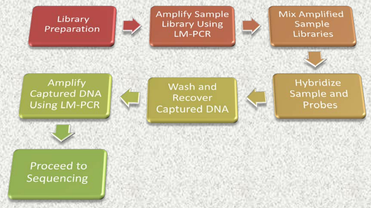

Fig.7. Workflow for SeqCap EZ Library experiments using Illumina sequencing instruments.

After sequencing, the bioinformatics pipeline was applied to analyze the capture efficiency and call SNPs. The following main results were obtained:

- mean mapping percentage is - > 99%;
- mean uniquely-mapping percentage - > 84%;
- on-target reads percentage of each sample - 67%-83%;
- mean on-target depth of coverage - 58X;
- mean percentage of bases with depth of coverage above 20X – 80%;
- mean percentage of non-captured bases on target regions - 2%.

The genotyping is still in progress. However, preliminary analysis showed ~2 million SNPs were discovered by GATK software (filtering conditions: the mapping quality of SNP≥30, ≥10X coverage in all the individuals and no insertion-deletion polymorphisms at the SNP loci in any individual).

Conclusions

1. High phenotypic variation exists in the ADEPT2 population.
2. NimbleGen capture is a very efficient method for the exome capture in loblolly pine. High genetic variation was observed that will be used to dissect complex climate change related traits.

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.

United States Department of Agriculture | National Institute of Food and Agriculture