Association Studies of Wood Development and Stress Response Genes in Loblolly Pine

Sreenath Palle[1], Candace Seeve[1], Andrew Eckert[2], Jill Wegrzyn[2], Patrick Cumbie[3], Barry Goldfarb[3], David Neale[2] & Carol Loopstra[1]

As part of the Plant Genome project, "Association Genetics of Natural Genetic Variation and Complex Traits in Pine", gene expression analysis of approximately 110 genes involved in xylem/wood development and 90 genes involved in stress response was done on an association population of loblolly pine (*Pinus taeda* L.). This population consists of 500 genotypes from across the natural range and was developed at North Carolina State University. Quantitative real-time PCR (qRT-PCR) was used to assay gene expression levels in 426 of these genotypes. Xylem development genes analyzed include those involved in cellulose biosynthesis, lignin biosynthesis, cell wall proteins, and genes involved in signal transduction in xylem tissue. Genes shown to be involved in drought and disease responses were included as stress response genes. The main objectives of this project are to analyze the selected genes by qRT-PCR to study gene expression differences among the different genotypes and to associate the gene expression data with single nucleotide polymorphism (SNP) genotypic data produced at the University of California-Davis. Positive associations were observed between SNPs and gene expression. Based on the results from this study, candidate genes may be further studied for association with phenotypic traits, used for development of molecular breeding tools, or included in future studies to further examine the molecular mechanisms of wood formation.

MATERIALS AND METHODS

Plant material: Three replicates of six-month-old loblolly pine rooted cuttings of 475 genotypes were obtained from the NCSU verification population. They were transplanted into pots and were grown for 4 more months in our green house. Stems, needles and roots were collected from each plant, frozen in liquid nitrogen and stored at -80°C. Some of the plants were lost due to *Fusarium* infection and at the time of harvesting there were 426 genotypes with at least two replicates.

Total RNA extraction and cDNA synthesis: Total RNA was extracted from the stems of all 426 genotypes (both replicates) using the method of Chang *et al*. (1993) except for an additional chlorofom extraction. The first strand cDNAs for each sample were made using random hexamers and Taqman reverse transcription reagents (Applied Biosystems, CA), following the manufacturer's recommendations.

Gene selection: Genes shown or hypothesized to be involved in lignin biosynthesis, cell wall development, xylem development and disease and drought responses were selected for the expression studies.

[1]Department of Ecosystem Science and Management, Molecular and Environmental Plant Sciences Program, Texas A&M University, College Station, TX 77843. [2]Department of Plant Sciences, University of California, Davis, CA 95616. [3]Department of Forestry and Environmental Resources, North Carolina State University, Raleigh, NC 27695.

Primer design: Bioinformatic approaches were used to identify putative orthologs in loblolly pine using the NCBI EST database (http://www.ncbi.nlm.nih.gov/blast/Blast.cgi) and the loblolly pine EST database at the University of Georgia (http://fungen.org/Projects/Pine/Pine.htm). Contigs were assembled from these EST sequences and gene specific primers were designed for qRT-PCR using Primer Express (Applied Biosystems).

Real-time quantitative RT-PCR: Transcript levels of the genes of interest were determined with qRT-PCR. Negative controls (no template, no reverse transcriptase) were used on some plates. Samples were run in duplicate on each plate using SYBR-Green PCR Master Mix (Applied Biosystems) on a GeneAmp 7900HT Sequence Detection System (Applied Biosystems), following the manufacturer's recommendations. SDS 2.3 software (Applied Biosystems) was used to collect the ΔΔCt values of all the genes for all the genotypes. Relative transcript levels for each sample were obtained using the 'relative standard curve method' (see User Bulletin #20 ABI PRISM 7900 Sequence Detection System for details), and was normalized to the transcript level of 18s rRNA or actin genes.

Association studies: The gene expression data was associated with SNP genotyping data (3939 SNPs), produced at UC-Davis, using Trait Analysis by aSSociation, Evolution and Linkage (TASSEL). Q-matrix was used to account for the structure in the population and General linear Model was used to do the association studies in TASSEL. Q-VALUE software was used to remove false positives.

RESULTS AND DISCUSSION

Gene expression differences usually ranging from 3-8 cycles were observed between the lowest and highest expression genotypes. The genes involved in lignin biosynthesis pathway showed 4-6 cycles of gene expression differences. These gene expression differences may be partially responsible for the phenotypic differences among genotypes. Correlations were sometimes observed between pairs of genes, suggesting their involvement in a pathway or co-regulation and co-expression (Fig.1).

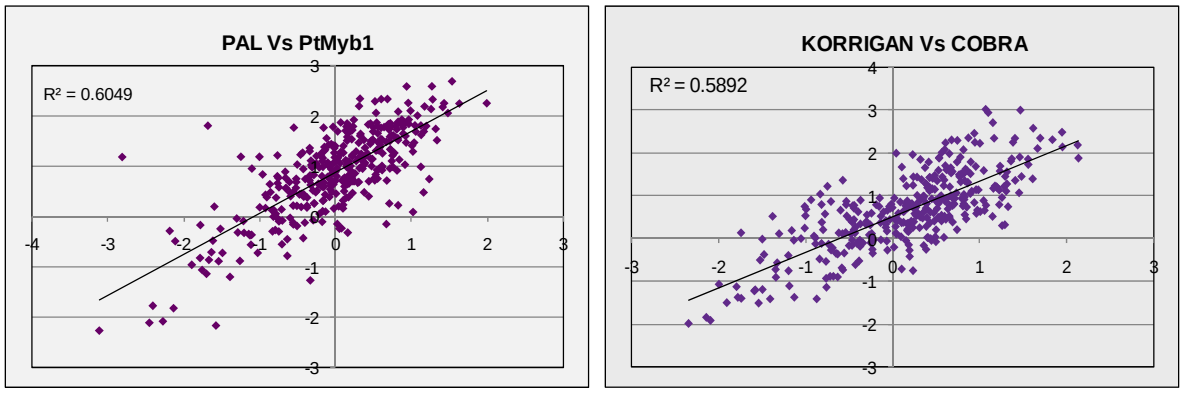

Fig.1 Correlations between ΔΔCT values of PAL Vs PtMyb1 and KORRIGAN Vs COBRA.

In order to dissect complex traits in forest trees, a candidate-gene-based association mapping strategy has been proposed by Neale and Savolainen (2004). The application of this association

mapping strategy in loblolly pine identified several SNPs from various genes that showed genetic association with gene expression data of genes thought to be involved in xylem development or disease and drought responses. Association studies resulted in the discovery of positive associations between the expression data of 80 genes and various SNPs. For example, expression of Laccase-5 has shown strong genetic association with a SNP in a gene homologous to Auxin response factor-11 (ARF-11) of *Arabidopsis thaliana*. Members of the ARF family are thought to be involved in vascular differentiation (Berleth *et al.*, 2000) and laccases are known to be involved in monolignol polymerization during lignin formation (Mayer and Staples 2002). As both of these genes are involved in xylem formation, the SNP in the ARF-11 homolog might have a direct or indirect effect on the expression of Laccase-5.

Some of the structural genes have been shown to be associated with SNPs in transcription factors and vice versa, suggesting that the transcription factor might be regulating the expression of that particular structural gene, either directly or indirectly. The gene expression data of a MADS box protein has shown strong association with a SNP in a gene homologous to the GRV2/ KATAMARI-2 heat shock protein of *A. thaliana*. Further analyses are being done to see if SNPs showing positive associations are synonymous or non-synonymous and whether they affect the protein conformation after translation.

Genes belonging to key pathways and those known to be involved in wood development and stress responses were evaluated in this association genetics study of loblolly pine. Ultimately it will be necessary to conduct association studies with virtually all the genes in the genome to have a complete understanding of the genetic architecture of a trait. Extension of the SNP discovery to the full-length DNA sequence, including promoter regions, is desirable because some genetic associations involving partially screened genes were not detected in this study due to the rapid decay of within-gene linkage disequilibrium in conifers.

REFERENCES

Berleth T, J Mattsson and CS Hardtke. 2000. Vascular continuity and auxin signals. Trends Plant Sci. 5: 387-393.

Chang S, J Puryear and JA Cairney. 1993. Simple and efficient method for isolating RNA from pine trees. Plant Mol Biol Rep. 11: 114–117.

Mayer AM and RC Staples. 2002. Laccase: new functions for an old enzyme. Phytochem. 60: 551-565.

Neale DB and O Savolainen. 2004. Association genetics of complex traits in conifers. Trends Plant Sci. 9: 325–330.