# TRANSCRIPT PROFILING OF DIFFERENTIATING XYLEM OF LOBLOLLY PINE (*Pinus taeda L.*)

A Dissertation

by

SUK HWAN YANG

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2004

Major Subject: Molecular and Environmental Plant Sciences

**TRANSCRIPT PROFILING OF DIFFERENTIATING XYLEM OF**

**LOBLOLLY PINE (*Pinus taeda L.*)**

A Dissertation

by

SUK HWAN YANG

Submitted to Texas A&M University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Approved as to style and content by:

| | |
|---|---|
| Carol A. Loopstra<br>(Chair of Committee) | Marla L. Binzel<br>(Member) |
| Jean H. Gould<br>(Member) | Keerti S. Rathore<br>(Member) |
| Charles T. Smith<br>(Head of Department) | Marla L. Binzel<br>(Chair of MEPS Faculty) |

December 2004

Major Subject: Molecular and Environmental Plant Sciences

# ABSTRACT

Transcript Profiling of Differentiating Xylem of Loblolly Pine (*Pinus taeda L.*).

(December 2004)

Suk Hwan Yang, B.S., Korea University;

M.S., Iowa State University

Chair of Advisory Committee: Dr. Carol A. Loopstra

Wood formation (xylogenesis) is a critical developmental process for all woody land plants. As an initial step to understand the molecular basis for temporal and spatial regulation of xylogenesis and the effect of the expression of individual genes on physical and chemical properties of wood, microarray and real-time RT-PCR analyses were performed to monitor gene expression during xylogenesis under various developmental and environmental conditions. The specific objectives established for this study were: Objective 1. Microarray analysis of genes preferentially expressed in differentiating xylem compared to other tissues of loblolly pine (see Chapter II); Objective 2. Microarray analysis of seasonal variation in gene expression for loblolly pines (*Pinus taeda L.*) from different geographical sources (see Chapter III); Objective 3. Real-time RT-PCR analysis of loblolly pine AGP and AGP-like genes (see Chapter IV). Based on the results from this study, candidate genes may be further studied for association with significant traits, used for genetic modification of wood properties, or included in future studies to further examine the molecular mechanisms of wood

formation.

This dissertation is dedicated to my parents,

Eun-Hee Koh and Doo-Hun Yang

## ACKNOWLEDGEMENTS

I sincerely thank my advisor, Dr. Carol Loopstra, for her help and guidance. She has been not only an advisor but also a great friend for me. I am also thankful to the other members of my graduate committee, Dr. Marla Binzel, Dr. Jean Gould, and Dr. Keerti Rathore for their valuable help and advice.

My special thanks goes to my family, especially to my lovely wife, Yoon-Jung. Without their love, patience, encouragement, and support through the years, I could not have made it this far.

This study was supported by U.S. National Science Foundation Award DBI-9975806 (P.I.: RR Sederoff, Genomics of Wood Formation In Loblolly Pine) and an award from the Texas Higher Education Coordinating Board Advanced Technology Program (000517-0119-2001). I thank Ernest Retzel and Martina Stromvik (Center for Computational Genomics and Bioinformatics, U. of Minnesota, MN) for their effort constructing the pine EST database (http://pinetree.ccgb.umn.edu/). I also thank members of Forest Biotechnology Group at NCSU for their effort on cDNA sequencing, kindly providing microarrays to us (special thanks to Dr. van Zyl), and providing advice and technical assistance.

I also thank Dr. Tom Byram and Dr. Larry Miller of Western Gulf Tree Improvement Program for their assistance and providing materials for this study.

# TABLE OF CONTENTS

CHAPTER Page

Page

# LIST OF FIGURES

# LIST OF TABLES

# CHAPTER I

## INTRODUCTION FOR THE TRANSCRIPT PROFILING OF DIFFERENTIATING XYLEM OF LOBLOLLY PINE (*Pinus taeda L.*)

Wood is one of the nation's most valuable industrial materials accounting for about 25% of the value of all major industrial materials (National Research Council, 1990). The value of delivered timber in the United States each year rivals or exceeds that of any other crop species including maize (McKeever and Howard, 1996). Loblolly pine is the most important tree species in the U.S. It is most widely planted each year and is a major source material for wood and pulp production in the U.S. (Moulton and Hernandez, 2000).

Wood is a model system to study plant cell wall biosynthesis. The cell wall is one of the significant differences between plant and animal cells and is thought to play important roles in the growth, development, adaptation and evolution of higher plants (Jones and Dangle, 1996). Pine wood is mostly composed of xylem tracheid cell walls and differentiating xylem is a rich source of RNAs and proteins involved in cell wall biosynthesis (Sederoff et al., 1994). Loblolly pine has become a model system to study wood formation (xylogenesis) in a gymnosperm. During the active growing season, the bark and the phloem layer of pine can be easily removed leaving the non-lignified

---

This dissertation follows the style of Plant Physiology.

differentiating xylem attached to the mature wood. Differentiating xylem can be separated by scraping the surface, while tracheids undergoing lignification and programmed cell death, as well as some mature tracheids and fully differentiated ray cells can be collected by deeper planing.

Wood is derived from the differentiation of two vascular cambial initials, fusiform initials and ray initials, which give rise to secondary phloem and xylem for translocation of nutrients and mechanical support. During xylogenesis, fusiform initials undergo several different steps including cell division, cell enlargement, cell wall thickening, lignification, and programmed cell death to become mature tracheids (Bailey, 1952). The primary wall is formed during cell enlargement. The thickness and composition of the wall in the subsequent stages is dependent on both the environmental and developmental factors. In other words, the properties of wood are determined by the activity of the genes and proteins expressed during xylogenesis and the variation in wood properties is caused by the regulation of these genes in response to developmental and environmental cues. The genes affecting wood properties such as wood specific gravity, microfibril angle, cell wall thickness, lumen diameter, fiber length, and chemical compositions of major cell wall components such as cellulose, lignin, and hemicellulose are of particular interest. These genes are the targets for genetic engineering of wood properties.

Trees are not an optimum experimental system due to their large size and long generation times. Pines have an additional disadvantage because of their large genome size, about 22 pg DNA per haploid genome (Wakamiya et al., 1996 ) which is about 200

times that of *Arabidopsis thaliana* (Somerville and Somerville, 1999). The pine genome is also redundant in repetitive DNA such as retrotransposons (Kriebel, 1985; Kamm et al., 1996; Kossack and Kinlaw, 1999). However, genomic approaches provide a new opportunity to overcome the disadvantages of pine as an experimental system because of the capability to work with large numbers of genes simultaneously. Now it is possible to analyze thousands of genes simultaneously using high throughput technologies such as microarray analysis to ask interesting biological questions that were previously impossible. The regulation, interactions, and roles in important biochemical pathways of those genes expressed during xylogenesis can be studied now.

Microarray analysis has already become a powerful tool for transcript profiling of conifers and poplars. Hertzberg et al. (2001) performed microarray analysis for different developmental stages of xylogenesis using a hybrid aspen unigene set consisting of about 3000 ESTs. Whetten et al. (2001) used the first-generation pine microarrays to study gene expression for juvenile vs. mature wood and compression vs. normal wood. Loblolly pine microarrays were used to study gene expression in different pine species (*Pinus sylvestris*) and Norway spruce (*Picea abies*) as well (van Zyl et al., 2002). Stress responses in pine were studied using microarrays containing 384 cDNAs (Heath et al., 2002) and the effects of PEG (polyethylene glycol) on gene expression during white spruce somatic embryogenesis was investigated using loblolly pine microarrays containing 2178 cDNAs (Stasolla et al., 2003). The effect of PEG on the transcript level of 512 stress-related genes during white spruce somatic embryogenesis was analyzed using loblolly pine cDNA microarrays as well (Stasolla et al., 2003).

One of the problems of cDNA microarraies is the cross-hybridization among different members of the multigene family. Complex gene families are relatively common in pine (Kinlaw and Neale, 1997) and each member of a multigene family could show a different pattern of gene expression in different environments and function differently. However, gene specific primers can be designed and real-time RT-PCR can be performed to examine the expression pattern of each member of a multigene family. Real-time RT-PCR analysis is a valuable tool to verify the overall validity of the microarray data as well.

There have been numerous studies to understand the mechanisms underlying xylogenesis (reviewed by Roberts and McCann, 2000; Chaffey, 2002). However, our understanding of wood formation at the molecular level is still elementary. Thus, the ultimate goal of this study is to understand how genes involved in wood formation eventually influence the final phenotype of wood. To achieve this goal, it is important to understand how genes are regulated and how they interact with each other during xylogenesis. As an initial step to understand the molecular basis for temporal and spatial regulation of xylogenesis and the effect of the expression of individual genes on physical and chemical properties of wood, microarray and real-time RT-PCR analyses were performed to monitor gene expression during xylogenesis for various developmental and environmental conditions. The specific objectives established for this study were: Objective 1. Microarray analysis of genes preferentially expressed in differentiating xylem compared to other tissues of loblolly pine (see Chapter II); Objective 2. Microarray analysis of seasonal variation in gene expression for loblolly

pines (*Pinus taeda L.*) from different geographical sources (see Chapter III); Objective 3. Real-time RT-PCR analysis of loblolly pine AGP and AGP-like genes (see Chapter IV).

# CHAPTER II

# MICROARRAY ANALYSIS OF GENES PREFERENTIALLY EXPRESSED IN DIFFERENTIATING XYLEM OF LOBLOLLY PINE (*Pinus taeda L.* )

## INTRODUCTION

Wood formation (xylogenesis) is a critical developmental process for all woody land plants for mechanical support and transport of water and minerals. A significant portion of the solar energy absorbed by the earth is stored in the form of cellulose in wood, making it the most abundant raw material in the world. Wood is economically important as well because it is a major building material and a source material for pulp production. In addition, wood is a specialized model to study plant cell wall biosynthesis (Jones and Dangle, 1996) and differentiating xylem is a rich source for DNA, RNA and proteins involved in cell wall biosynthesis (Sederoff et al., 1994).

Wood formation begins at a highly specialized tissue called the vascular cambium and includes five steps: cell division, cell enlargement, cell wall thickening, lignification and programmed cell death (Bailey, 1952). In each step, the expression of specific sets of genes is tightly controlled by environmental and developmental cues (Roberts and McCann, 2000; Hertzberg et al., 2001; Mellorowicz et al., 2001; Whetten et al., 2001) influencing the composition and morphology of the xylem cell wall which is the most important factor for determining wood properties (Megraw, 1985).

There have been numerous studies to understand the mechanisms underlying xylogenesis. However, our understanding of wood formation at the molecular level is still elementary. Our ultimate goal is to understand how genes involved in wood formation eventually influence the final phenotype of wood. To achieve this goal, it is important to understand how genes are regulated by developmental and environmental factors during xylogenesis. In addition, understanding the molecular basis of wood formation will accelerate the improvement of wood properties by specifically targeting genes affecting useful traits such as specific gravity, microfibril angle or content of cell wall components by genetic modification. It will also be useful for traditional breeding by using a candidate gene approach for marker-assisted selection.

Thanks to recently emerging high-throughput techniques such as large scale EST sequencing (Allona et al., 1998; Sterky et al., 1998; Whetten et al., 2001), microarray analysis (Hertzberg et al., 2001; Whetten et al., 2001; Heath et al., 2002; van Zyl et al., 2002), 2-D PAGE (Costa et al., 1999; Mijnsbrugge et al., 2000), and SAGE (Lorenz and Dean, 2002), it is now possible to better understand the molecular basis of wood formation on a genome wide scale. As an initial step to understand the molecular basis for temporal and spatial regulation of xylogenesis and the effect of the expression of individual genes on physical and chemical properties of wood, we compared transcript profiles of differentiating loblolly pine xylem with those of needles, megagametophytes, and embryos. A microarray containing approximately 1500 ESTs with putative functions of interest, selected from several loblolly pine xylem partial cDNA libraries was made. Genes preferentially expressed in xylem were identified and real-time RT-PCR was used

to test the overall validity of the microarray data.

## MATERIALS AND METHODS

### Plant materials

Loblolly pine (*Pinus taeda L.*, Texas source S6PT7) seeds were surface sterilized using a 20% bleach solution for 30 minutes followed by three washes with dd$H_2O$. After being nicked at the end, pine seeds were stratified in 1% $H_2O_2$ for 3 weeks. The 1% $H_2O_2$ was replaced everyday during the stratification. After embryo radicles reached several millimeters, the embryo radical and megagametophyte were separated and frozen in liquid nitrogen. Newly differentiating xylem was isolated from fast growing loblolly pines approximately 12 to 15 years old. First, the bark and phloem layers were peeled off and differentiating xylem tissue was collected and frozen immediately in liquid nitrogen. Needles were collected from 5 month-old loblolly pine seedlings growing in the greenhouse.

### Microarray preparation

1500 ESTs with putative functions of interest were selected from normal wood, compression wood, late wood and planings (deeper cell layers) xylem cDNA libraries made from loblolly pine (Whetten et al., 2001; Kirst et al., 2003). Putative classifications (cell wall-related, stress-related, etc) were made based on blast searches. Clones were inoculated into 96-well blocks containing 1.3 ml magnificent broth (MB) and 100 μg/ml

ampicillin and were incubated for 16 hrs at 37°C in a shaker incubator (220 rpm). Plasmids were isolated using a R.E.A.L. Prep 96 Plasmid Kit (QIAGEN, Maryland) or a Qiaprep 96 Turbo Miniprep Kit. After the isolation, plasmids stocks in 96-well plates were dried and redissolved in 10 μl molecular grade water. A dilution containing 1 μl of plasmid stock and 99 μl $ddH_2O$ was vortexed briefly and stored at –20 °C until needed. The cDNA inserts were PCR amplified using M13 forward and M13 reverse universal primers. PCR was performed in a 50 μl reaction containing 39.1 μl $ddH_2O$, 5 μl 10x PCR buffer, 1 μl dNTPs (10mM each), 1 μl M13 forward primer (10 μM), 1 μl M13 reverse primer (10 μM), 0.4 μl TAQ polymerase (5U/μl), and 2.5 μl of dilute plasmid. The PCR conditions were 1 minute of pre-denaturation at 94 °C, 35 cycles of 30 seconds at 94 °C, 1 min at 57 °C, and 4 minutes at 72°C, followed by a 10 minutes final extension at 72 °C. The quantity and the quality of the amplified PCR products were checked on a 1% agarose gel. PCR products were purified with 96-well multiscreen filter plates (Millipore Corp. MA) following the manufacturer's recommendation. Purified cDNA fragments (40 μl final volume) were transferred to a 96-well plate and an equal volume of DMSO was added to each well, vortexed for 30 seconds and spun down. Twenty μl of each spotting solution was transferred to 384-well plates and sealed and stored at -20 °C until needed. The cDNA fragments were spotted onto Corning CMT-GAPS™ coated slides (Corning, NY) using a GMS 417 Arrayer (Genetic MicroSystems, MA), with a center to center spacing of 285 μm, following the manufacturer's recommendations. The resulting arrays contained 6144 spots. After the

printing, DNA fragments were UV cross-linked at 250mJ, baked at 75°C for 2 hrs and stored in slide racks in the dark until needed.

**Probe synthesis and hybridization**

Total RNAs were extracted from differentiating xylem, developing embryos, megagametophytes and needles using the method of Chang et al. (1993). Each RNA was treated with *DNaseI* after the extraction to make sure no residual DNA remained. For the first strand cDNA synthesis, 2 µl of oligo dT (500ng/µl) and dNTPs at 10mM were added to 40 µg of total RNA from each tissue. $ddH_2O$ was added bringing the total volume of each reaction up to 24 µl. After an incubation at 65°C for 5 minutes, chilling on ice for 1 min and a brief centrifugation, 8 µl of first strand buffer, 4 µl 0.1 M DTT and 2 µl of RNase Out (40U/µl, Ambion) were added. Following a gentle mixing, a brief spin, and incubation at 42°C for 2 min, Superscript II reverse transcriptase (1µl of 200U/µl, Invitrogen) was added, the reaction was mixed again gently with a pipette, and incubated at 42°C for 50 minutes. After heat inactivation of the Superscript II at 70°C for 15 min, the reaction was collected at the bottom of the tube by a quick spin and put on ice. The first strand cDNAs were precipitated by adding an equal volume of 2-propanol. After washing with 70% ethanol, the pellets were dried in a speed vac and dissolved in 53 µl $ddH_2O$. The cDNAs were denatured at 95°C for 5 min and placed on ice. The rest of the procedure was performed in minimal light due to the light sensitivity of Cy3 and Cy5. Twenty µl of nucleotide labeling mix (1200 µM dATP, dGTP and dCTP each; 150 µM dTTP, 150 µM Cy3-dUTP or Cy5-dUTP), 10 µl of reaction 2 buffer

(Invitrogen) and 2 μl of Klenow (5U/μl) were added, mixed well and placed on ice. The reaction was then incubated in the dark at 37°C for 1 hr. The Cy3 and Cy5 labeled probes were combined and cleaned up using a QIAquick PCR Purification Kit (Qiagen) following the manufacturer's recommendations. After the final elution, probes were dried in a speed vac for 60 min in the dark. Each dried probe was redissolved in 30 μl of hybridization buffer containing 0.5% SDS, 5x SSC, 5x Denhardt's, 50% formamide, 500ng poly A RNA, and 500ng calf thymus DNA. After a denaturation at 95°C for 2 min and a brief centrifugation, hybridization solution was added to the microarray slides and a cover slip was placed on the top. Slides were then placed in a Corning incubation chamber (Corning, NY) and 20 μl water was added to reduce any evaporation during the hybridization. The hybridization was performed in a 42°C water bath for 20 hrs in the dark. After hybridization, the slides were transferred to a disposable tube containing 50 ml of pre-heated (42°C) wash 1 solution (1x SSC, 0.2% SDS). The slides were washed twice in wash 2 solution (0.1x SSC, 0.2% SDS), and three times in wash 3 solution (0.1 % SDS) all in a shaking 42°C incubator for 4 min. The slides were transferred to a slide rack and dried by centrifuging at 500 rpm for 5 min. All the slides were scanned within one hour.

**Image analysis, quantification and data analysis**

After the washing steps, the slides were scanned with both channels using a ScanArray 3000 (GSI Lumonics, Watertown, MA) at a 10 μm resolution. The laser power and photo-multiplier tube setting were adjusted to approximately balance both

channels while scanning. The intensities of each spot for both channels were quantified using QuantArray (GSI Lumonics, Watertown, MA). For the microarray data analysis, we adopted the fixed effect ANOVA model of Kerr and Churchill (2001a) using MA-ANOVA, a software package for the analysis of spotted cDNA microarray experiments by Wu, Kerr, Cui and Churchill (http:\\www.jax.org\research\Churchill). Measured spot intensities for each of the channels for each of the tissue comparisons were pre-normalized following the method of Yang et al. (2001) using the "malowess" function in MA-ANOVA. An ANOVA analysis using the "fitmaanova" function and a permutation F-test for the null hypothesis assuming zero VG effect were performed for each gene using the "make_Ftest" function in MA-ANOVA (2001b). The permutation F-test was performed twice using both a less conservative p-value (p=0.01) and a more conservative p-value (p=0.001) to reduce the number of false positives. Further data analysis was also performed using GeneSpring (Silicon Genetics, CA).

**Real-time quantitative RT-PCR**

To test the overall validity of our microarray data, clones with values ranging from the top to the bottom xylem vs megagametophyte log ratio were selected and analyzed by real-time quantitative RT-PCR (Heid et al., 1996). Gene specific primers for the selected clones were designed based on cDNA sequences from the NSF Pine EST Database (http://web.ahc.umn.edu/biodata/nsfpine/) using Primer Express (Applied Biosystems, CA). Real-time RT-PCR was repeated three times for each of the selected clones with SYBR-Green PCR Master Mix (Applied Biosystems, CA) on a GeneAmp

7900 Sequence Detection System (Applied Biosystems, CA) following the manufacturer's recommendations. A dissociation curve was used to identify a temperature where only amplicon, and not primer dimers, accounted for SYBR-bound fluorescence. All data were normalized to internal loblolly pine 18S rRNA quantities ($\Delta$-$C_T$) and the $\Delta$-$C_T$ value for the xylem sample was used as the calibrator to the $\Delta$-$C_T$ values for other tissue samples ($\Delta\Delta$-$C_T$ analysis, see User Bulletin #2" ABI PRISM 7900 Sequence Detection System, pp11-15 for details).

## RESULTS AND DISCUSSION

### A microarray from differentiating xylem

To identify genes preferentially expressed in differentiating xylem tissue, a microarray containing an approximately 1500-clone unigene set (list is available under Supplementary Data - 2004 Plant Science at http://forestry.tamu.edu/newweb/people/faculty&staff/c-loopstra) was constructed. As part of a NSF pine genomics project, several partial cDNA libraries were made at North Carolina State University using differentiating xylem tissues from loblolly pine including normal wood, compression wood, latewood and planings (deeper cell layer) (Whetten et al., 2001; Kirst et al., 2003). These cDNAs were partially sequenced from the 5' ends and overlapping sequences were assembled into contigs representing unigene sequences. After Blastx analysis of sequences of each contig, approximately 1500 EST contigs with putative functions of interest were selected to be included on the array. The selected contigs

Table 1. Functional classification of genes preferentially expressed in xylem (p-value=0.01) compared to the functional classification of total genes on the array.

| Functional Class | Xylem Preferential Genes | Total Genes On the | *Percent preferentially |
|---|---|---|---|
| cell wall-related | 66 (20%) | 153 (10%) | 43 |
| no hit | 52 (15%) | 212 (15%) | 24 |
| intermediate metabolism | 48 (14%) | 250 (17%) | 19 |
| translation | 22 (7%) | 154 (7%) | 14 |
| other | 22 (7%) | 148 (7%) | 15 |
| DNA/RNA-binding | 19 (6%) | 106 (6%) | 18 |
| stress-related | 18 (5%) | 43 (5%) | 42 |
| cytoskeleton | 16 (5%) | 32 (2%) | 50 |
| signal transduction | 14 (4%) | 56 (4%) | 25 |
| transport | 12 (4%) | 68 (5%) | 18 |
| inducible | 12 (4%) | 37 (3%) | 32 |
| methionine biosynthesis | 11 (3%) | 37 (3%) | 30 |
| turn over | 7 (2%) | 62 (4%) | 30 |
| lipid metabolism | 5 (1%) | 30 (2%) | 17 |
| disease-related | 5 (1%) | 23 (2%) | 22 |
| e-transfer | 4 (1%) | 53 (4%) | 7 |
| surface proteins | 4 (1%) | 13 (1%) | 30 |

* Percentage of genes on the array that are preferentially expressed in xylem (last column).

represented cell wall proteins, proteins for intermediate metabolism, stress-related proteins, DNA-RNA binding proteins, hormone responsive proteins, disease responsive proteins, transporters, and proteins for lipid metabolism (Table 1). In addition to contigs with putative functions of interest, contigs without any putative function assigned ("no hits" after a Blastx search using an E-value cutoff of $10^{-5}$) were included on the array as well. To reduce nonspecific hybridization, EST clones located near the 3' end of each contig and larger than 200 bp were selected to be amplified by PCR. For the negative controls, no templates were added during the PCR reaction included on the array as well.

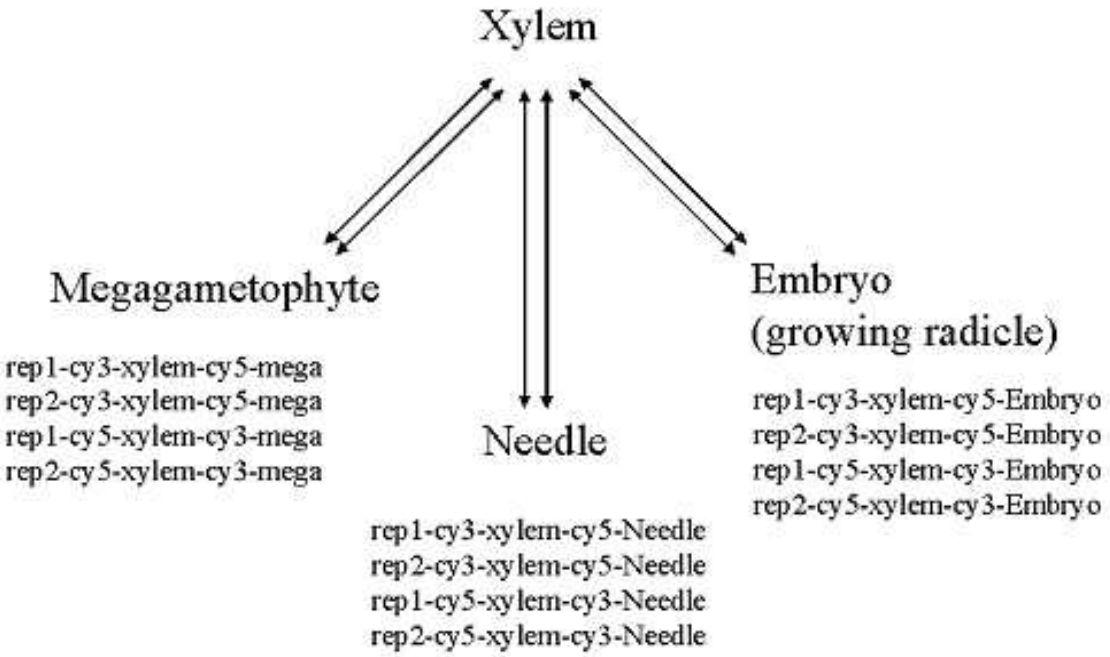


Figure 1. Double dye-flip reference design

To reduce nonspecific hybridization, EST clones located near the 3' end of each contig and larger than 200 bp were selected to be amplified by PCR. For the negative controls, no templates were added during the PCR reaction making blank spots on the array. Each clone was spotted four times on each slide producing an array containing 6144 spots.

**Identification of genes preferentially expressed in xylem**

Lee et al. (2000) pointed out that there is inherent variability in most microarray data even after removing systematic sources of variation. Thus, it is critical to have an experimental design with sufficient replications to control this inherent noise and to produce consistent and reliable results (Lee et al., 2000; Kerr and Churchill 2001a,b). In this study, each of the ESTs was spotted four times on the array and a double dye-flip reference experimental design (Figure 1) was adopted providing 32 data points for each of the ESTs for each of the tissue comparisons. The fixed effect ANOVA model by Kerr and Churchill (2001a) was used for microarray data analysis. One of the advantages of using this ANOVA approach is that normalization of the data for various factors such as array, dye, variety, gene effects, and their interaction effects on the signal is an integral part of the data analysis. However, often more elaborate pre-transformation of the data is necessary before ANOVA analysis to meet a fundamental assumption of ANOVA, an existence of a scale on which various effects are additive (Kerr and Churchill, 2001a). Thus, we adopted the lowess adjustment method of Yang et al. (2001) to smooth the data before ANOVA analysis. Figure 2 shows one example of pre-normalization. For each of the spots on the array, the log2(R/G) (red/green) value is plotted against the log2(R*G) value. If there is no bias between the two different dyes and no differential gene expression in each of the tissues, the spots should reside on a zero horizontal line. However, a curvature at the low end in the upper plot (below 22 on the X-axis) is evident. This curvature is straightened by using the method of Yang et al. (2001) using the “malowess” function in MA-ANOVA (lower plot). After pre-normalization, the

majority of the genes reside close to the zero horizontal line indicating that most of the genes expressed in the differentiating xylem are not tissue-specific (Figure 2). However, we could identify outliers that are far from the zero horizontal line indicating that these genes are differentially expressed in one of the tissues. We also checked the normality in residual distribution for the model adopted in each of the tissue comparisons and found no significant evidence against homoscedasticity in each of the residual plots (data not shown).

Differentially expressed genes were identified by hypothesis testing using two different models, an alternative model and a null model. During the permutation F-test, both the less conservative p-value (0.01) and more conservative p-value (0.001) were used. Using a p-value=0.01, 440 genes in the xylem vs. megagametophyte, 401 genes in the xylem vs. embryo and 191 genes in the xylem vs. needle tissue comparisons were identified as differentially expressed genes (gene lists are available under Supplementary Data - 2004 Plant Science at http://forestry.tamu.edu/newweb/people/faculty&staff/c-loopstra). However, using a p-value=0.001, these numbers were reduced to 104 genes in the xylem vs. megagametophyte, 57 genes in the xylem vs. embryo and 30 genes in the xylem vs. needle tissue comparisons. Because we are interested in identifying genes with preferential expression in xylem, a Venn diagram was generated to show the number of genes preferentially expressed in xylem among differentially expressed genes identified in each of the tissue comparisons (Figure 3). Using a p-value=0.01, 204 genes in the xylem vs. megagametophyte, 178 genes in the xylem vs. embryo and 85 genes in the xylem vs. needle tissue comparisons were identified as genes preferentially expressed in

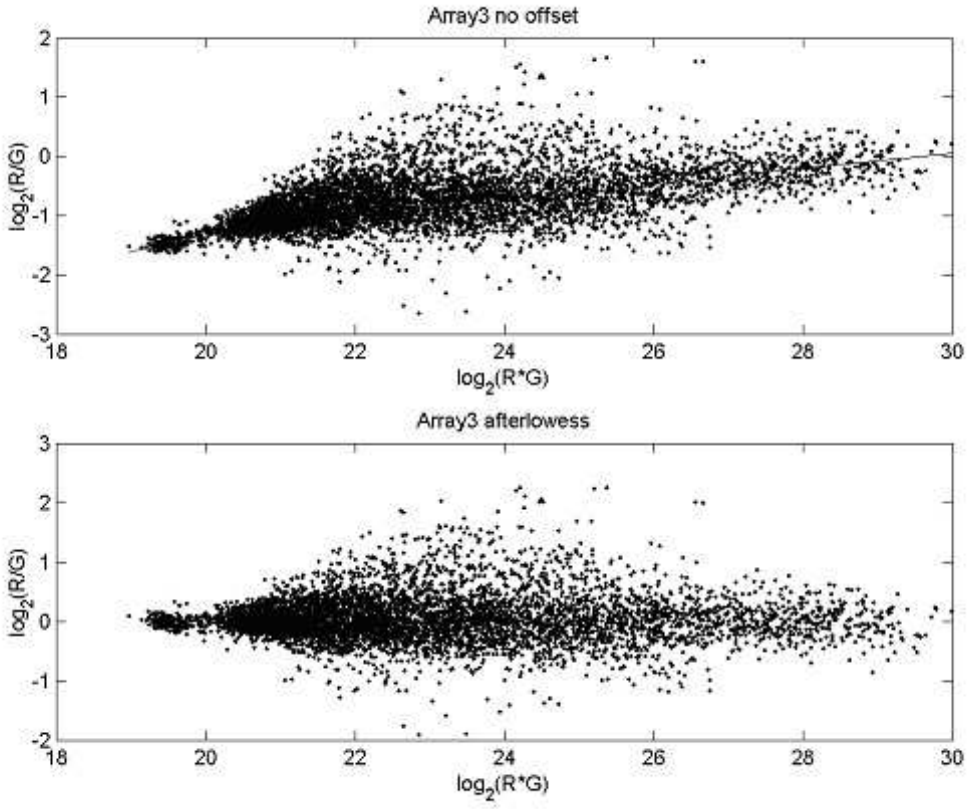


Figure 2. For each spot on the array, the $\log_2(R/G)$ value is plotted against the $\log_2(R*G)$ value. If there is no bias in labeling of the two dyes and no differential gene expression in each tissue, most spots should reside in a zero horizontal line. However, the curvature at the low end in the upper plot (below 22 on the X-axis) is evident. This curvature is straightened by using the method of Yang et al. (2001) using the "malowess" function in MA-ANOVA (lower plot).

xylem. Xylem is the most extensively vascularized tissue examined, whereas megagametophyte was the only non-vascularized tissue examined. Thus, it was expected that xylem and megagametophyte would be the most distinct of all the pairwise comparisons. Many genes preferentially expressed in xylem compared to one tissue also

showed the same preferential expression pattern when compared to the other tissues (overlapping regions) and 30 genes (highlighted in red in the supplementary data) were preferentially expressed in xylem in all three comparisons (Figure 3 A). Using a p-value=0.001, 56 genes in the xylem vs. megagametophyte, 36 genes in the xylem vs. embryo and 16 genes in the xylem vs. needle tissue comparison were identified as genes preferentially expressed in xylem (http://forestry.tamu.edu/newweb/people/faculty&staff /c-loopstra). There was only one gene (NXNV_149_F10) preferentially expressed in xylem in all three comparisons (Figure 3 B) and there was no putative function assigned for it ("no hits" after a Blastx search using an E-value cutoff of $10^{-5}$). A tblastX search of the NR database revealed it is similar to a gene down-regulated by drought stress in *Pinus pinaster* (E-value=2e-08). The genes that consistently showed preferential expression in xylem in all comparisons are strong candidates for having important roles in xylogenesis.

**Functional classification and characterization of xylem preferential genes**

Genes identified as preferentially expressed in xylem in this study were classified by functional groups and compared to the functional class of the total ESTs on the array (Table 1). In general, percentages of classes of xylem preferential genes reflected the percentages of the array. However, there was a significant increase in the percentage of genes that are preferentially expressed in xylem that are cell wall-related (20%) compared to the percentage on the array (10%). Approximately 43% of the cell wall genes represented on the array are preferentially expressed in xylem (Table 1). The high

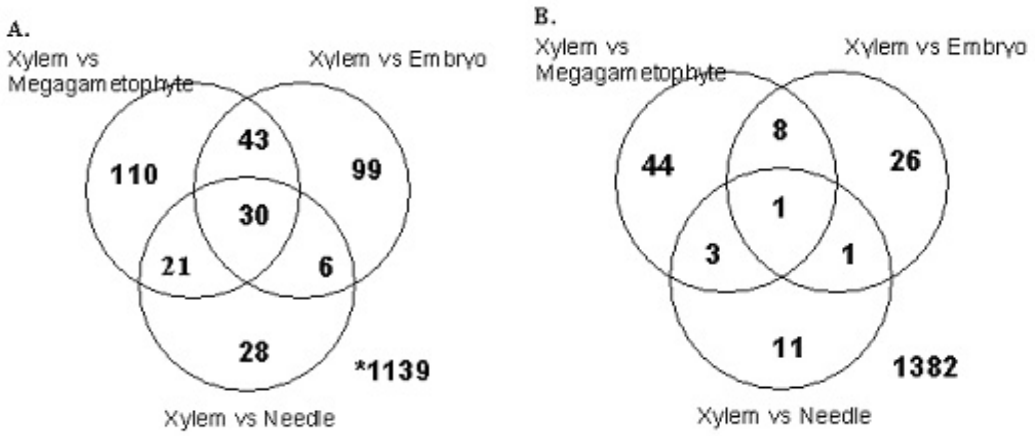


Figure 3. Venn diagrams of the numbers of overlapping and non-overlapping genes preferentially expressed in xylem compared to other tissues. Genes were identified using a p-value=0.01 (A) and a p-value=0.001 (B) after ANOVA analysis and permutation F-tests from each of the tissue comparisons. * Numbers out side of the Venn diagrams indicate the number of genes that are not included on any of the categories.

number of cell wall-related genes in our selected gene list confirms the overall reliability of the current microarray data because we expect some cell wall proteins to be highly expressed in xylem. Other functional classes of which the percentage of the xylem preferential genes was increased were “cytoskeleton” and “inducible” (Table 1).

Genes for the cellulose synthase complex were expected to be preferentially expressed in xylem. According to the current model (Delmer and Amor, 1995), cellulose microfibrils are synthesized on the plasma membrane by a cellulose synthase complex, which is composed of multiple cellulose synthase subunits (particle rosettes) and sucrose synthase. Previous studies showed that transcripts for these genes are highly abundant in wood forming tissues of both gymnosperms and angiosperms (Allona et al., 1998;

Sterky et al., 1998; Whetten et al., 2001) and their expression was coordinated with late cell expansion and secondary wall formation (Hertzberg et al., 2001). There are multiple members of the cellulose synthase gene family and they are differentially regulated (Allona et al., 1998; Sterky et al., 1998; Wu et al., 2000). The xylem loblolly pine EST database contains nine different contigs for cellulose synthase (contigs 747, 2303, 3052, 4859, 5673, 6822, 7800, 7822 and 7904 in http://web.ahc.umn.edu/biodata/nsfpine/contig_dir16/). According to Whetten et al. (2001), EST clones with more than 98% sequence similarity are sometimes placed into different contigs when PHRAP (http://www.phrap.org) paramers used for contig assembly are a minimum mismatch of 40 and a minimum score of 80. PHRAP parameters used for contig assembly for the xylem loblolly pine EST database (http://web.ahc.umn.edu/biodata/nsfpine/contig_dir16/) were a minimum match of 50, a minimum score of 100 and a minimum length of 100. Kinlaw and Neale (1997) suggested that complex gene families are relatively common in pine. Thus, these different contigs for cellulose synthase could represent allelic variation or different members of the loblolly pine cellulose synthase gene family. Our array contained clones that represent six cellulose synthase contigs (contigs 4859, 5673, 6822, 7800, 7822 and 7904). Among these, only contig 7822 was preferentially expressed in xylem for all three comparisons ($p=0.01$). Contig 7800 was preferentially expressed in xylem compared to megagametophytes but not in other two comparisons. Contig 7904 was preferentially expressed in xylem compared to needles but not in other two comparisons. This result indicates that different members of the loblolly pine cellulose synthase gene family are also differentially regulated. Sucrose synthase is responsible

for channeling UDP-glucose into the growing cellulose chain (Delmer and Amor, 1995). This array contained clones that represent three sucrose synthase contigs (contigs 7783, 3756 and 3167). Contig 7783 was preferentially expressed in xylem compared to embryos but not in the other two comparisons (p=0.01). Contig 3756 was preferentially expressed in xylem compared to megagametophytes but not in the other two comparisons.

Previous studies (Allona et al., 1998; Sterky et al., 1998; Whetten et al., 2001) indicate that among the most abundant transcripts in wood forming tissues are the genes for cell wall re-structuring and expansion such as expansins, XETs (xyloglucan endo-transglycosylases) and endo-glucanases. Expansins cause wall creep by loosening hydrogen bonding between cellulose microfibrils and glycan cross-linkers (Cosgrove, 1997). The loblolly pine xylem EST database contains 11 different contigs for expansin (contigs 838, 2753, 4335, 4844, 6512, 6603, 6774, 7754, 7859, 7879 and 7897 in http://web.ahc.umn.edu/biodata/nsfpine/contig_dir16/). Our array contained clones that represent three expansin contigs (contigs 4844, 6512 and 7754). Among these, only contig 7754 was preferentially expressed in xylem for all three comparisons (p=0.01). XETs cut and stitch xyloglucan polymers into new configurations (Smith and Fry, 1991). Four different members of the loblolly pine XET family were identified previously and their expression was differentially regulated (Allona et al., 1998). We identified 32 different contigs for XET in the loblolly pine xylem EST database and our array contained clones that represent seven XET contigs (contigs 2807, 4609, 5809, 7263, 7313, 7209 and 8012). Among these, only contig 7709 was preferentially

expressed in xylem for all three comparisons (p=0.01). Contigs 7313 and 5809 were preferentially expressed in xylem compared to embryos but not in the other two comparisons. Contig 7263 was preferentially expressed in xylem compared to megagametophytes and embryos but not compared to needles. We also identified 59 different contigs for glucanases in the loblolly pine xylem EST database and our array contained clones that represent four glucanase contigs (contigs 5350, 5383, 7176 and 7518). Among these, only contig 7518 was preferentially expressed in xylem compared to embryos but not in the other two comparisons (p=0.01).

Previous studies also showed that transcripts for genes involved in the lignin biosynthetic pathway are among the most abundant transcripts in wood forming tissues (Allona et al., 1998; Sterky et al., 1998; Whetten et al., 2001; Lorenz and Dean, 2002). In some cases, expression is up-regulated in compression wood compared to normal wood (Whetten et al., 2001) and different members of gene families are differentially regulated during wood formation (Hertzberg et al., 2001). Many genes involved in lignin biosynthesis were also identified as preferentially expressed in xylem compared to other tissues in this study. Laccase is a diphenol oxidase involved in monolignol polymerization (Bao et al., 1993). Our array contained clones that represent 12 contigs for laccase (contigs 305, 4929, 4944, 5681, 5931, 6071, 6127, 6980, 7139, 7467, 7582 and 8024 in http://web.ahc.umn.edu/biodata/nsfpine/contig_dir16/). Among these, only contig 6071 was preferentially expressed in xylem in all three comparisons (p=0.01). Contig 305 was preferentially expressed in xylem compared to megagametophytes but not in the other two comparisons. Contigs4944, 5681 and 8024 were preferentially

expressed in xylem compared to megagametophytes and embryos but not compared to needles. Contigs5931 and 6980 were preferentially expressed in xylem compared to megagametophytes and needles but not compared to embryos. This result also indicates that members of the loblolly pine laccase gene family are differentially regulated during wood formation.

Transcripts for cell wall structural proteins such as arabinogalactan-proteins (AGPs), glycine-rich proteins and proline-rich proteins are among the most abundant transcripts in wood forming tissues and are expected to be preferentially expressed in differentiating xylem tissue compared to other tissues (Loopstra and Sederoff, 1995; Allona et al., 1998; Sterky et al., 1998; Loopstra et al., 2000; Zhang et al., 2000; Whetten et al., 2001; Lorenz and Dean, 2002). AGPs are highly glycosylated proteoglycans or glycoproteins (2-10% protein by weight) which contain highly repetitive sequences and are frequently rich in proline/hydroxyproline, serine, threonine, alanine, and glycine. General features of AGPs include a signal peptide for targeting to the endoplasmic reticulum, a hydroxyproline-rich domain and a GPI anchor site for plasma membrane attachment. Six different loblolly pine AGP genes encoding AGP-like proteins have been identified (Loopstra et al., 2000; Zhang et al., 2000) and they are among the most abundant transcripts in differentiating xylem tissue (Allona et al., 1998; Whetten et al., 2001; Lorenz and Dean, 2002). The loblolly pine xylem EST database (http://web.ahc.umn.edu/biodata/nsfpine/contig_dir16/) contains numerous contigs for different AGPs (contigs 7942, 8039 and 8025 for AGP6; contigs 459, 1204, 6663, 6586, 7600, 7711, 6466 and 7834 for AGP5; contigs 7422, 7945 and 7948 for AGP4; contigs

16, 269, 300, 7679, 7869 and 7984 for 3H6, contigs 7693, 8006 and 8022 for 14A9; contigs 6333, 7444, 7704, 7934 and 8019 for AGP3). Our array contained one clone for AGP6 (contig 8025), four clones for AGP5 (contigs 1204, 6663, 7711 and 7834), one clone for 3H6 (contig 7869), two clones for 14A9 (contigs7693 and 8022) and three clones for AGP3 (contigs 6333, 7444 and 8019). Among these, only contig 8025 for AGP6 and contig 8019 for AGP3 showed preferential expression for all three comparisons ($p=0.01$) but most of the other contigs for different AGPs were also preferentially expressed in xylem compared to other tissues (see supplementary data for details).

Other cell wall related genes that were very strongly preferentially expressed in xylem include a basic blue protein, UDP-glucuronosyl transferase, and lp6 from loblolly pine. Interestingly, some of the genes with unknown function ("no hits" after a Blastx search using an E-value cutoff of $10^{-5}$) showed very strong xylem preferential expression (Table 2). These "no hits", especially contigs with sufficient sequence size, could be genes unique to conifers, gymnosperms or wood forming tissue. However, according to Kirst et al. (2003), the number of pine wood ESTs showing "no hits" decreased when longer contigs were used. Thus, some "no hits", especially contigs with a short sequence such as singletons, could be unrecognized contaminants from various sources (Kirst et al., 2003).

Lorenz and Dean (2002) used SAGE, serial analysis of gene expression, to profile secondary xylem gene expression and to identify genes that are differentially expressed along the axial developmental gradient in loblolly pine xylem. Genes

Table 2. Examples of genes expressed preferentially in xylem.

| Putative Function | Clone ID | Contig # | Accession # | p-value X/M | p-value X/E | p-value X/N |
|---|---|---|---|---|---|---|
| **Cell wall-related proteins** | | | | | | |
| cellulose synthase | NXNV_181_B11 | Contig7822 | BE431357 | p<0.001 | p<0.01 | p<0.01 |
| xyloglucan endotransglycosylase | NXNV_164_H08 | Contig7709 | BE209174 | p<0.001 | p<0.001 | p<0.01 |
| AGP3 (5n2d) | NXNV_156_G02 | Contig8019 | BE123812 | p<0.01 | p<0.01 | p<0.001 |
| laccase | NXCI_094_C09 | Contig6071 | BF010890 | p<0.001 | p<0.01 | p<0.01 |
| UDP-glucuronosyltransferase | NXNV_079_G02 | Contig7928 | AW698020 | p<0.01 | p<0.001 | p<0.001 |
| lp6 protein - loblolly pine | NXSI_129_F11 | Contig8048 | BQ702534 | p<0.01 | p<0.01 | p<0.01 |
| AGP6 | NXSI_040_D02 | Contig8039 | BF609096 | p<0.01 | p<0.01 | p<0.001 |
| expansin 9 | NXCI_132_F04 | Contig7754 | BF186094 | p<0.01 | p<0.01 | p<0.01 |
| cellulase | NXNV_120_C02 | Contig7361 | AW869945 | p<0.001 | | p<0.001 |
| **Intermediate metabolism** | | | | | | |
| UDP-glucosyltransferase | NXCI_083_A06 | Contig7463 | BE762178 | p<0.01 | p<0.01 | p<0.01 |
| myo-inositol-1-phosphate synthase | NXSI_103_A10 | Contig7995 | BG039701 | p<0.001 | | p<0.001 |
| **DNA/RNA binding proteins** | | | | | | |
| transcription factor Hap5a | NXSI_054_E11 | Contig6389 | BF610085 | p<0.001 | p<0.01 | p<0.01 |
| **Translation** | | | | | | |
| translation initiation factor eIF-4A | NXSI_057_C07 | Contig7311 | BF610274 | p<0.01 | p<0.01 | p<0.01 |
| ribosomal protein L2 | NXSI_045_A04 | Contig6890 | BF609324 | p<0.001 | | p<0.001 |
| **Inducible** | | | | | | |
| ABA induced protein | NXSI_130_B01 | Contig4478 | BQ702570 | p<0.001 | p<0.001 | |
| **Stress-related** | | | | | | |
| heat shock protein 82 | NXSI_116_B04 | Contig7801 | BG040868 | p<0.001 | p<0.001 | |
| 18.2 KDA class I heat shock protein | NXSI_139_G02 | Contig3759 | BG275507 | p<0.001 | p<0.01 | |
| **Disease-related** | | | | | | |
| disease resistance responsive gene | NXNV_106_E03 | Contig7742 | AW984958 | p<0.001 | p<0.01 | p<0.01 |
| metallothionein-like protein | NXSI_001_E11 | Contig7169 | BF516649 | p<0.01 | p<0.001 | p<0.01 |
| **Transport** | | | | | | |
| hexose transporter | NXCI_114_B08 | Contig7821 | BF185986 | p<0.001 | | p<0.001 |
| **Signal transduction** | | | | | | |
| putative calmodulin | NXNV_015_H07 | Contig5843 | CD026887 | p<0.001 | | |
| **Lipid biosynthesis** | | | | | | |
| Beta-ketoacyl-ACP synthetase I-2 | NXSI_116_H07 | Contig2178 | BG040941 | p<0.001 | p<0.001 | |
| **Cytoskeleton** | | | | | | |
| tubulin beta -2 chain | NXSI_125_H09 | Contig7992 | BQ702187 | p<0.001 | p<0.01 | p<0.01 |
| actin | NXSI_067_G03 | | BF777317 | p<0.001 | p<0.001 | |
| actin | NXNV_072_D07 | Contig7905 | AW697948 | p<0.001 | p<0.01 | |
| **No Hits** | | | | | | |
| | NXCI_133_E11 | Contig7624 | CD017749 | p<0.001 | p<0.001 | p<0.01 |
| | NXSI_063_E04 | Contig6824 | BQ701379 | p<0.01 | p<0.01 | p<0.001 |
| | NXSI_086_D12 | Contig7554 | BF778735 | p<0.001 | p<0.01 | p<0.01 |
| | NXNV_149_F10 | Contig2323 | BE123558 | p<0.001 | p<0.001 | p<0.001 |
| | NXSI_054_A09 | Contig5671 | BF610040 | p<0.001 | p<0.001 | p<0.01 |

Clone IDs and contig numbers are from the loblolly pine xylem EST database (http://web.ahc.umn.edu/biodata/nsfpine/contig_dir16/).

identified as preferentially expressed in xylem compared to other tissues (p=0.01) in this study were compared with the 40 most abundant transcripts in loblolly pine xylem identified by SAGE. We found that 17 of the 21 most abundant transcripts identified by SAGE with known putative functions were on our list of preferentially expressed genes (p=0.01). We also compared our gene list (p=0.01) with the 100 contigs that have the highest number of clones aligned in the loblolly pine xylem EST database (http://web.ahc.umn.edu/biodata/nsfpine/contig_dir16/). We found that 23 of these contigs were included on our list. This number is significant because not all of the 100 most abundant contigs are included on our current array and not all of them are expected to be xylem-specific. Table 3 indicates examples of genes that are preferentially expressed in xylem (this study) and are among the most abundant transcripts based on SAGE (Lorenz and Dean, 2002) or the pine EST database.

**Real-time quantitative RT-PCR**

To test the overall validity of our microarray data, 11 clones ranging from the top to the bottom X/M ratio were selected and analyzed by real-time quantitative RT-PCR (Heid et al., 1996). For each of the selected clones, real-time RT-PCR was repeated three times. To produce a ratio relative to xylem, the $\Delta\Delta$-$C_T$ analysis method (see materials and methods for the details) was used. After the analysis, the $Log_2$ X/M ratio from real-time RT-PCR analysis was plotted correlation coefficient of 0.93 confirms the overall validity of the microarray data produced by this study (Figure 4). against the $Log_2$ X/M ratio from the microarray analysis. The calculated correlation coefficient of 0.93

Table 3. Genes preferentially expressed in xylem (this study) that are highly expressed based on EST sequencing (http://web.ahc.umn.edu/biodata/nsfpine/contig_dir16/) and SAGE data (Lorenz and Dean, 2002).

| Contig # | Accession # | EST rank | SAGE rank | Putative function |
|---|---|---|---|---|
| Contig8067 | BQ701188 | 4 | | Unknown |
| Contig8063 | BE431364 | 8 | | Allergen-like protein BRSN20 |
| Contig8057 | AW984986 | 14 | 9 | Aquaporin |
| Contig8054 | CD026621 | 17 | | SAM synthetase |
| Contig8051 | BF169951 | 20 | | Adenosylhomocysteinase |
| Contig8048 | BQ702534 | 23 | | Lp6 protein |
| Contig8046 | AW698113 | 25 | 15 | Phenylcoumaran benzylic ether reductase |
| Contig8045 | BG040930 | 26 | | Photoassimilate-responsive protein PAR-1c precursor |
| Contig8042 | BE761849 | 29 | | H+-PYROPHOSPHATASE (EC 3.6.1.1) |
| Contig8039 | BF609096 | 32 | | AGP (AGP 6) |
| Contig8037 | BF777810 | 34 | | dTDPglucose 4,6-dehydratase (EC 4.2.1.46) |
| Contig8035 | BF010655 | 36 | | Trans-cinnamate 4-hydroxylase (EC 1.14.13.11) |
| Contig8030 | BQ701804 | 41 | | Phytocyanin |
| Contig8029 | CD021087 | 42 | | Tubulin alpha chain |
| Contig8024 | BQ702308 | 47 | | Laccase |
| Contig8022 | BE644038 | 49 | 7 | AGP (14A9) |
| Contig8019 | BE123812 | 52 | 4 | AGP (AGP3) |
| Contig8016 | AW784086 | 55 | | Polyubiquitin |
| Contig8010 | BF610091 | 61 | | Glycine-rich protein |
| Contig8007 | AW985236 | 64 | | UDP-glucose dehydrogenase |
| Contig7999 | BF777966 | 72 | 23 | Actin-depolymerizing factor |
| Contig7995 | BG039701 | 76 | | Myo-inositol-1-phosphate synthase |
| Contig7993 | BF186313 | 78 | | Water-stress-inducible protein LP3 |
| Contig7992 | BQ702187 | 79 | | Tubulin beta-2 chain |
| Contig7991 | BQ702339 | 80 | | ABC transporter |
| Contig7317 | AW758919 | 754 | 13 | Alpha-galactosidase |
| Contig7169 | BF516649 | 902 | 20 | Metallothionein-like protein |

confirms the overall validity of the microarray data produced by this study (Figure 4).

Previous northern blot analyses of two AGPs (Ptx3H6 and Ptx14A9) usingxylem, needles, megagametophytes, and embryos showed that they are preferentially expressed in xylem. Almost no expression in the megagametophyte and only trace amounts in embryos and needles could be detected (Loopstra and Sederoff, 1995). A similar northern blot analysis for another AGP showed that PtaAGP3 was xylem-specific compared to other tissues as well (Loopstra et al., 2000). Allona et al. (1998) also performed northern blot analyses of cell wall-related cDNAs including xyloglucan endotransglycosylase, glycine-rich proteins and proline-rich proteins in various tissues and confirmed their high xylem preferential expression. The relative expression ratios ($\log_2$ ratio) for genes identified as preferentially expressed in xylem in our microarray data were much smaller than expected (see supplementary data for $\log_2$ ratio) from these studies. The comparison between microarray and real-time RT-PCR data (Figure 4) also indicates that there was an apparent compression or underestimation of differential expression in the microarray data.

There could be several possible reasons for compression of the microarray data. Yue et al. (2001) and Li et al. (2002) pointed out that the concentration of the spotted DNA on the slide could have a significant impact on the signal intensity and reproducibility, and insufficient DNA concentration on the spot (less than 0.2 μg/μl) could cause compression in differential expression. During the microarray fabrication, cDNAs were PCR amplified and PCR products were checked on a 1% agarose gel. However, the amount of the PCR products in the spotting solution was not measured.

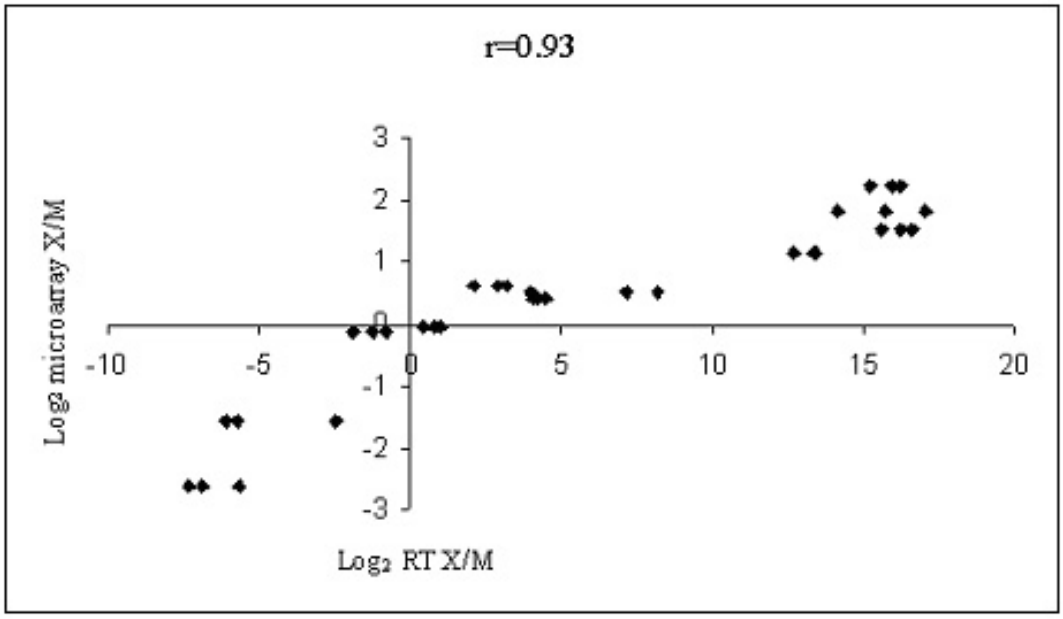


Figure 4. A scatter plot representing the correlation (r=0.93) between the microarray data and real-time RT-PCR data. The $Log_2$ X/M ratio from microarray analysis (Y axis) was plotted against the $Log_2$ X/M ratio from real-time RT-PCR (X axis).

Also, several factors such as probe concentration, ionic strength, and temperature during the hybridization may need to be further optimized to get a higher stringency with maximum signal and minimum background. Kinlaw and Neale (1997) suggested that complex gene families are relatively common in pine and investigation of the current loblolly pine xylem EST database also suggests existence of multiple members of many gene families. Cross-hybridization among different members of the same gene family could also be responsible for compression of the microarray data especially when using cDNA microarrays. Finally, a combination of pre-normalization and ANOVA is a

powerful tool to remove systematic variation associated with microarray data and to find differentially expressed genes with statistical confidence. However, elaborate smoothing techniques such as the lowess adjustment method (Yang et al., 2001) used in this study could have over-fit the data causing a compression of the signal and the underestimation of differential expression. The lowess adjustment method is a global normalization procedure assuming the majority of genes are expressed constantly regardless of the treatment and only a small portion of the genes on the array are regulated by the treatment. Our array does not represent all genes expressed in xylem and may have been enriched for xylem-preferred transcripts. Thus, a global adjustment such as the lowess adjustment would be expected to diminish the magnitude of differential expression for authentically regulated genes. However, dye effects seem inevitable in most microarray data and normalization to remove the unbalance between the two dyes is critical (Kerr and Churchill, 2001b; Yang et al., 2001). Currently there are several normalization methods available and many other normalization strategies are being developed (Li et al., 2002). It may always be a compromise between being conservative using a powerful normalization tool and perhaps over-fitting the data and being less conservative using a somewhat less elaborate method and having a less desired effect allowing crude fitting. However, due to the existence of inherent noise and inevitable dye effects associated with microarray data, a more conservative approach would be preferable. A combination of elaborate pre-normalization such as the lowess adjustment method of Yang et al. (2001) to smooth data and the ANOVA method (Kerr and Churchill, 2001a) seems very powerful to remove systematic variation associated with microarray data and to produce

a reliable and reproducible result with statistical confidence.

# CHAPTER III

## SEASONAL VARIATION IN GENE EXPRESSION FOR LOBLOLLY PINES (*Pinus taeda L.*) FROM DIFFERENT GEOGRAPHICAL SOURCES

### INTRODUCTION

Commercially important characteristics such as wood specific gravity are known to differ with seed source. For example, when grown on a common site, the specific gravity of Arkansas trees is greater than that of Louisiana trees and Texas trees have a greater specific gravity than Atlantic coast sources (Byram and Lowe, 1988). Differences in wood specific gravity among trees from different seed sources are likely due to the amount of earlywood vs. latewood. The more latewood, the greater the wood specific gravity (Byram and Lowe, 1988). Specific gravity is the single most important physical property of wood. Most mechanical properties of wood such as the strength, stiffness, yield of pulp per unit, heat transmission, and heat per unit volume produced in combustion are closely correlated to specific gravity (Haygreen and Bowyer, 1996).

Tracheid cells formed early in the growing season (earlywood) differ from those formed later in the year (latewood) in their chemical composition and physical characteristics. Latewood cells have greater density, smaller lumen, a smaller radial diameter and thicker cell walls than earlywood cells (Zimmermann and Brown, 1971 p.96). Previous studies suggested that several factors are responsible for differences

between earlywood and latewood. The formation of cells with large diameter (earlywood) is highly dependent on the availability of soil moisture (Kozlowski et al., 1991). On the other hand, the latewood transition is positively correlated to the cessation of height growth (Jayawickrama et al., 1997) suggesting that development of latewood cells with thicker cell walls is dependent on the availability of photosynthate later in the year (Zimmermann and Brown, 1971 p.96). Early in the growing season, most of the available sugars are used for the shoot and needle growth leaving only a little for cell wall thickening during xylogenesis. However, later in the growing season, the apical growth at the shoot tips and needles has mostly ceased and most of the photosynthates produced by the now full-grown needles are available for cell wall thickening, producing tracheid cells with a small diameter and a thicker cell wall (Zimmermann and Brown, 1971 p.96). Latewood cells can also be induced by several environmental stimuli such as drought, photoperiod and temperature (Larson, 1969). Plant hormones, especially auxin, have been suggested as important factors involved as well (Larson, 1962; Whitmore and Zahner, 1996; Dodd and Fox, 1990; Uggla et al., 2001; Mellorowicz et al., 2001). We are interested in genes preferentially expressed in latewood because latewood has more desirable wood properties than earlywood. Pulp yield and cohesiveness are positively correlated to alpha-cellulose and hemicellulose content and pulping cost is negatively correlated to lignin content (Smook and Kocurek, 1988). Significantly higher holocellulose and lower lignin contents were observed for latewood compared to that of earlywood of loblolly pine (Sewell et al., 2002). In addition, microfibril angle has a significant effect on lumber strength, stiffness and dimensional

stability and latewood cells typically have a smaller microfibril angle that is more desirable (Megraw, 1985). Thus, increasing latewood percent would have a more positive effect on wood properties such as wood density and wood specific gravity than earlywood (van Buijtenen, 1964; Gilmore et al., 1966; Byram and Lowe, 1988; Zobel and Jett, 1995; Jayawickrama et al., 1997; Sewell et al., 2002).

To examine variation in gene expression among trees from different geographical sources when grown on a common site and seasonal variation in gene expression within each seed source, transcript profiles of differentiating xylem from two different loblolly pine seed sources, South Arkansas and South Louisiana, and from earlywood and latewood within each seed source, were compared. Microarrays containing 2171 ESTs with putative functions of interest, selected from several loblolly pine xylem partial cDNA libraries and a shoot tip library were used. Many genes preferentially expressed in latewood compared to earlywood were for proteins involved in cell wall biosynthesis. Variation in gene expression among trees from the two seed sources within each growing season suggests that there may be more differences between South Arkansas trees and South Louisiana trees within latewood than within earlywood. Finally, variation in gene expression among trees from different regions may reflect adaptation to different environments.

# MATERIALS AND METHODS

## Plant materials

The Western Gulf geographic seed source study was established in the 1970s to have insight into loblolly pine seed movement within the states of Arkansas, Louisiana, Mississippi, Oklahoma, and Texas. Each seed source consisted of five open-pollinated families that came from parent trees selected for superior properties such as growth, form, and specific gravity from each of the Western Gulf provenances (Byram and Lowe, 1988). For this study, differentiating xylem samples collected from trees in one of the 28 plantings of the Western Gulf seed source study, a planting near Hudson, TX, were used for analysis. Newly differentiating xylem was isolated from loblolly pines from South Arkansas and South Louisiana, the two seed sources that showed the biggest difference in wood specific gravity when grown on a common site (Byram and Lowe, 1988). Three families per seed source were selected and samples were collected from five trees per family. Differentiating xylem was collected separately in early July and October for earlywood and latewood comparisons. The same trees were sampled each time. A hole saw was used to remove the bark and phloem. The differentiating xylem was scraped from the stem and frozen in liquid nitrogen. Five millimeter increment cores were taken from bark to bark and examined under a light microscope to check the status of earlywood latewood development.

**Microarray preparation**

The microarrays used in this study were made by the Forest Biotechnology Group at North Carolina State University and kindly provided to us. A total of 1936 contigs with putative functions of interest were selected as well as 236 contigs without any putative function assigned ("no hits" after a Blastx search using an E-value cutoff of $10^{-5}$). The contigs contain EST sequences from normal wood, side wood, compression wood, latewood, planings (deeper cell layers), root wood xylem and shoot tip cDNA libraries made from loblolly pine (Whetten et al., 2001; Kirst et al., 2003). To reduce nonspecific hybridization, EST clones located near the 3' end of each contig and larger than 200 bp were selected to be amplified by PCR. The microarray preparation procedures used in this study were described in Chapter II.

**Probe synthesis and hybridization**

Total RNAs were extracted from differentiating xylem using the method of Chang et al. (1993). Residual DNA was removed using DNA-free™ (Ambion Inc. Austin, TX) and RNA was further purified using MEGAclear™ (Ambion Inc. Austin, TX ). Due to the limited number of microarray slides available, an equal amount of RNA was pooled from each sample and combined to synthesize probes following the experimental design in Figure 5A. Briefly, about 300 milligrams of ground differentiating xylem tissue from five trees from each family and growing season were combined and RNAs were purified. The RNA purifications were repeated with separate

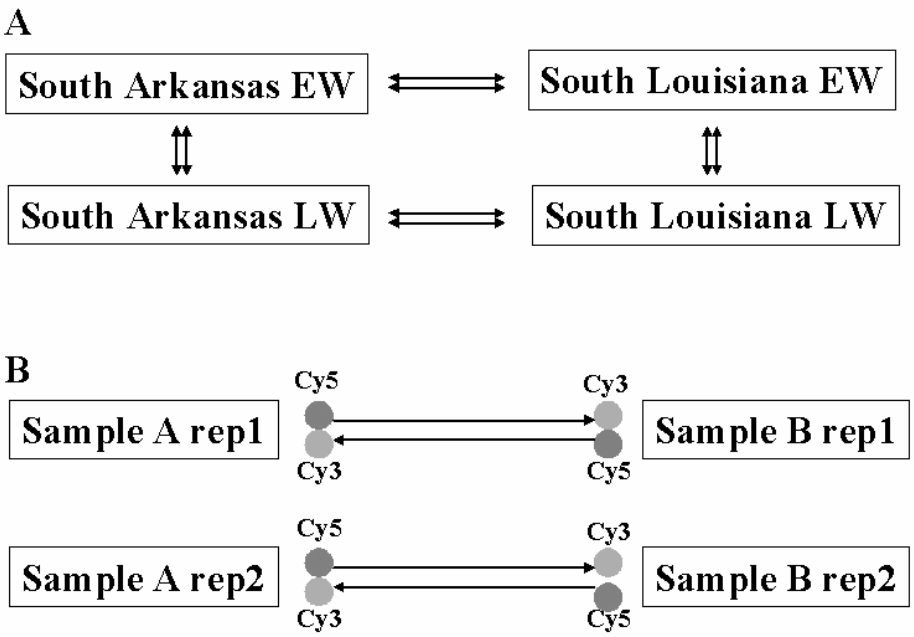


Figure 5. Loop experimental design used for probe labeling and hybridization (A). Double dye-flip method used for each comparison (B). EW, earlywood; LW, latewood

samples to produce biological replications (Figure 5B). An equal amount of RNA was pooled from each family and combined into four samples as shown in Figure 1A. For the first strand cDNA synthesis and labeling, the FluoroScript$^{TM}$ cDNA Labeling System (Invitrogen Corp. CA) was used following the manufacturer's instructions. The hybridization and washing procedures used were described in Chapter II.

**Image analysis, quantification and data analysis**

After the washing steps, the slides were scanned with both channels using a ScanArray 3000 (GSI Lumonics, Watertown, MA) at a 10 μm resolution. While scanning, the laser power and photo-multiplier tube setting were adjusted to approximately balance both channels. The intensities of each spot for both channels were quantified using QuantArray (GSI Lumonics, Watertown, MA). For the microarray data analyses, we adopted the fixed effect ANOVA model of Kerr and Churchill (2001a,b) using MA-ANOVA, a software package for the analysis of spotted cDNA microarray experiments by Wu, Kerr, Cui and Churchill (http:\\www.jax.org\research\Churchill). Measured spot intensities for each of the channels were pre-normalized following the method of Yang et al. (2001) using the "malowess" function in MA-ANOVA. Separate statistical tests were performed for seasonal variation within each seed source and variation between the two seed sources within each growing season. An ANOVA was carried out using the "fitmaanova" function to estimate effects of various factors on each spot intensity. A permutation F-test for the null hypothesis assuming zero VG effect was performed for each spot using the "make_Ftest" function in MA-ANOVA. The multiple testing-adjustment method (Westfall and Young 1993, Oleksiak et al., 2002) was applied to provide tighter control of type I errors.

**Real-time quantitative RT-PCR**

Real-time quantitative RT-PCR (Heid et al., 1996) was used to test the overall validity of our microarray data. Gene specific primers for the selected clones were

designed based on sequences from the NSF Pine EST Database (http://web.ahc.umn.edu/biodata/nsfpine/) using Primer Express (Applied Biosystems, CA). Real-time RT-PCR was repeated three times for each of the selected clones with SYBR-Green PCR Master Mix (Applied Biosystems, CA) on a GeneAmp 7900 Sequence Detection System (Applied Biosystems, CA) following the manufacturer's recommendations. A dissociation curve for each amplicon was examined to confirm the specificity of the primer pair used. The $C_T$ (threshold cycle) value obtained after each reaction was normalized to the $C_T$ value of 18S rRNA (see User Bulletin #2" ABI PRISM 7900 Sequence Detection System, pp 11-15 for details). A paired t-test was used on normalized $C_T$ values ($\Delta C_T$) to test seasonal variation in gene expression within each source and an ANOVA (nested factorial model) was used on $\Delta C_T$ values to test variation between the two seed sources and among families within each seed source.

**RESULTS AND DISCUSSION**

A microarray analysis was performed to examine variation in gene expression among trees from different geographical sources when grown on a common site and seasonal variation in gene expression within each seed source. The gene list for the microarray used in this study is available under Supplementary Data – Tree Physiology 2004 at http://forestry.tamu.edu/newweb/people/faculty&staff/c-loopstra). A total of 1936 genes with putative functions of interest were included on the array. In addition, 236 genes without any putative function assigned ("no hits" after a Blastx search using

an E-value cutoff of $10^{-5}$) were included on the array.

According to Lee et al. (2000), there is inherent variability in most microarray data even after removing systematic sources of variation. Thus, it is important to have an experimental design with a sufficient number of replications to control inherent noise and to produce consistent and reliable results (Kerr and Churchill, 2001a,b; Lee et al., 2000). In this study, each of the ESTs was spotted four times on the array and a loop design (Figure 5A) with a double dye-flip method (Figure 5B) was adopted providing 32 data points for each of the ESTs for each of the comparisons. The fixed effect ANOVA model (Kerr and Churchill, 2001a,b) was adopted to estimate effects of various factors on each spot intensity. The normality in residual distribution for the model adopted was also examined and no significant evidence against homoscedasticity was found (data not shown). Genes with significant variation in expression were identified using a permutation F-test (p-value=0.01) and multiple testing-adjustment was applied to provide tighter control of type I errors (Westfall and Young, 1993; Oleksiak et al, 2002).

**Genes with seasonal variation in expression within each source**

The expression of 110 genes was significantly different between earlywood and latewood within trees from South Arkansas. Within trees from South Louisiana, the expression of 87 genes was significantly different between earlywood and latewood. Among these, the expression of 53 genes was significantly different for both seed sources (gene lists are available under Supplementary Data - Tree Physiology 2004 at http://forestry.tamu.edu/newweb/people/faculty&staff/c-loopstra). Many genes

Table 4. Examples of genes expressed preferentially in latewood compared to earlywood within each seed source.

| Putative Function | Clone ID | Contig # | Accession # | p-value SAR | p-value SLA |
|---|---|---|---|---|---|
| Putative dehydrine | 14 D07 | | AW010819 | p<0.01(2.5)* | p<0.01(1.7) |
| AGP5 | NXSI_079_E12 | Contig2860 | BF777983 | p<0.01(1.8) | p<0.01(1.1) |
| Glycine-rich protein | NXSI_054_F05 | Contig7987 | BF610091 | p<0.01(1.8) | p<0.01(0.1) |
| Adenylate kinase | NXCI_156_C12 | Contig7286 | BF221235 | p<0.01(1.8) | p<0.01(1.0) |
| no hit | NXSI_074_B07 | Contig6989 | CD025671 | p<0.01(1.7) | p<0.01(1.6) |
| Cellulose Synthase-1 | NXSI_024_H01 | Contig7766 | BG039035 | p<0.01(1.7) | p<0.01(1.3) |
| Fructose-bisphosphate aldolase | 07 E05 | | AW010498 | p<0.01(1.5) | p<0.01(1.0) |
| Tubulin alpha-1 chain | 22 F09 | | AW011596 | p<0.01(1.4) | p<0.01(1.0) |
| Lp6 protein - loblolly pine | NXSI_129_F11 | Contig8028 | BQ702534 | p<0.01(1.4) | p<0.01(0.9) |
| AGP5 | NXSI_103_C04 | Contig7819 | BG039717 | p<0.01(1.2) | p<0.01(1.7) |
| Expansin9 precursor | 21 H02 | | AW011524 | p<0.01(1.1) | p<0.01(0.7) |
| Putative importin | NXNV_145_F12 | Contig5507 | BE123594 | p<0.01(1.1) | p<0.01(0.6) |
| Transcription factor Hap5a | NXSI_054_E11 | Contig6314 | BF610085 | p<0.01(1.1) | p<0.01(0.4) |
| Tubulin beta-2 chain | NXSI_125_H09 | Contig7966 | BQ702187 | p<0.01(1.0) | p<0.01(0.7) |
| Cellulose synthase | NXSI_087_D09 | Contig7766 | BF778814 | p<0.01(1.0) | p<0.01(0.5) |
| Putative SF16 protein | NXCI_095_D10 | Contig6082 | BF010830 | p<0.01(1.0) | p<0.01(1.0) |
| 1,4-benzoquinone reductase | NXSI_031_E03 | Contig5720 | BF517905 | p<0.01(1.0) | p<0.01(0.7) |
| SAM Synthetase 2 | 40 D10 | | AW065163 | p<0.01(0.9) | p<0.01(0.6) |
| Isoflavone reductase homolog | NXNV_127_E04 | Contig5231 | AW887964 | p<0.01(0.9) | p<0.01(0.6) |
| MADS box protein AGL2 | ST06D02 | | AW010398 | p<0.01(0.8) | p<0.01(0.9) |
| Pinoresinol-lariciresinol reductase | NXSI_013_B10 | Contig7578 | BF517330 | p<0.01(0.7) | p<0.01(0.3) |
| Phenylcoumaran benzylic ether reductase | NXNV_066_E09 | Contig7621 | AW698113 | p<0.01(0.6) | p<0.01(0.4) |
| Aluminum-induced protein | NXSI_134_D02 | Contig7817 | BQ702940 | p<0.01(0.6) | p<0.01(0.2) |
| no hit | 17 A10 | | AW011130 | p<0.01(2.2) | |
| Sucrose Synthase | NXSI_007_H12 | Contig1276 | BF517003 | p<0.01(1.7) | |
| Polyubiquitin | NXSI_081_D01 | Contig8038 | BF778050 | p<0.01(1.7) | |
| Beta tubulin | NXCI_001_A06 | Contig7830 | CD015920 | p<0.01(1.4) | |
| no hit | 15 F10 | | AW011016 | p<0.01(1.4) | |
| Cinnamyl-alchohol-dehydrogenase | NXNV_162_F07 | Contig7123 | BE187332 | p<0.01(1.2) | |
| SAM Synthetase | NXNV008F05 | Contig8032 | AW289947 | p<0.01(1.2) | |
| Putative SAM Synthetase | NXSI_012_H05 | Contig7945 | BF517311 | p<0.01(1.1) | |
| no hit | NXSI_089_E04 | Contig6991 | BF778882 | p<0.01(1.1) | |
| 14A9 | NXCI_053_H01 | Contig8004 | BE644038 | p<0.01(1.0) | |
| Cellulose synthase | NXSI_108_H05 | Contig7881 | BG040499 | p<0.01(0.9) | |
| Trans-cinnamate 4-hydroxylase | NXCI_087_F07 | Contig8013 | BF010655 | p<0.01(0.4) | |
| Laccase | NXCI_094_C09 | Contig6075 | BF010890 | | p<0.01(0.9) |

Clone IDs and contig numbers are from the loblolly pine xylem EST database (http://web.ahc.umn.edu/biodata/nsfpine/contig_dir20/) and a shoot tip library. SAR, South Arkansas; SLA, South Louisiana. * Figures in the parentheses are log ratios from the microarray analyses.

preferentially expressed in latewood were for proteins involved in cell wall biosynthesis (Table 4). These genes are of particular interest because latewood has more desirable wood properties than earlywood for many purposes and increasing latewood percent in a wood would have a desirable effect (Megraw, 1985; Sewell et al., 2002). We identified two genes (NXSI_054_E11 and ST06D02) for transcription factors that could be involved in the preferential expression of genes in latewood. Both of them are expressed preferentially in latewood compared to earlywood within both seed sources. A previous study in our lab also showed that one of them (NXSI_054_E11) is preferentially expressed in xylem compared to megagametophytes, needles and embryos (Yang et al., 2004). The discovery of trans-acting factors for cis-elements shared by genes preferentially expressed in latewood could be very important because most of the economically important quantitative traits in pines are thought to be controlled by the collective action of multiple genes (Sewell and Neale, 2000). Trans-acting factors regulate many genes sharing the matching cis-elements simultaneously (pleiotropism). Targeting upstream components in the signal transduction pathway rather than final responsive genes for genetic modification could have a more significant effect on the final phenotype. Thus, these two genes could be key regulators of the genes that play important roles during latewood formation and could be candidate genes for important quantitative traits.

Our array contained clones that represent six cellulose synthase contigs (contigs 7881, 7803, 7766, 6864, 5828 and 5607). Among these, only contig 7766 was preferentially expressed in latewood for both of the seed sources ($p=0.01$). Contig 7881

was preferentially expressed in latewood only within trees from South Arkansas. Cellulose microfibrils are synthesized on the plasma membrane by a cellulose synthase complex, which is composed of multiple cellulose synthase subunits (particle rosettes) and sucrose synthase (Delmer and Amor, 1995). Previous studies showed that transcripts for these genes are highly abundant in wood forming tissues of both gymnosperms and angiosperms (Allona et al., 1998; Sterky et al., 1998; Whetten et al., 2001) and their expression was coordinated with late cell expansion and secondary wall formation (Hertzberg et al., 2001). There are multiple members of the cellulose synthase gene family expressed in differentiating xylem and they are differentially regulated (Allona et al., 1998; Sterky et al., 1998; Wu et al., 2000; Yang et al., 2004). Many of them are preferentially expressed in differentiating xylem compared to other tissues (Yang et al., 2004). So far, we have identified 12 different sequences for cellulose synthase in loblolly pine public databases, 11 from loblolly pine xylem EST libraries (contigs 7881, 7803, 7766, 6864, 5828, 5607, 5258, 3670, 2633, 1343 and 993 in http://web.ahc.umn.edu/biodata/nsfpine/contig_dir20/) and one from a shoot tip library (Gene Bank Acc# AW056552). According to Whetten et al. (2001), EST clones with more than 98% sequence similarity are sometimes placed into different contigs when PHRAP (http://www.phrap.org) parameters used for contig assembly are a minimum mismatch of 40 and a minimum score of 80. PHRAP parameters used for contig assembly for the xylem loblolly pine EST database (http://web.ahc.umn.edu/biodata/nsfpine/contig_dir20/) were a minimum match of 50, a minimum score of 100 and a minimum length of 100. Thus, these different contigs for cellulose synthase could represent allelic variation or

different members of the loblolly pine cellulose synthase gene family. The level of sequence similarity among these contigs and a previous study in our lab (Yang et al., 2004) strongly suggest that these are sequences for different members of the cellulose synthase gene family. Our array also contained clones that represent three sucrose synthase contigs (contigs 7744, 3656 and 1276). Contig 1276 was preferentially expressed in latewood compared to earlywood within trees from South Arkansas ($p=0.01$). Sucrose synthase is responsible for channeling UDP-glucose into the growing cellulose in a cellulose synthase complex (Delmer and Amor, 1995).

Our array contained a total of 12 clones representing the loblolly pine tubulin multigene family. Seven clones represented the beta-tubulin gene family (contigs 7995, 7966, 7830, 7162, 6405, 4803 and 4624) and 5 clones represented the alpha-tubulin gene family (contigs 8045 and 8008 from xylem cDNA libraries and Gene Bank Acc# AW011596, AW042942 and AW010543 from a shoot tip library). Among the 12 tubulin genes represented, contig 7966 and clone AW011596 were expressed preferentially in latewood within trees from both seed sources. Contig 7830 and clone AW010543 were expressed preferentially in latewood only within trees from South Arkansas. Microtubules are heterodimeric polymers of the alpha- and beta-tubulins that are encoded by multigene families in plants (Ludwig et al., 1987, Hussey et al., 1990, Snustad et al., 1992). The orientation of newly deposited cellulose microfibrils is determined by cortical microtubules that are cross-linked to the cytoplasmic face of the plasma membrane creating channels within the membrane for the directional movement of cellulose synthase complexes along the membrane during cellulose microfibril

synthesis (Giddings and Staehelin, 1988; Baskin et al., 1994; Nick, 2000). Thus, these tubulin genes could be candidate genes for S2 layer microfibril angle in the secondary cell wall of tracheid cells.

Most of the known genes involved in lignin biosynthesis were represented on the microarray. In this study, several genes involved in lignin biosynthesis including trans-cinnamate 4-hydroxylase (contig 8013), cinnanmyl-alcohol dehydrogenase (contig 7123) and laccase (contig 6075) were preferentially expressed in latewood. Several SAM synthase genes that are thought to be involved in methyl transfer during cross-linking of the monolignols were preferentially expressed in latewood as well (Table 4). Lignin is a phenolic polymer derived from irregular cross-linking of aromatic alcohol subunits (monolignols) such as *p*-coumaryl alcohol, coniferyl alcohol and synapyl alcohol. These monolignols are synthesized from phenylalanine by the phenylpropanoid pathway and secreted into the apoplast where they are cross-linked and become a major reinforcing matrix of the secondary cell wall where cellulose microfibrils are embedded (reviewed by MacKay et al., 1997). Previous studies showed that transcripts for genes involved in the lignin biosynthetic pathway are among the most abundant transcripts in wood forming tissues (Allona et al., 1998; Sterky et al., 1998; Whetten et al., 2001; Lorenz and Dean, 2002). In some cases, expression is up-regulated in compression wood compared to normal wood (Whetten et al., 2001) and different members of gene families are differentially regulated during wood formation (Hertzberg et al., 2001). Many genes involved in lignin biosynthesis are preferentially expressed in differentiating xylem compared to other tissues as well (Yang et al., 2004). The preferential expression of

genes involved in lignin biosynthesis in latewood compared to earlywood in this study is puzzling because previous studies showed lower lignin content in latewood than earlywood (Saka and Goring, 1985; Sewell et al., 2002). Differences could be due to properties of the samples used in this study (differentiating xylem) and samples used in previous studies (inner rings of 5-mm radial wood core). Alternatively, even if lignin content is lower in latewood relative to other cell wall components, the total amount of lignin needed may be greater due to thicker cell walls.

Several genes encoding putative cell wall proteins were represented on the arrays. Transcripts for cell wall structural proteins such as arabinogalactan-proteins (AGPs), glycine-rich proteins and proline-rich proteins are among the most abundant transcripts in wood forming tissues and are preferentially expressed in differentiating xylem tissue compared to other tissues (Loopstra and Sederoff, 1995; Allona et al*.,* 1998; Sterky et al., 1998; Loopstra et al., 2000; Zhang et al., 2000; Whetten et al., 2001; Lorenz and Dean, 2002; Yang et al., 2004). AGPs are highly glycosylated proteoglycans or glycoproteins (2-10% protein by weight) which contain highly repetitive sequences and are frequently rich in proline/hydroxyproline, serine, threonine, alanine, and glycine. General features of classical AGPs include a signal peptide for targeting to the endoplasmic reticulum, a hydroxyproline-rich domain and a GPI anchor site for plasma membrane attachment. Six different loblolly pine AGP-like genes or gene families have been identified (Loopstra et al., 2000; Zhang et al., 2000) and they are among the most abundant transcripts in differentiating xylem tissue (Sterky et al., 1998; Whetten et al., 2001; Lorenz and Dean, 2002). The loblolly pine xylem EST database (http://web.ahc.

umn.edu/biodata/nsfpine/contig_dir20/) contains numerous contigs for different AGP-like genes (contigs 8005, 8026 and 7885 for ptaAGP6; contigs 7819, 7620, 7601, 6582, 6524, 6283, 2860 and 221 for ptaAGP5; contigs 7932, 7916 and 7402 for ptaAGP4; contigs 7842, 7704, 5448, 80 and 16 for ptx3H6, contigs 8004, 7979 and 7689 for ptx14A9; contigs 7990, 7873, 7675, 7646 and 6417 for ptaAGP3). Careful examination of DNA and protein sequence similarity among contigs for each AGP-like gene suggest that there are at least seven different members of the ptaAGP5 gene family. Our array contained one clone for ptaAGP6 (contig 8026), four clones for ptaAGP5 (contigs 7819, 7601, 6582 and 2860), one clone for ptx3H6 (contig 7842), two clones for ptx14A9 (contigs 8004 and 7689) and three clones for ptaAGP3 (contigs 7990, 7675 and 6417). Among these, contigs for two different ptaAGP5 genes (contigs 7829 and 2860) were preferentially expressed in latewood compared to earlywood within trees from both seed sources. One contig for ptx14A9 (contig 8004) was preferentially expressed in latewood within trees from South Arkansas. Two contigs for glycine-rich proteins were preferentially expressed in latewood compared to earlywood as well. Contig 7987 was preferentially expressed in latewood within trees from both seed sources and contig 8038 was preferentially expressed in latewood within trees from South Arkansas.

Some of the genes with no putative functions assigned ("no hits" after a Blastx search using an E-value cutoff of $10^{-5}$) also showed preferential expression in latewood. These "no hits", especially contigs with sufficient sequence size, could be genes unique to conifers, gymnosperms or wood forming tissue. However, according to Kirst et al. (2003), the number of pine wood ESTs showing "no hits" decreased when longer contigs

were used. Thus, some “no hits”, especially contigs with a short sequence and singletons, could be unrecognized contaminants from various sources (Kirst et al., 2003).

Interestingly, many stress-related genes, especially genes considered to be induced by drought stress, were preferentially expressed in earlywood compared to latewood within each seed source (Table 5). This made data analysis difficult because up-regulation of these genes in earlywood could be due to drought stress received during the early growing season not representing typical differences between earlywood and latewood. Maximum temperatures in the week preceding the earlywood harvest ranged from 30 °C to 34.4 °C with 23 mm of rain in the preceding two weeks. Maximum temperatures in the week preceding the latewood harvest ranged from 25 °C to 30.6 °C with 56 mm of rain in the preceding two weeks. However, a dehydrine, belonging to a group of well-known drought stress-induced genes, was strongly preferentially expressed in latewood among trees within both seed sources. Thus, preferential expression of these stress-related genes in earlywood might not be simply due to drought stress and more complicated unknown stress-related regulation might be involved. These stress-related genes could play a significant role during earlywood formation as well.

**Genes with variation in expression among trees from different geographical sources within each growing season**

Within latewood, 131 genes with significant variation in gene expression between South Arkansas and South Louisiana trees were identified. However, within earlywood, only 51 genes with variation in gene expression between the two seed

Table 5. Examples of stress-related genes with variation in expression between earlywood and latewood within each seed source and among trees from the two seed sources within each growing season.

| | | | | p-value (LW vs EW) | | p-value (SAR vs SLA) | |
|---|---|---|---|---|---|---|---|
| Putative Function | Clone ID | Contig # | Accession # | SAR | SLA | LW | EW |
| Low molecular weight heat shock protein | 40 F04 | | AW065178 | p<0.01(-0.8)* | p<0.01(-0.8) | p<0.01(0.8) | p<0.01(0.8) |
| Class 1 heat shock protein | 14 B10 | | AW010802 | p<0.01(-0.8) | p<0.01(-0.9) | p<0.01(0.7) | p<0.01(0.6) |
| Heat shock protein 82 | NXSI_011_G01 | Contig7776 | BF517211 | p<0.01(-0.7) | p<0.01(-0.6) | | p<0.01(0.7) |
| Heat shock protein 70 | NXSI_126_A05 | Contig5612 | BQ702194 | p<0.01(-0.7) | p<0.01(-0.6) | | p<0.01(0.8) |
| 18.2 KDA class 1 heat shock protein | NXSI_139_G02 | Contig3618 | BG275507 | p<0.01(-0.6) | | | |
| Heat shock protein 20 | NXSI_100_C03 | Contig6525 | BG039545 | p<0.01(0.9) | | | p<0.01(0.4) |
| Water-stress-inducible protein LP3 | NXCI_132_H04 | Contig7971 | BF186115 | p<0.01(-0.4) | p<0.01(-0.3) | p<0.01(0.4) | |
| Water stress inducable protein LP3 | 34 H09 | | AW064717 | p<0.01(-0.1) | p<0.01(-0.3) | | |
| Water stress inducable protein Lp3 | 23 A08 | | AW042659 | p<0.01(-0.8) | p<0.01(-0.3) | | p<0.01(0.4) |
| Putative dehydrine | 14 D07 | | AW010819 | p<0.01(2.5) | p<0.01(1.7) | p<0.01(1.3) | |
| Putative dehydrine | NXSI_058_G02 | Contig7918 | BF610459 | p<0.01(1.5) | p<0.01(0.7) | p<0.01(1.0) | |
| Putative dehydrine | NXSI_071_D09 | Contig7967 | BF777423 | | | p<0.01(0.9) | |
| Putative dehydrine | NXCI_002_C10 | Contig6080 | BE451794 | | | p<0.01(0.6) | |
| Probable aquaporin | NXSI_048_C09 | Contig7765 | BF609696 | p<0.01(-0.6) | p<0.01(-0.4) | p<0.01(-0.3) | |
| Aquaporin | 07 A03 | | AW010453 | p<0.01(-0.2) | p<0.01(-0.4) | | |
| Drought-induced protein | NXNV_123_E11 | | AW870057 | p<0.01(-0.8) | | | |

Clone IDs and contig numbers are from the loblolly pine xylem EST database (http://web.ahc.umn.edu/biodata/nsfpine/contig_dir20/) and a shoot tip library. * Figures in the parentheses are log ratios from the microarray analyses. Positive log ratio values (>0) represent preferential expression in latewood and South Arkansas trees compared to earlywood and South Louisiana trees. SAR, South Arkansas; SLA, South Louisiana; EW, earlywood; LW, latewood.

sources were identified. Table 6 shows examples of genes differentially expressed between the two seed sources within each growing season. Among these, the expression of 29 genes was significantly different for both earlywood and latewood (gene lists are available under Supplementary Data - Tree Physiology 2004 at http://forestry.tamu.edu/newweb/people/faculty&staff/c-loopstra). The difference in the number of genes identified between earlywood and latewood implies that there might be a greater difference in latewood than earlywood between trees from South Arkansas and South Louisiana. These are the two seed sources that showed the largest difference in wood specific gravity when grown on a common site in a previous study (Byram and Lowe, 1988). Previous studies suggested that latewood would have more significant effects on wood properties such as wood density and wood specific gravity than earlywood (van Buijtenen, 1964; Gilmore et al., 1966; Byram and Lowe, 1988; Zobel and Jett, 1995; Jayawickrama et al., 1997).

The native range of loblolly pine is widely distributed, spanning 15 states from New Jersey to central Florida and westward to Texas (Baker and Landon, 1990). Following the withdrawal of glaciers after the climax of the Wisconsin glaciation 13,000 years ago, loblolly pines migrated northward following optimum environments (Watts, 1983). Thus, the current population structure of loblolly pine in its natural range has been established for a long period of time evolving adaptations to the new environments. Loblolly pines are out-crossing (wind-pollinated) in nature. Gene flow by pollen dispersal is highly extensive (DiGiovanni et al., 1996) and selfed progenies are inferior to out-crossed ones for competition in nature (Kraus and Squillace, 1964). Breeding

Table 6. Examples of genes differentially expressed between the two seed sources within each growing season.

| | | | | p-value | |
|---|---|---|---|---|---|
| Putative Function | Clone ID | Contig # | Accession # | LW | EW |
| No hit | NXSI_021_D01 | Contig7400 | BF517419 | p<0.01(2.2)* | p<0.01(1.0) |
| No hit | 17 A10 | | AW011130 | p<0.01(2.1) | p<0.01(1.2) |
| Glycine-rich protein homolog | NXSI_054_F05 | Contig7987 | BF610091 | p<0.01(1.5) | p<0.01(0.7) |
| Hypothetical protein | 16 C05 | | AW011062 | p<0.01(1.4) | p<0.01(1.2) |
| Metallothionine-like protein EMB 30 | 38 A10 | | AW064970 | p<0.01(1.4) | p<0.01(0.8) |
| Metallothionine like protein | 35 A01 | | AW064720 | p<0.01(1.4) | p<0.01(1.1) |
| Hypothetical protein | NXSI_021_B12 | Contig7600 | BF517412 | p<0.01(1.3) | p<0.01(0.9) |
| Putative dehydrine | 14 D07 | | AW010819 | p<0.01(1.3) | |
| Alluminum induced prot | 37 A06 | | AW064886 | p<0.01(1.3) | p<0.01(0.8) |
| Putative importin | NXNV_145_F12 | Contig5507 | BE123594 | p<0.01(1.2) | |
| Dormancy associated protein | 40 D05 | | AW065158 | p<0.01(1.2) | |
| Pathogenesis-related protein PR-1 | NXNV_082_C11 | Contig7141 | CD020656 | p<0.01(1.1) | |
| Putative casein kinase | NXCI_048_E07 | Contig2796 | BE643881 | p<0.01(1.1) | |
| Hypothetical ORF-6 protein | 17 C04 | | AW011146 | p<0.01(1.1) | |
| TCTP-like protein | 08 A07 | | AW010558 | p<0.01(1.1) | |
| Putative dehydrin | NXSI_058_G02 | Contig7918 | BF610459 | p<0.01(1.0) | |
| Actin | 06 G12 | | AW010441 | p<0.01(1.0) | |
| Polyubiquitin | NXNV_123_D06 | Contig6253 | CD020985 | p<0.01(1.0) | |
| No hit | NXSI_054_A09 | Contig6104 | BF610040 | p<0.01(0.9) | p<0.01(1.1) |
| SAM synthase | NXSI_023_F01 | Contig8042 | BQ701188 | p<0.01(0.9) | |
| Hydroxyproline enriched glycoprotein | 15 B09 | | AW010974 | p<0.01(0.9) | |
| Adenosyl methionine syhthetase | 08 F07 | | AW010600 | p<0.01(0.9) | |
| Laccase (diphenol oxidase) | NXNV_066_B07 | Contig5714 | AW698095 | p<0.01(-0.2) | p<0.01(-0.2) |
| probable UDP-glucuronosyltransferase | NXNV_079_G02 | Contig7913 | AW698020 | p<0.01(-0.4) | p<0.01(-0.5) |
| Proline-rich protein | NXCI_021_D03 | Contig7921 | BE496599 | p<0.01(-0.4) | p<0.01(-0.2) |

Clone IDs and contig numbers are from the loblolly pine xylem EST database (http://web.ahc.umn.edu/biodata/nsfpine/contig_dir20/) and a shoot tip library. * Figures in the parentheses are log ratios from the microarray analyses. Positive log ratio values (>0) represent preferential expression in South Arkansas trees compared to South Louisiana trees. LW, latewood; EW, earlywood.

efforts have recently accelerated the rate of genetic change (Wells 1985, Lambeth 1984, Li et al., 1999). However, the natural population of loblolly pine is highly heterozygous and still maintains much of its original variation. Several previous studies have

suggested climate-related selection pressure in the evolution of many traits of loblolly pine. Variation of such traits generally shows a north to south pattern implying latitude-related climatic variables. In general, compared to northern populations, southern populations have smaller seed, worse first year planting survival, faster growth at the nursery and later in the out plantings, and more resistance to fusiform rust (Wells and Wakeley, 1966; Goggans et al., 1972; Slunder, 1980; Wells et al., 1991). Among various latitude-related climatic variables, variation in precipitation is of particular interest. Latitude-related variation in several traits shows a significant correlation to the cline of decreasing precipitation with increasing latitude (Wells et al., 1991). Byram and Lowe (1988) showed that the average specific gravity of trees from the provenance with lower precipitation (South Arkansas) is greater than that of trees from provenances with higher precipitation (South Louisiana) when grown on a common site. The results from this study support the results from the previous studies. Many stress-related genes, especially drought-stress induced genes, with variation in expression between the two seed sources were identified in this study and most of them are expressed more highly in South Arkansas trees than South Louisiana trees during both growing seasons (Table 5). These differences in gene expression level should be due to genetics because these pines were exposed to the same environment on the same site regardless of the seasonal change. South Arkansas has a lower average precipitation (1170mm-1420mm) than South Louisiana (1420mm-1725mm). Pines native to these regions may have evolved adaptations to the different environments and some differences in gene expression may be due to adaptation to drought stress.

The cis-elements in regulatory sequences are highly polymorphic (Stephens et al., 2001) and these naturally occurring polymorphisms (SNPs) can change in vivo transcription rates causing substantial variation in gene expression between individuals within and among natural populations (Koivisto et al., 1994; Beaty et al., 1995; Crawford et al., 1999; Segal et al., 1999; Oleksiak et al., 2002). Oleksiak et al. (2002) suggested that variation in gene expression among individuals within and among natural populations could be an important mechanism for evolution by natural selection, especially directional selection. The results from this study also support adaptation to the different environments through variation in gene expression.

**Real-time quantitative RT-PCR**

To test the overall validity of our microarray data, selected genes were further examined by real-time quantitative RT-PCR (Heid et al., 1996). Unlike the microarray analyses, samples for different families within each seed source and within different growing seasons were analyzed separately using real-time RT-PCR. A paired t-test was used on $\Delta C_T$ values to test seasonal variation in gene expression within each source. In general, real-time RT-PCR results agree well with our microarray data (Table 7). The log ratio for each gene from the microarray analysis in Table 7 represents the degree of differential expression between earlywood and latewood within each seed source. Positive log ratio values (>0) represent preferential expression in latewood compared to earlywood. The differences in $\Delta C_T$ values between earlywood and latewood ($EW\Delta C_T$ - $LW\Delta C_T$) from the real-time PCR analysis also represent the degree of differential

expression between earlywood and latewood within each seed source. Positive values (>0) represent preferential expression in latewood compared to earlywood. Log ratio values from the microarray data agree relatively well with real-time RT-PCR $\Delta C_T$ mean difference values (Table 7).

An ANOVA was used on $\Delta C_T$ values for each gene selected to test variation in expression between the two seed sources and among families within each seed source within each growing season. In general, real-time RT-PCR results agree well with our microarray data for variation between the two seed sources within each growing season even though an additional factor, different families within each seed source, were included during real-time RT-PCR result analyses (Table 8). Significant variation in expression among families within each seed source was observed for several genes tested. The adjusted $R^2$ values in Table 8 represent the fitness of the nested-factorial model used for the ANOVA explaining how much of the variation in expression for each gene tested was explained by the family structure for each seed source within each growing season. Byram and Lowe (1988) showed that there is a consistent pattern of variation in specific gravity among families within each seed source and little GxE effect on family-within-seed source. The variation in expression of the genes between the two seed sources and among families within each seed source observed in this study also supports the findings from the previous study.

Table 7. Comparison of microarray and real-time RT-PCR results for selected genes identified in the earlywood vs. latewood study.

| | | SAR | | | | SLA | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Microarray | | Real-Time RT-PCR | | Microarray | | Real-Time RT-PCR | |
| Clone ID | Putative Function | Pr > F | log ratio | Pr > \|t\| | ΔCt (EW-LW) | Pr > F | log ratio | Pr > \|t\| | ΔCt (EW-LW) |
| 14 D07 | Putative dehydrine | <0.01 | 2.5 | <.0001 | 2.3 | <0.01 | 1.7 | <.0001 | 2.37 |
| 17 A10 | No hit | <0.01 | 2.2 | <.0001 | 2.61 | | | 0.001 | 1.25 |
| NXSI_079_E12 | AGP 5 | <0.01 | 1.8 | <.0001 | 2.5 | <0.01 | 1.1 | <.0001 | 2.34 |
| NXCI_156_C12 | Adenylate kinase | <0.01 | 1.8 | <.0001 | 1.05 | <0.01 | 1.0 | <.0001 | 1.19 |
| NXSI_054_F05 | Glycine-rich protein homolog | <0.01 | 1.8 | <.0001 | 3.18 | <0.01 | 0.1 | <.0001 | 2.17 |
| NXSI_074_B07 | No hit | <0.01 | 1.7 | <.0001 | 2.1 | <0.01 | 1.6 | <.0001 | 2.56 |
| NXCI_062_B10 | Probable gamma-thionin precursor SPI1 | <0.01 | 1.4 | <.0001 | 2.85 | | | 0.018 | 1.64 |
| NXNV_127_E04 | Isoflavone reductase homolog | <0.01 | 0.9 | <.0001 | 3.17 | <0.01 | 0.6 | <.0001 | 2.7 |
| NXSI_134_F04 | Cellulase (EC 3.2.1.4) 1 precursor | <0.01 | -0.5 | <.0001 | -1.24 | | | 0.848 | 0 |
| NXSI_116_B04 | Heat shock protein 82 | <0.01 | -0.8 | <.0001 | -3.32 | <0.01 | -0.4 | <.0001 | -2.3 |
| 40 F04 | Low molecular weight heat shock protein | <0.01 | -0.8 | <.0001 | -0.76 | <0.01 | -0.8 | 0.012 | -0.52 |
| 14 B10 | Class 1 heat shock protein | <0.01 | -0.8 | <.0001 | -1.64 | <0.01 | -0.9 | <.0001 | -0.95 |
| NXSI_036_F01 | Ethylene-responsive transcriptional coactivator | <0.01 | -1.1 | <.0001 | -4.61 | <0.01 | -0.9 | <.0001 | -3.55 |
| NXNV_156_G02 | AGP3 | | | 0.003 | -0.65 | <0.01 | 0.3 | 0.008 | 0.68 |

Positive log ratio values (>0) represent preferential expression in latewood compared to earlywood within each seed source. Positive ΔCt (EW-LW) values (>0) also represent preferential expression in latewood compared to earlywood within each seed source. SAR, South Arkansas; SLA, South Louisiana; ΔCt (EW-LW), mean difference between earlywood and latewood ΔCt values (=EWΔCt – LWΔCt).

Table 8. Comparison of microarray and real-time RT-PCR results for selected genes identified in the South Arkansas vs. South Louisiana study.

| | | SAR vs SLA | | | | Among families within each seed source | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | EW (Pr > F) | | LW (Pr > F) | | SAR (Pr > F) | | SLA (Pr > F) | | adjusted $R^2$ | |
| Clone ID | Putative Function | Microarray | RT | Microarray | RT | EW | LW | EW | LW | EW | LW |
| NXSI_054_F05 | Glycine-rich protein homolog | <0.01 | <.0001 | <0.01 | <0.0001 | <.0001 | 0.107 | <.0001 | <0.0001 | 0.95 | 0.94 |
| 35 A01 | Metallothionine like protein | <0.01 | <.0001 | <0.01 | <0.0001 | 0.001 | <0.0001 | <.0001 | 0.003 | 0.91 | 0.96 |
| 17 A10 | No hit | <0.01 | <.0001 | <0.01 | <0.0001 | <.0001 | 0.195 | 0.612 | 0.001 | 0.88 | 0.95 |
| NXNV_079_G02 | Probable UDP-glucuronosyltransferase | <0.01 | <.0001 | <0.01 | <0.0001 | 0.394 | 0.001 | <.0001 | 0.073 | 0.89 | 0.84 |
| NXCI_062_B10 | Probable gamma-thionin precursor SPI1 | <0.01 | 0.009 | <0.01 | <0.0001 | 0.688 | 0.095 | 0.018 | 0.404 | 0.39 | 0.89 |
| 40 F04 | Low molecular weight heat shock protein | <0.01 | 0.005 | <0.01 | <0.0001 | 0.483 | 0.052 | 0.058 | 0.001 | 0.37 | 0.66 |
| NXCI_021_D03 | Proline-rich protein. | <0.01 | 0.035 | <0.01 | <0.0001 | 0.654 | <0.0001 | 0.018 | <0.0001 | 0.33 | 0.98 |
| 14 D07 | Putative dehydrine | | 0.104 | <0.01 | 0.05 | 0.005 | <0.0001 | 0.005 | <0.0001 | 0.53 | 0.71 |

SAR, South Arkansas; SLA, South Louisiana; EW, earlywood; LW, latewood; RT, real-time RT-PCR.

# CHAPTER IV

# REAL-TIME RT-PCR ANALYSIS OF LOBLOLLY PINE (*Pinus taeda L.*) ARABINOGALACTAN-PROTEIN-LIKE GENES

## INTRODUCTION

Arabinogalactan-proteins (AGPs) are a class of a large hydroxyproline-rich glycoproteins (HGRPs) found in almost all plant species including angiosperms, gymnosperms, and lower plants such as bryophytes (reviewed by Fincher et al*.,* 1983; Showalter, 1993; Nothnagel, 1997; Majewska-Sawka and Nothnagel, 2000). Protein cores of AGPs, which comprise only two to 10% of the total mass, are abundant in hydroxyproline (Hyp), alanine (Ala), serine (Ser), and threonine (Thr) and are intensively glycosylated mostly by galactose and arabinose (Fincher et al*.,* 1983; Showalter and Varner, 1989; Du et al*.,* 1996). AGPs share common characteristic domains rather than having sequence identify. In general, classical AGP protein cores contain a signal peptide for the secretion pathway, a Hyp-rich domain predicted to be O-glycosylated, and a GPI anchor signal sequence for plasma membrane attachment (Youl et al., 1998; Schultz et al*.,* 1998). Some classical AGPs contain a short domain rich in basic amino acids within a Hyp-rich domain (Gao et al., 1999; Zhang et al*.,* 2003). There are also non-classical AGPs that contain extra Cys-rich or Asn-rich domains in addition to Hyp-rich domains. Non-classical AGPs do not contain a GPI anchor signal sequence

(Majewska-Sawka and Nothnagel, 2000). There are also short AG-peptides (Schultz, 2002). AGPs have been implicated in various plant growth and developmental processes such as somatic embryogenesis (Kreuger and van Holst, 1993, 1995; Thomson and Knox, 1998; Champman et al., 2000), pollen tube growth (Cheung et al., 1995; Roy et al., 1998; Wu et al., 2000), cell proliferation (Serpe and Nothnegel, 1994; Thompson and Knox, 1998), cell expansion (Willats and Knox, 1996; Ding and Zhu, 1997), and cell differentiation (Pennell and Roberts, 1990; Knox et al., 1991). However, specific roles of AGPs in these processes remain to be elucidated.

AGPs are particularly interesting with respect to pine xylem development. A total of six AGP-like genes or gene families have been cloned from differentiating pine xylem. Among them, only AGP3 (Loopstra et al., 2000) and AGP6 (Zhang et al., 2003) have been verified experimentally to be AGPs. Others are classified as AGP-like because they share general characteristics of known AGPs. Two AGP-like genes, ptx3H6 and ptx14A9, were cloned by differential screening of a xylem cDNA library (Loopstra and Sederoff, 1995). They are preferentially expressed in xylem compared to other tissues (Loopstra and Sederoff, 1995) and might be regulated by plant hormones during seedling development (No and Loopstra, 2000). There is some experimental evidence using the genes tagged with the c-myc epitope to transform tobacco that ptx3H6 and ptx14A9 are AGPs (No and Loopstra, unpublished data). PtaAGP3 was cloned as part of an EST project and identified as an AGP following protein purification and N-terminal sequencing. It is preferentially expressed in xylem compared to other tissues as well (Loopstra et al., 2000). PtaAGP4, ptaAGP5, and ptaAGP6, were cloned

from loblolly pine xylem as part of EST on sequencing project (Allona et al., 1998; Zhang et al., 2000). PtaAGP6 was highly expressed in immature xylem from vertical or bent wood. PtaAGP4 was expressed more in xylem from compression wood and side wood than in xylem from vertical wood. PtaAGP5 was expressed more in side wood, less in vertical wood and least in compression wood (Zhang et al., 2000). An immuno-localization study showed that ptaAGP6 expression is restricted to a file of cells that just precede secondary cell wall thickening (Zhang et al., 2003). Most of these pine AGPs or AGP-like genes are among the most abundant transcripts in differentiating xylem tissue (Loopstra et al., 1995; Allona et al., 1998; Whetten et al., 2001; Lorenz and Dean, 2002).

Numerous potential roles of AGPs during xylogenesis have been proposed. For example, AGPs may act as carrier or shuttle molecules that bind newly synthesized wall polymers, keeping them soluble during transport to the cell wall (Gilbeaut and Carpita, 1991). AGPs may also have roles in secondary cell wall initiation and lignification, acting as matrices for orderly addition of nascent wall precursors to the growing wall matrix (Kieliszewski and Lamport, 1994). AGPs are thought to be involved in cell expansion as well (Zhu et al., 1993; Kielizewski and Lamport, 1994; Jauh and Lord, 1996; Willats and Knox, 1996). They may also play a role in programmed cell death (Schindler et al., 1995; Greenberg, 1996; Buckner et al., 1998), cell adhesion (Johnson et al., 2003), and signaling pathways through GPI anchors (Schultz et al., 1998) during xylogenesis. However, the specific role of each AGP during pine xylogenesis remains to be elucidated. It is possible that different AGPs have similar roles but at different times or in different locations. Alternatively, they may have very different functions during

xylogenesis. To better understand the roles of pine AGPs during xylogenesis, gene-specific primers were designed and relative transcript levels of 11 loblolly pine AGP-like genes (ptx3H6, ptx14A9, ptaAGP3, ptaAGP4, ptaAGP6 and six members of the ptaAGP5 multigene family) were examined using real-time RT-PCR analysis in this study. Expression was examined in different tissues, earlywood and latewood, compression, opposite, and vertical woods, drought stressed roots, and *in vitro* cultured cells induced for lignification. The different loblolly pine AGP-like genes had varying expression patterns under the different conditions suggesting different functions for each loblolly pine AGP.

## MATERIALS AND METHODS

### Plant materials

#### 1. Tissue specificity

See materials and methods in Chapter II.

#### 2. Season and origin

See materials and methods in Chapter III.

**3. Drought stress**

Within the natural range of loblolly pine, two populations, The Lost Pines representing pines from a dry area and a South Louisiana population representing pines from a relatively wet area, were selected. The Lost Pines are from Bastrop, TX which has the lowest precipitation (865mm-965mm annual precipitation) within the range of loblolly pine. On the other hand, the South Louisiana pines are from the wettest part of the range (1500mm-1700mm annual precipitation). Six-month-old pine seedlings, six trees per family and three families per each seed source, were drought stressed in the green house. Control pines were watered daily. Pines with three different levels of drought stress, control (0.1-0.5 MPa), moderate (1-1.5 MPa) and highly stressed (2.0-2.5 MPa), were attained over the treatment period. Roots from each pine were harvested after measuring pre-dawn water potential using a Scholander's pressure bomb chamber and frozen immediately in liquid nitrogen and stored at –80 °C until further use.

**4. Compression wood**

Two-year-old pine seedlings were bent for three weeks to produce compression wood. The bark and phloem layers were peeled off and differentiating xylem tissues from compression wood (red-colored) and opposite wood (white-colored) were collected and frozen immediately in liquid nitrogen and stored at –80 °C until further use.

**5. Cell culture**

Dr. Stasolla (U. of Manitoba, Canada) kindly performed the cell cultures and provided the RNAs to us. Suspension cultured cells of *Pinus taeda L.* (*CAD/cad*) were initiated from shoot tips (2 cm in length) using methods described by Brown and Lawrence (1968). Maintenance of cells and induction of secondary cell wall formation and lignin deposition were carried out following Eberhardt et al. (1993). Briefly, cells were maintained in a proliferation medium containing 2,4-dichlorophenoxyacetic acid (11 μM) as a sole source of growth regulators and were subcultured into fresh medium at 7 d intervals. Lignification was initiated by transferring the cells (2.5 ml packed cell volume) into 50 ml of NAA (11 μM)-containing medium (lignification medium). Cells were harvested at day 0, 7, 14, 21, 35 and frozen immediately in liquid nitrogen and stored at –80 °C until further use.

**Identification and analysis of AGP sequences**

"BLASTN reports in nsfpine" were searched with "arabinogalactan" as the query. The contigs for the resulting EST hits were identified in the loblolly pine xyem EST database (http://web.ahc.umn.edu/biodata/nsfpine/contig_dir20/). DNA and protein sequence similarity were analyzed with the ClustralW alignment program (http://www.ebi.ac.uk/clustalw/). N-terminal signal peptide sequences for each AGP were predicted with SignalP (http://www.cbs.dtu.dk/services/SignalP/; Nielsen et al., 1997). The C-terminal trans-membrane domain and GPI anchor attachment for each AGP were predicted with PSORT (http://psort.nibb.ac.jp/form.html; Nakai and Horton

Table 9. Gene specific primers designed for 11 loblolly pine AGP and AGP-like genes for real-time RT-PCR.

| Gene ID | Forward primer | Reverse primer |
|---|---|---|
| pta3h6 | 5'-CAGGTGGTGAAACAATGGCTC-3' | 5'-AGAGGCTGAAGGAGACTGCG-3' |
| pta14a9 | 5'-CCTGTTCGTTCCTGCTTCGT-3' | 5'-CTGTCTGCAACGGAATTCGA-3' |
| ptaAGP3 | 5'-TCCATTGCTGTTTGGCAGATC-3' | 5'-GGCCAAAATGTAGCTCCAGG-3' |
| ptaAGP4 | 5'-AAAGTTGATGATGGCCCCAC-3' | 5'-GATTCCACCTGGGCTGATTCT-3' |
| ptaAGP5A | 5'-GCAGACAAGATGGGCCGAT-3' | 5'-TTCGGCAAAAGTGAGGGTG-3' |
| ptaAGP5B | 5'-GGTTGTGAGTGCTACCCCTAATCT-3' | 5'-GAACGACCCATTATACCAATTAAAGG-3' |
| ptaAGP5C | 5'-AAACTCCGGCATCTGGTCC-3' | 5'-AGAGCCATCTTCTCCATGCTG-3' |
| ptaAGP5D | 5'-CTGCCTCGAAAAACCTCTTCA-3' | 5'-GCTGTGATCAAAAGATACTAGTGGAAA-3' |
| ptaAGP5E | 5'-TTATTCTTCCTGGGCAACGTG-3' | 5'-CTGGTTGTTGCTGACAAATGATAAT-3' |
| ptaAGP5F | 5'-TTTCTCTTGGGCAGATTTGCTT-3' | 5'-TGTCTGCTGCTTGCTGGC-3' |
| ptaAGP6 | 5'-TGGCTCTGCATTGCAAGTTT-3' | 5'-GCAGTTGTGGGTGGCTTAGC-3' |
| 18s rRNA | 5'-AAGACGGACCACTGCGAAAG-3' | 5'-ATCCCTGGTCGGCATCGT-3' |

1999). Additional analyses for GPI anchor sites were performed with big-PI plant predictor (http://mendel.imp.univie.ac.at/sat/gpi/plant_server.html; Eisenhaber et al., 2003).

**Real-time quantitative RT-PCR**

Total RNAs were extracted from each sample using the method of Chang et al. (1993). Residual DNA was removed using DNA-free™ (Ambion Inc. Austin, TX) and RNA was further purified using MEGAclear™ (Ambion Inc. Austin, TX ). 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-specific primers for each AGP were designed based on

sequences from the NSF pine xylem EST Database using Primer Express (Applied Biosystems, CA) (Table 9). Samples and standards were run in duplicate on each plate and repeated on at least two plates using SYBR-Green PCR Master Mix (Applied Biosystems, CA) on a GeneAmp 7900 Sequence Detection System (Applied Biosystems, CA) following the manufacturer's recommendations. Real-time RT-PCR was performed in a 10 µl reaction containing 3.5 µl $ddH_2O$, 5 µl 2x PCR mix, 0.5 µl forward primer (1 µM), 0.5 µl reverse primer (1 µM), and 0.5 µl of template cDNA (10ng/µl). The PCR conditions were two minutes of pre-incubation at 50°C, 10 minute of pre-denaturation at 94 °C, 40 cycles of 15 seconds at 95 °C and one min at 60 °C, followed by steps for dissociation curve generation (30 seconds at 95 °C, 60 seconds at 60 °C and 30 seconds at 95 °C). For data collection, SDS 2.1 (Applied Biosystems, CA) was used. Dissociation curves for each amplicon were carefully examined and each amplicon producing a single dissociation peak was sequenced to confirm the specificity of the primer pair used. Relative transcript levels for each sample were obtained using the "relative standard curve method" (see User Bulletin #2" ABI PRISM 7900 Sequence Detection System for details) and were normalized to the transcript level of 18s RNA of each sample. A paired t-test and an ANOVA were used on normalized transcript levels to test for variation in gene expression among samples for each experiment.

Table 10. Sequence similarity among different members of the ptaAGP5 gene family at the amino acid level.

| % Identity | ptaAGP5A | ptaAGP5B | ptaAGP5C | ptaAGP5D | ptaAGP5E | ptaAGP5F | ptaAGP5G |
|---|---|---|---|---|---|---|---|
| ptaAGP5A | | 66 | 63 | 61 | 62 | 72 | 57 |
| ptaAGP5B | | | 64 | 60 | 76 | 72 | 64 |
| ptaAGP5C | | | | 68 | 66 | 71 | 64 |
| ptaAGP5D | | | | | 60 | 64 | 54 |
| ptaAGP5E | | | | | | 73 | 62 |
| ptaAGP5F | | | | | | | 70 |
| ptaAGP5G | | | | | | | |

Protein sequence similarity was analyzed with the ClustralW alignment program (http://www.ebi.ac.uk/clustalw/).

## RESULTS

### Loblolly pine AGP-like proteins

The loblolly pine xylem EST sequencing project deposited 59,447 ESTs in the database as of November 2003 (http://pinetree.ccgb.umn.edu/). ESTs for AGP-like genes were sought querying “arabinogalactan” in the “BLASTN reports in nsfpine”. The contigs for the resulting EST hits were identified in the loblolly pine xyem EST database (http://web.ahc.umn.edu/biodata/nsfpine/contig_dir20/). Numerous contigs for different AGPs were identified (contigs 8005, 8026 and 7885 for ptaAGP6; contigs 7819, 7620, 7601, 6582, 6524, 6283, 2860 and 221 for ptaAGP5; contigs 7932, 7916 and 7402 for ptaAGP4; contigs 7842, 7704, 5448, 80 and 16 for ptx3H6, contigs 8004, 7979 and 7689 for ptx14A9; contigs 7990, 7873, 7675, 7646 and 6417 for ptaAGP3). Careful

examination of DNA and protein sequence similarity among the contigs for each AGP suggests that there are at least seven different members of the AGP5 gene family with between 54% and 73% similarity at the amino acid level (Table 10). The AGP5 gene family members share several features in common in addition to overall sequence similarities (Figure 6). All of them are predicted to contain putative N-terminal signal peptides. Some of them are predicted to be GPI anchored and contain a hydrophobic C-terminal transmembrane domain, typical of GPI anchored proteins. In addition, Pro residues, which are thought to be post-translationally modified to Hyp and become O-glycosylated *in vivo*, are well conserved among the different members of ptaAGP5 indicating common patterns of glycosylation (Figure 6). The predicted structures of each loblolly pine AGP or AGP-like protein core including those previously described (Loopstra and Sederoff, 1995; Loopstra et al., 2000; Zhang et al., 2000) are summarized in Table 11. Overall, our results are consistent with the previous studies. However, ptaAGP4 was predicted to have a signal peptide and C-terminal transmembrane domain by Zhang et al. (2000). Both the SignalP (http://www.cbs.dtu.dk/ services/SignalP/; Nielsen et al., 1997) and PSORT (http://psort.nibb.ac.jp/form.html; Nakai and Horton 1999) programs used in this study did not predict a signal peptide or a C-terminal transmembrane domain for ptaAGP4. The numbers of ESTs for each putative loblolly pine AGP and AGP-like gene in six different xylem EST libraries are summarized in Table 12.

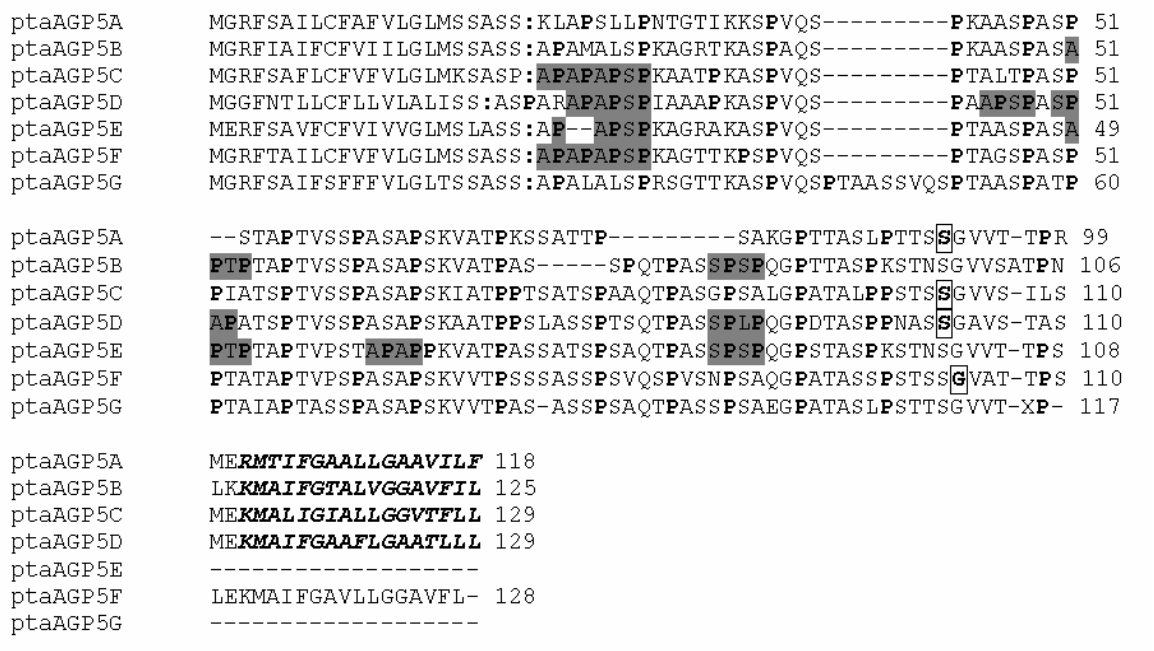


Figure 6. Alignment of predicted amino acid sequences of the ptaAGP5 gene family. The ClustralW alignment program (http://www.ebi.ac.uk/clustalw/) was used for sequence alignment. The C-terminal end sequences for two ptaAGP5 family members, ptaAGP5E and ptaAGP5G, were incomplete. Predicted cleavage sites by signal peptidase for each gene is marked by a colon. Predicted GPI anchored residues are boxed. Predicted C-terminal transmembrane domains are bold and italicized. N-terminal signal peptide sequences for each AGP were predicted with SignalP. The C-terminal trans-membrane domain and GPI anchor attachment for each AGP were predicted with PSORT (http://psort.nibb.ac.jp/form.html; Nakai and Horton 1999).Gray boxes, known glycosylation sequence (A/S P A/S P); bold Ps, Pro residues thought to be post-translationally modified to Hyp and become O-glycosylated *in vivo*.

Table 11. Summary of predicted protein structure of each pine AGP.

| Gene ID | Length (aa) | N-SP | C-TMS | GPI anchored |
|---|---|---|---|---|
| ptx3h6 | 168 | Yes (1-23) | Yes (152-168) | Yes (145) |
| ptx14a9 | 264 | Yes (1-21) | Yes (247-263) | Yes (235) |
| ptaAGP3 | 140 | Yes (1-25) | No | Yes (118) |
| ptaAGP4 | 184 | No | No | No |
| ptaAGP5A | 118 | Yes (1-22) | Yes (102-118) | Yes (92) |
| ptaAGP5B | 125 | Yes (1-22) | Yes (109-125) | No |
| ptaAGP5C | 129 | Yes (1-22) | Yes (113-129) | Yes (103) |
| ptaAGP5D | 129 | Yes (1-19) | Yes (113-129) | Yes (103) |
| ptaAGP5E | 108* | Yes (1-22) | | |
| ptaAGP5F | 128 | Yes (1-22) | No | Yes (104) |
| ptaAGP5G | 117* | Yes (1-22) | | |
| ptaAGP6 | 235 | Yes (1-22) | Yes (215-231) | Yes (208) |

N-terminal signal peptide sequences for each AGP were predicted with SignalP (http://www.cbs.dtu.dk/services/SignalP/; Nielsen et al., 1997). The C-terminal trans-membrane domain and GPI anchor attachment for each AGP were predicted with PSORT (http://psort.nibb.ac.jp/form.html; Nakai and Horton 1999). Additional analyses for GPI anchor sites were performed with big-PI plant predictor (http://mendel.imp.univie.ac.at/sat/gpi/plant_server.html; Eisenhaber et al., 2003). * The C-terminal sequences of ptaAGP5E and ptaAGP5G in the EST database were incomplete. N-SP; N-terminal signal peptide. C-TMS; C-terminal transmembrane sequence.

Table 12. Contigs for each loblolly pine AGP and AGP-like gene identified in the xylem EST database (http://web.ahc.umn.edu/biodata/nsfpine/contig_dir20/) and the number of EST hits for each gene in six different xylem partial cDNA libraries.

| Gene ID | Contig# | EST libraries | | | | | |
|---|---|---|---|---|---|---|---|
| | | NXCI(9333)* | NXLV(10244) | NXNV(8490) | NXPV(9642) | NXRV(10184) | NXSI(11904) |
| ptx3h6 | 7704, 7842, 5448, 8016 | 7(12.7)** | 3(5.4) | 15(27.2) | 12(21.8) | 6(10.9) | 12(21.8) |
| ptx14a9 | 8004, 7979 | 69(56.5) | 0.0 | 11(9.0) | 1(0.8) | 13(10.6) | 28(22.9) |
| ptaAGP3 | 7990, 7873, 7675, 6417, 7646 | 13(9.4) | 28(20.4) | 29(21.1) | 1(0.7) | 39(28.5) | 27(19.7) |
| ptaAGP4 | 7932, 7916, 7402 | 14(36.8) | 1(2.6) | 7(18.4) | (0) | 7(18.4) | 9(23.6) |
| ptaAGP5A | 7819 | 5(18.5) | 0 | 4(14.8) | (0) | 10(37.0) | 8(29.6) |
| ptaAGP5B | 7620 | 2(11.7) | (0) | 1(5.8) | (0) | 10(58.8) | 4(23.5) |
| ptaAGP5C | 7601 | 4(23.5) | 1(5.8) | 2(11.7) | (0) | 8(47.6) | 2(11.7) |
| ptaAGP5D | 6582 | 1(14.3) | (0) | 3(42.8) | 1(14.3) | (0) | 2(28.6) |
| ptaAGP5E | 6524 | 1(14.3) | (0) | 1(14.3) | (0) | 5(71.4) | (0) |
| ptaAGP5F | 6283, 221 | 2(28.6) | (0) | 3(42.8) | 1(14.3) | (0) | 1(14.3) |
| ptaAGP5G | 2860 | 1(50) | (0) | (0) | (0) | (0) | 1(50) |
| ptaAGP6 | 8026, 8005, 7885 | 102(49.5) | 1(0.5) | 16(7.7) | 4(1.9) | 22(10.6) | 61(29.6) |
| | total | 221(34.4) | 34(5.2) | 92(14.3) | 20(3.1) | 120(18.7) | 155(24.1) |

* Figures in the parentheses are the total number of ESTs in each library. **Figures in the parentheses are the percent from each library. NXCI: juvenile compression wood; NXLV: transitional/mature latewood; NXNV: mature normal wood; NXPV: transitional planings wood; NXRV: root wood; NXSI: juvenile side wood.

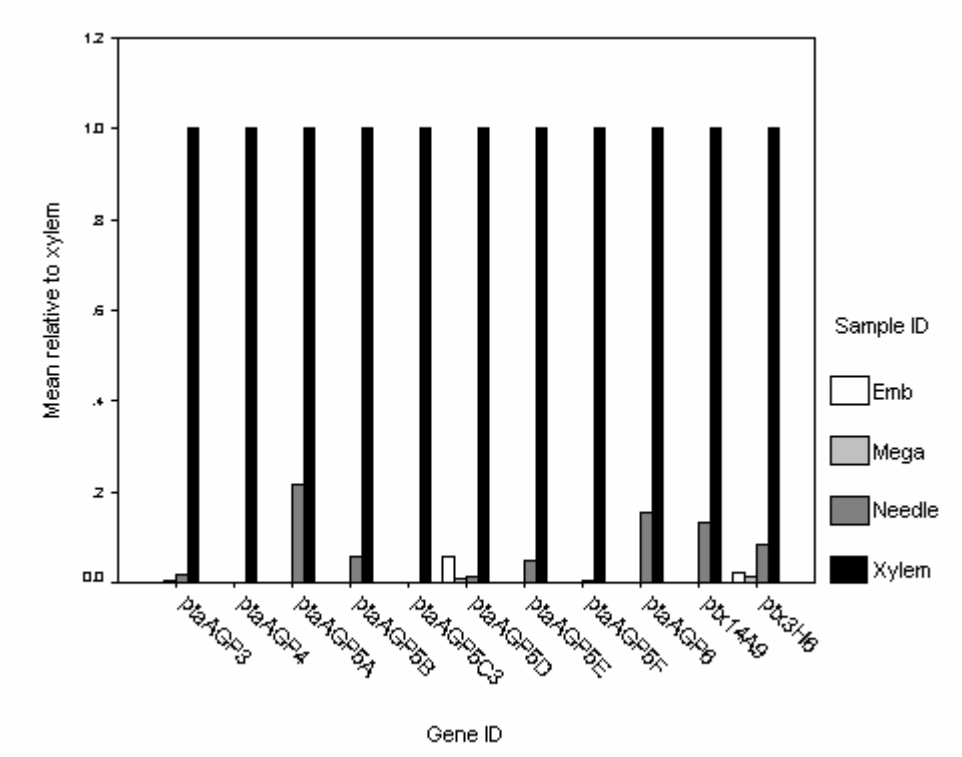


Figure 7. Real-time RT-PCR analysis of loblolly pine AGP and AGP-like genes in different tissues. Emb: embryo; Mega: megagametophyte.

**Real-time RT-PCR analysis of loblolly pine AGPs**

**1. Tissue specificity**

Tissue specific cell wall proteins could play important roles in development of specific cell types or formation of specific structural features of the cell (Keller, 1993; Showalter, 1993; Kieliszewski and Lamport, 1994; Cassab, 1998; Sommer-Knudsen et al., 1998). AGP-like genes are among the most abundant transcripts in wood forming

tissues of loblolly pine and are expected to be preferentially expressed in differentiating xylem tissue compared to other tissues (Loopstra and Sederoff, 1995; Allona et al., 1998; Loopstra et al., 2000; Zhang et al., 2000;Whetten et al., 2001; Lorenz and Dean, 2002; Yang et al., 2004). Transcript levels of 11 different loblolly pine AGP or AGP-like genes from differentiating xylem were compared in xylem, needles, megagametophytes, and embryos using real-time RT-PCR (Figure 7). An ANOVA was used on normalized transcript levels to test the statistical significance of variation in expression among different tissues for each gene. As expected, all of the AGP- like genes were preferentially expressed in differentiating xylem compared to other tissues implying their significant roles during xylogenesis ($P<0.001$).

**2. Season and origin**

Earlywood tracheid cells differ from latewood cells in their chemical composition and physical characteristics. Latewood cells have greater density, smaller lumen, smaller radial diameters and thicker cell walls than earlywood cells (Zimmermann and Brown, 1971, p.96). Commercially important characteristics such as wood specific gravity are known to differ with seed source. For example, when grown on a common site, the specific gravity of Arkansas trees is greater than that of Louisiana trees (Byram and Lowe, 1988). Transcript levels of 11 AGP and AGP-like genes in differentiating xylem from earlywood and latewood within two loblolly pine seed sources, South Arkansas and South Louisiana, were examined using real-time RT-PCR (Figure 8). A paired t-test and an ANOVA were used on normalized transcript levels to

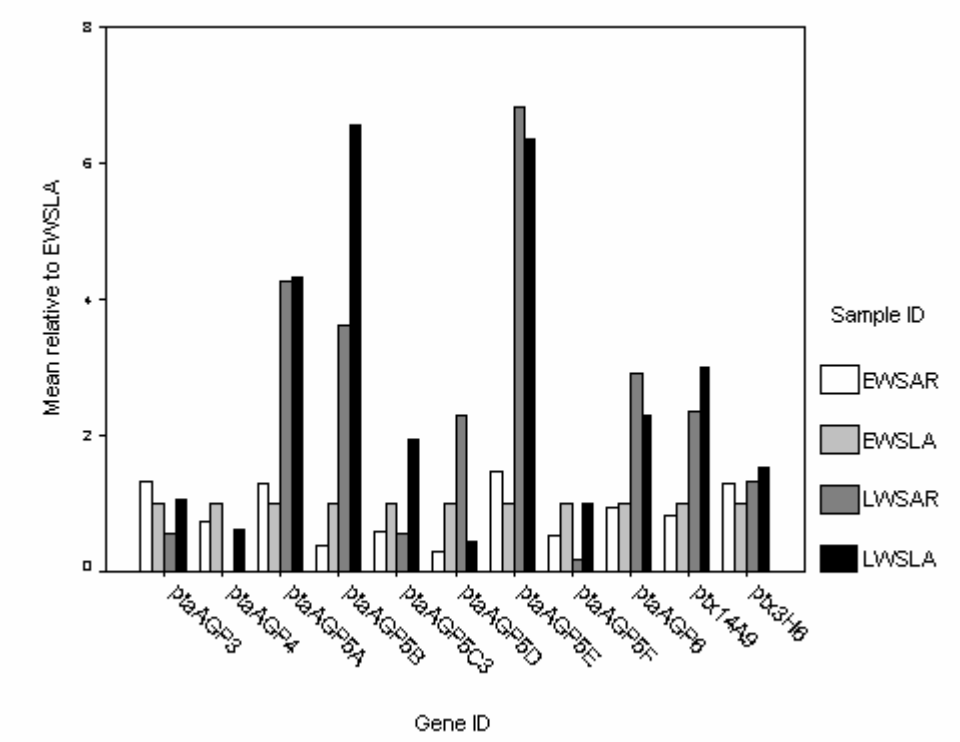


Figure 8. Real-time RT-PCR analysis of loblolly pine AGP-like genes in differentiating xylem during different growing seasons and from different seed sources grown on a common site. EWSAR: earlywood, South Arkansas; EWSLA: earlywood, South Louisiana; LWSAR: latewood, South Arkansas; LWSLA: latewood, South Louisiana.

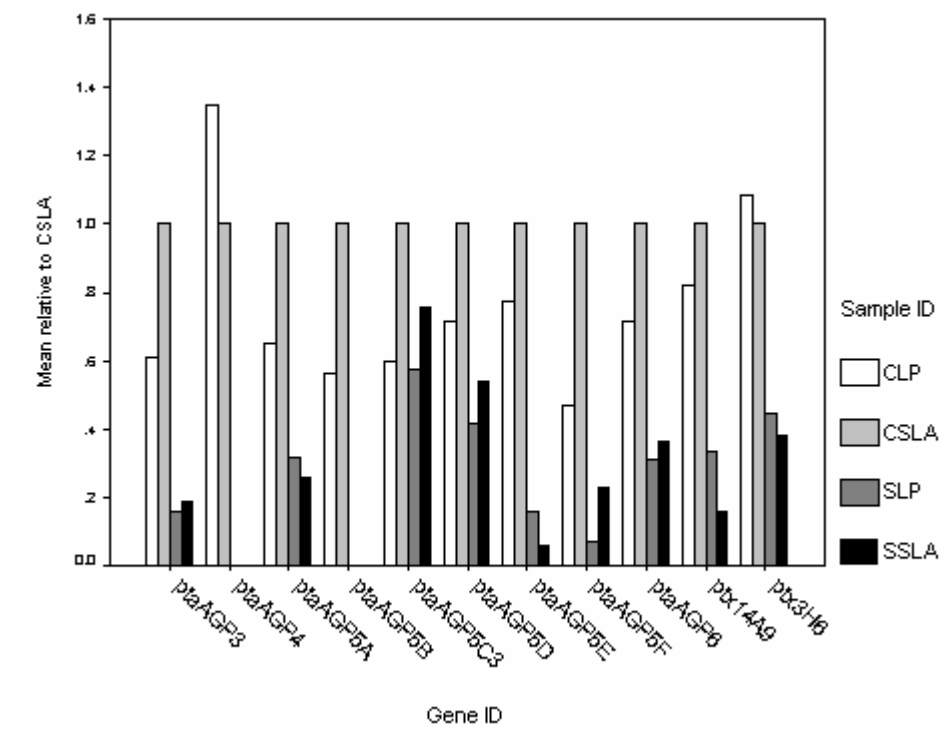


Figure 9. Real-time RT-PCR analysis of loblolly pine AGP and AGP-like genes in drought stressed roots from different seed sources. CLP: control, Lost Pines; CSLA: control, South Louisiana; SLP: stressed, Lost Pines; SSLA: stressed, South Louisiana.

test the statistical significance of variation in expression among the different samples for each gene. A total of nine AGP-like genes within South Arkansas trees (ptx14A9, ptaAGP3, ptaAGP4, ptaAGP5A, ptaAGP5B, ptaAGP5D, ptaAGP5E, ptaAGP5F, and ptaAGP6) and eight genes within South Louisiana trees (ptx3H6, ptx14A9, ptaAGP4, ptaAGP5A, ptaAGP5B, ptaAGP5C, ptaAGP5E, and ptaAGP6) showed significant seasonal variation in expression ($P<0.001$). Most were preferentially expressed in latewood compared to earlywood implying their significant roles during latewood

development (Figure 8). On the other hand, three AGP-like genes within earlywood (ptx3H6, ptaAGP5C, and ptaAGP5F) and eight genes within latewood (ptx3H6, ptaAGP3, ptaAGP4, ptaAGP5B, ptaAGP5C, ptaAGP5D, ptaAGP5F, and ptaAGP6) showed significant variation in expression between the two seed sources ($P<0.001$). In most cases, expression is higher in South Louisiana trees.

**3. Drought stress**

Drought stress can cause serious problems to the forest industry. For example, seedlings are very susceptible to drought stress and significant numbers (up to 65%) of loblolly pine seedlings planted in Texas are lost due to drought during bad years (Phillips, 1998). The growth rate during droughts is far less than normal and repeated droughts can cause severe reductions in yield. Transcript levels of 11 AGPs and AGP-like genes in well-watered seedling roots were compared with drought-stressed seedling roots from two loblolly pine populations, The Lost Pines and South Louisiana, using real-time RT-PCR (Figure 9). The Lost Pines are from Bastrop, TX which has the lowest precipitation (865mm-965mm annual precipitation) within the range of loblolly pine. On the other hand, South Louisiana pines are from wettest part of the range (1520mm-1720mm annual precipitation). An ANOVA was used on normalized transcript levels to test the statistical significance of variation in expression among different samples for each gene. The transcript levels of all the genes except ptaAGP5C were significantly decreased in drought stressed roots compared to control roots within both populations ($P<0.001$). The transcript levels of seven AGP-like genes (ptaAGP3, ptaAGP4,

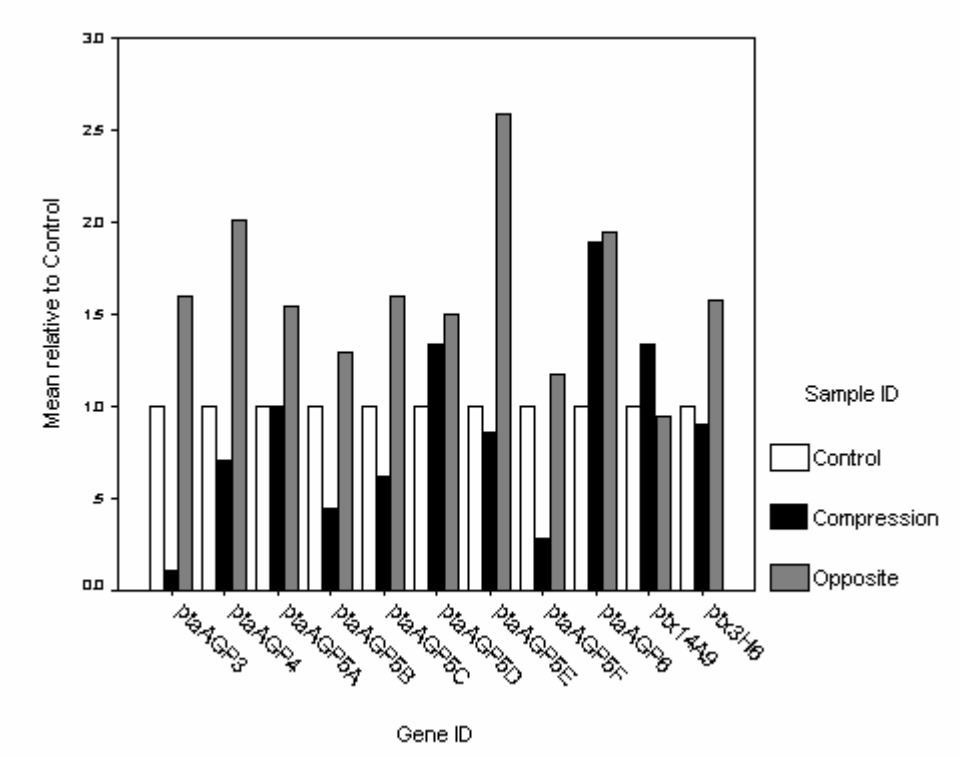


Figure 10. Real-time RT-PCR analysis of loblolly pine AGP and AGP-like genes in differentiating xylem under mechanical and gravitational stress.

ptaAGP5A, ptaAGP5B, ptaAGP5D, ptaAGP5F, and ptaAGP6) were significantly different between the two populations within control roots. Expression was greatest in South Louisiana trees for all of these genes except ptaAGP4. On the other hand, the transcript levels of none of the genes were significantly different between the two populations within stressed seedling roots ($P<0.001$).

**4. Compression wood**

Compression wood produced by mechanical and gravitational stress differs from normal wood in cell shape, wall thickness, morphology of intercellular spaces, lignin composition and lignin content (Timell, 1986). Previous studies showed that transcript levels of several loblolly pine AGP-like proteins were modified in response to mechanical and gravitational stimuli (Zhang et al*.,* 2000; Whetten et al., 2001). In this study, transcript levels of 11 AGP and AGP-like genes in differentiating xylem from control seedlings (vertical wood) were compared with compression wood and opposite wood using real-time RT-PCR (Figure 10). An ANOVA was used on normalized transcript levels to test the statistical significance of variation in expression among different samples. Transcript levels of three genes (ptaAGP3, ptaAGP5B, and ptaAGP6) were significantly different between control and compression wood ($P<0.001$). Among them, only ptaAGP6 was up-regulated in compression wood compared to the control. On the other hand, transcript levels of five genes (ptx3H6, ptaAGP3, ptaAGP5A, ptaAGP5E and ptaAGP6) were significantly different between control and opposite wood ($P<0.001$) with expression higher in the opposite wood. Between compression and opposite woods, transcript levels of five genes (ptx3H6, ptaAGP3, ptaAGP5A, ptaAGP5B and ptaAGP5E) were significantly different ($P<0.001$). Overall, these AGP and AGP-like genes were up-regulated in opposite wood compared to compression wood and control wood (Figure 10).

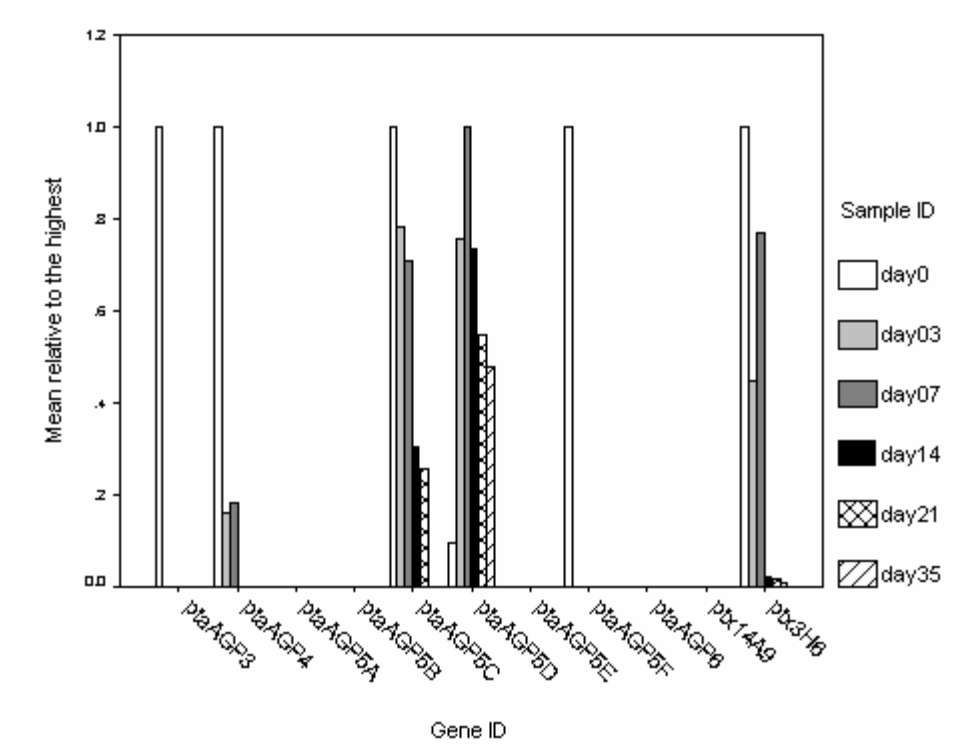


Figure 11. Real-time RT-PCR analysis of loblolly pine AGP and AGP-like genes in cell culture system after the induction for lignification.

**5. Cell culture**

A previous study (Eberhardt et al., 1993) showed that secondary cell wall formation and lignin deposition can be induced in a loblolly pine cell culture system through manipulation of the culture medium (see materials and methods for details). Cells in the liquid culture contain only a thin primary cell wall before the induction (day 0). However, cells begin to accumulate phenolic compounds in the cell wall after the induction (Day 14) and increase in thickness of cell wall and lignification occur between days 21 and 35 (Stasolla et al., 2003). In this study, changes in transcript levels of

loblolly pine AGP and AGP-like genes in cultured cells after the induction of lignification were monitored using real-time RT-PCR (Figure 11). An ANOVA was used on normalized transcript levels to test the statistical significance of variation in expression among different samples harvested on different days following induction. Interestingly, five genes (ptx14A9, ptaAGP5A, ptaAGP5B, ptaAGP5E, and ptaAGP6) were not expressed at all in the cultured cells, even after the induction. Transcripts of some of these AGP-like genes are expected to be among the most abundant transcripts in wood forming tissues. Two genes (ptaAGP3 and ptaAGP5F) were expressed in the maintaining media (day 0) but were rapidly shut down after the induction (day 3). On the other hand, three genes (ptx3H6, ptaAGP4, and ptaAGP5C) were expressed in the maintaining media (day 0) but were more gradually shut down after the induction. PtaAGP5D was the only one whose expression was increased after the induction. The transcript level of ptaAGP5D increased gradually after the induction until day 7 and decreased gradually afterwards (Figure 11).

**Ptx14A9 orthologs in other species**

To find proteins with significant sequence similarity to pine AGPs and AGP-like proteins from other species, the predicted amino acid sequences of each loblolly pine AGP-like protein were blasted against the public databases using BLASTP (http://www4.ncbi.nlm.nih.gov/BLAST/). The "Filtering low complexity regions" option was not used due to the highly repetitive nature of AGP-like protein sequences. Many putative orthogs of ptx14A9 with significant sequence similarities (using an E-value

cutoff of $10^{-20}$) from other species were identified. However, only a few proteins with significant sequence similarities were identified for the other AGP-like proteins using the same restrictions or with even less stringent conditions. The BLASTP search results also identified a known conserved domain, FAS1 (Four repeated domains in the Fasciclin I family of proteins) within a pta14A9 protein core (Figure 12A). The fasciclin 1 (Fas1) from fruit fly is thought to be involved in homophilic adhesion (Elkins et al., 1990). The ClustralW alignment program (http://www.ebi.ac.uk/clustalw/) was used with manual adjustment for multiple sequence alignment of ptx14A9 with putative orthologs from other species. Figure 12B shows a multiple sequence alignment of fasciclin domains of ptx14A9 and putative ptx14A9 orthologs from cotton, poplar, arabidopsis and rice. FAS1 domain sequences from these species were highly conserved and previously identified conserved regions common to all fasciclin domains, H1 and H2 (Johnson et al., 2003), could also be identified. Northern blot analysis previously performed in our lab showed that a putative ptx14A9 ortholog in poplar (Pop 14A9, GeneBank Acc# AF183809) is highly preferentially expressed in differentiating xylem compared to phloem or leaves. A putative ptx14A9 ortholog in Arabidopsis (AtFLA12, GeneBank Acc# NM_125442) was expressed in stems but not in leaves (data not shown). This results suggest that ptx14A9 orthologs from other plant species may play similar role as ptx14A9 especially during xylogenesis.

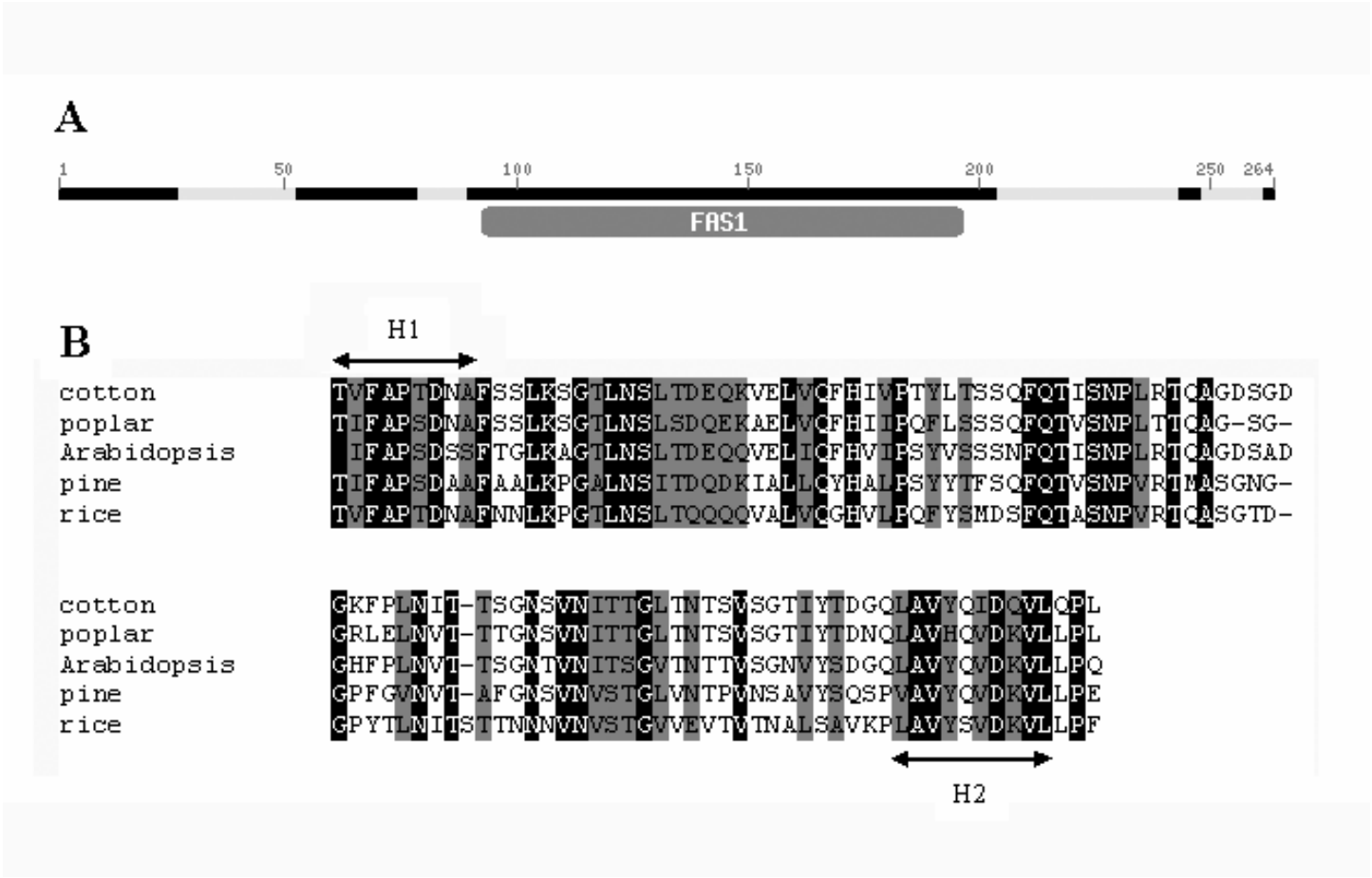


Figure 12. Multiple sequence alignment of fasciclin domains of ptx14A9 and putative ptx14A9 orthologs from several other plant species. Amino acids identical in all 5 proteins are shaded in black and conserved amino acids are shaded in dark gray. H1 and H2: previously identified conserved regions common to all fasciclin domains (Johnson et al., 2003). Cotton = AGP (Accession # AAO92753), Poplar = Pop14A9 (Accession # AF183809), Arabidopsis = FLA12 (Accession # NM_125442), Pine = ptx14A9 (Accession # U09556), Rice = AGP-like protein (Accession # NM_191933).

## DISCUSSION

Complex gene families are relatively common in pine (Kinlaw and Neale, 1997) and different members of a multigene family may be regulated differently in different tissues and at different developmental stages. We identified multiple contigs for each loblolly pine AGP or AGP-like gene (Table 12). These contigs could represent different alleles of the gene or different members of a multigene family (Whetten et al*.,* 2001). There could be as many as 12 different alleles for each gene within the loblolly pine xylem EST project if each of the six libraries was made from one heterozygous individual. Allelic variation within pine is very high making it difficult to distinguish allelic variation from different members of a multigene family without genetic segregation data (Whetten et al., 2001). After careful examination of DNA and protein sequence similarities among the contigs for each gene, we concluded that there are at least seven different members of the ptaAGP5 gene family with between 54% and 73% identity at the amino acid level (Table 10). The putative ptaAGP5 multigene family members share several features in common in addition to overall sequence similarities (Figure 6). Gene specific primers were designed for six different members of ptaAGP5. Real-time RT PCR analysis suggests different temporal and spatial regulation for the members of the ptaAGP5 multigene family. Duplicated ptaAGP5 genes might have evolved new regulatory sequences for new patterns of gene expression in different tissues at different developmental stages or in response to different environmental signals. Multigene families are believed to provide tighter and more sophisticated

control in response to the environment (Pickett and Meeks-Wagner, 1995; McAdams and Arkin, 1999; Meagher et al., 1999). It was also suggested that functional gene duplications could strongly influence the direction of evolution and adaptation (Kinlaw and Neale, 1997).

AGPs are a class of proteins that share common characteristic domains instead of having sequence identity. Thus, different AGPs could have similar roles but at different times or in different locations. Alternatively, they could have very different functions. The different expression patterns of the loblolly pine AGPs observed in this study suggest different functions for each loblolly pine AGP. An examination of publicly available Arabidopsis microarray data also suggested that different AGPs could have different functions (Schultz et al., 2002). Previous studies showed that loblolly pine AGPs are differentially regulated in response to various stimuli such as gravitational stress, hormone inhibitors and seasonal change (Loopstra and Sederoff, 1995; Whetten et al*.,* 1998; No and Loopstra, 2000; Loopstra et al., 2000; Zhang et al*.,* 2000; Whetten et al*.,* 2001; Zhang et al*.,* 2003; Yang et al*.,* 2004). Some expression studies performed previously were reconfirmed in this study. PtaAGP5 gene family members are newly identified and were examined for the first time in this study.

Numerous potential roles of AGPs during xylogenesis have been proposed. For example, AGPs may act as carrier or shuttle molecules that bind newly synthesized wall polymers, keeping them soluble during transport to the cell wall (Gilbeaut and Carpita 1991). Latewood cells have greater density, smaller lumen, smaller radial diameter and thicker cell wall than earlywood cells (Zimmermann and Brown, 1971, p.96). Thus,

AGPs expressed preferentially in latewood compared to earlywood (Figure 8) could be involved in these differences.

Kieliszewski and Lamport (1994) proposed that AGPs may play a role in secondary cell wall initiation by intercalating phenolics for subsequent orderly polymerization. Compression wood cells are different from normal wood cells in that they have a round shaped cross-section compared to the rectangular shape of normal wood cells, a higher ratio of secondary wall thickness to cell diameter, decreased cell length, increased microfibril angle, increased p-hydroxy phenyl subunit content in lignin, and increased lignin content (Timell, 1986). Thus, AGPs expressed preferentially in compression wood compared to normal wood (Figure 10) may play a role in lignification. However, an immuno-localization study showed that ptaAGP6 is not tightly linked to lignification (Zhang et al*.,* 2003) even though it was preferentially expressed in both compression wood and opposite wood compared to normal wood (Figure 10). Thus, the genes preferentially expressed in compression wood compared to normal wood could play a role in stress responses as well (Zhang et al*.,* 2003). A previous study showed that the composition of lignin produced by in vitro loblolly pine cell culture is very similar to the lignin composition in vivo (Stasolla et al*.,* 2003). Thus, AGPs with a change in transcript levels in cultured cells near the induction of lignification could play a role in lignification. The transcript level of ptaAGP5D increased gradually after the induction until day 7 and decreased gradually afterward (Figure 11). However, five AGP-like genes (ptx14A9, ptaAGP5A, ptaAGP5B, ptaAGP5E, and ptaAGP6) were not expressed at all in the cultured cells even after the

induction (Figure 11). Transcripts of some of these AGP or AGP-like genes are expected to be among the most abundant transcripts in wood forming tissues. Thus, the cell culture system used in this study could be an excellent system to study lignin biosynthesis *in* vitro (Stasolla et al., 2003) but the cells might not have been transdifferentiated into true xylem tracheids after the induction and might not represent the true nature of tracheid cells during xylogenesis. In addition, differentiating xylem and developing cotton fibers are very similar in that both developmental processes involve cell elongation, deposition of a secondary cell wall and programmed cell death except that cotton fibers are not lignified. Thus, AGPs expressed during cotton fiber development play other roles besides lignification. A putative ptx14A9 ortholog exists in cotton and is expressed in fibers.

The transcript levels of most of the loblolly pine AGP and AGP-like genes were significantly decreased in drought stressed roots compared to control roots within both populations examined in this study. Plant hormones, especially auxin, have been suggested as important factors involved during xylogenesis (Uggla et al*.,* 2001; Mellerowicz et al*.,* 2001). In addition, a previous study in our lab showed that two AGP-like genes, ptx3H6 and ptx14A9, are differentially regulated during seedling development and these differences may be mediated by different hormonal signaling (No and Loopstra, 2000). Abscisic acid (ABA) is an important mediator of responses to drought stress throughout the tree (Roberts and Dumbroff, 1986) and drought inhibits tracheid production and reduces tracheid radial width in *Pinus radiata* (Jenkins, 1974). ABA is a well known antagonist against auxin and giberellin as well. Previous studies

also showed that some arabidopsis AGPs are differentially regulated by ABA (Schultz et al., 2002; Johnson et al., 2003). Thus, decreases in transcript levels of these AGP-like genes in drought stressed roots could be mediated by ABA. However, a previous study also showed no ABA mediated reduction in expression of ptx3H6 and ptx14A9 (No and Loopstra, 2000). Thus, the decrease in transcript levels of some AGPs in response to drought stress could be mediated by ABA but could also be a general response to environmental stress.

General features of protein cores of AGPs include a signal peptide for entering into the secretion pathway, a Hyp-rich domain for glycosylation, and a GPI anchor site for plasma membrane attachment. All the loblolly pine proteins in this study except ptaAGP4 were predicted to contain N-terminus signal peptides (Table 11). However, ptaAGP4 was predicted to have signal peptide and C-terminal transmembrane domain by Zhang et al. (2000). All the known AGPs contain N-terminus signal peptides. Thus, further verification of the identity of ptaAGP4 might be necessary. Most of the pine AGP and AGP-like proteins analyzed in this study were predicted to contain C-terminus transmembrane domains and GPI anchor sites except ptaAGP4, ptaAGP5B, ptaAGP5E, and ptaAGP5G (Table 11). However, C-terminus sequences in the current database for ptaAGP5E and ptaAGP5G are incomplete. Therefore, they may also contain GPI anchor sites. The C-terminus hydrophobic transmembrane domains of GPI-anchored proteins are thought to function as a recognition site for a transamidase, which cuts the hydrophobic domain and transfers the protein to a prefabricated GPI anchor (Kinoshita and Inoue, 2000). The GPI anchor sites of these AGPs form a covalent bondage via

phosphoethanolamine and a conserved glycan to phosphatidylinositol or ceramide (Kinoshita and Inoue, 2000). The GPI anchor can be cleaved by a specific phospholipase releasing the protein into the extra-cellular matrix (ECM) as well (Griffith and Ryan, 1999). Thus, these GPI anchored AGPs are attached to the plasma membrane or at least transiently located on the plasma membrane (Schultz et al., 1998). Several possible roles of GPI anchoring of proteins were reviewed by Borner et al. (2002) including targeting and signal transduction. The pattern of glycosylation, especially O-linked AG glycosylation, has been implicated to play a critical role in AGP function because many putative AGP functions have been proposed using monoclonal antibodies against AGP carbohydrate epitopes or Yariv reagents (Yariv et al., 1967; Knox et al., 1991; Pennell 1992). AG glycomodules in AGPs contain large, heterogeneous, and highly branched glycan chains. The precise structure and biosynthetic mechanism of these carbohydrate chains has not yet been elucidated (Cassab, 1998; Majewska-Sawka and Nothnagel 2000). According to the contiguity hypothesis (Kieliszewski and Lamport, 1994), clustered noncontiguous Hyps in the AG glycomodules provide sites for complex AG-type chain attachment and clustered contiguous Hyps provide sites for arabinosylated glycomodules. This hypothesis was proven by Shpak et al. (1999; 2001) using synthetic oligopeptides. Several possible functions of AG glycosylation were reviewed by Borner et al. (2002). It could function as a physical link between the protein and ECM (Kohorn 2000) or could provide sites for substrates and ligand binding. AG glycosylation could function as a protective shield from other proteins such as proteases (Kielizewski, 2001) as well. In addition, the carbohydrate chain could provide a signal for correct targeting

of proteins through the secretion pathway to a final destination such as a specific part of the cell surface (Benting et al. 1999; Hauri et al. 2000).

One of the disadvantages of using conifers as a model system is that it is almost impossible to confirm the function of a gene of interest using reverse genetics although transient assays, such as GUS assays for promoter studies, are plausible. However, there is significant homology of expressed genes from wood-forming tissues of loblolly pine with arabidopsis genes (Kirst et al., 2003). Functional analyses using other model systems such as Arabidopsis or poplar could be useful if there are orthologs for pine candidate genes in these species. However, conifer-specific or gymnosperms-specific genes should be analyzed in pine. Schultz et al. (2002) identified 13 classical AGPs, 10 AG-peptides, three basic AGPs that contain a short lysine-rich domain, and 21 fasciclin-like AGPs (FLAs) in the arabidopsis genome. Mutants for these genes may play important roles in determining the role of AGPs. Nam et al. (1999) identified one AGP mutant, *rat1* (resistant to Agrobacterium transformation), with T-DNA tagging. There are also two Arabidopsis mutants with decreased AGP content, *dim* (dimmuto) (Takahashi et al., 1995) and *reb1-1* (root epidermal bulger) (Ding and Zhu 1997). A recently identified mutation in FLA4, *sos5* (salt overly sensitive), showed root swelling and root growth arrest when grown on a high-salt medium (Shi et al., 2003). To find proteins from other species with significant sequence similarity to the pine AGP-like proteins, predicted amino acid sequences of each loblolly pine AGP-like protein were blasted against the public database using BLASTP (http://www4.ncbi.nlm.nih.gov/BLAST/). Many putative orthologs of ptx14A9 with significant sequence similarities

(using an E-value cutoff of $10^{-20}$) from other species were identified. Since AGP sequences are very diverse and gymnosperms and angiosperms are thought to have last shared a common ancestor over 300 million years ago (Bousquet et al*.,* 1992), it is somewhat surprising to find the high levels of similarity. Putative ptx14A9 orthologs in other species could be found in different cell types. One arabidopsis ortholog encodes a putative pollen surface protein, a maize ortholog was endosperm-specific, and a tomato ortholog was expressed in the ovary. A previous study in our lab showed that, like ptx14A9, a putative ptx14A9 ortholog in poplar (Pop 14A9, GeneBank Acc# AF183809) was preferentially expressed in differentiating xylem. In arabidopsis, there are multiple putative ptx14A9 orthologs (21 FLAs) with varying degrees of sequence similarity to ptx14A9. AtFLA12 (GeneBank Acc# NM_125442) was preferentially expressed in stems indicating primary expression in vascular tissues like ptx14A9 and pop14A9. These observations suggest that some ptx14A9 orthologs from other plant species may have a similar role as ptx14A9, especially during xylogenesis. The ptx14A9 protein core contains a putative cell adhesion domain, FAS1 (four repeated domains in the Fasciclin I family of proteins), in addition to a Hyp-rich glycosylation region (Figure 12A). The FAS1 domain consists of two long repeats and contains two highly conserved regions (H1 and H2) of approximately 10 amino acids each (Kawamoto et al*.,* 1998; Schultz et al., 2000; Johnson et al., 2003). Figure 12B shows a multiple sequence alignment of fasciclin domains of ptx14A9 and putative ptx14A9 orthologs from several other plant species. FAS1 domain sequences from these species were highly conserved among these species suggesting the conservation of cell adhesion function (Figure 12B).

# CHAPTER V

# CONCLUSIONS

Xylogenesis is a complicated process where tight temporal and spatial regulation of the expression of specific sets of genes by various factors are involved. Our knowledge about this important biological process, especially the molecular basis of it, has been limited. This study is an effort to understand the molecular basis for wood formation on a genome-wide scale. In this study, the effects of various developmental and environmental factors on gene expression during xylogenesis were examined using microarray and real-time RT-PCR analyses.

In Chapter II, many genes preferentially expressed in differentiating xylem compared to other tissues such as megagametophyte, needle, and embryo were identified. There was a significant increase in the percentage of cell wall-related genes that are preferentially expressed in xylem (20%) compared to the percentage on the array (10%). In Chapter III, the effects of two important factors, different growing seasons and geographical sources, on gene expression during xylogenesis were examined. Many genes preferentially expressed in latewood compared to earlywood were for proteins involved in cell wall biosynthesis. Variation in gene expression among trees from the two seed sources within each growing season suggests that there may be more differences between South Arkansas trees and South Louisiana trees within latewood than within earlywood. Finally, variation in gene expression among trees from different

regions may reflect adaptation to different environments. In Chapter IV, seven different members of the ptaAGP5 gene family with between 54% and 73% identity at the amino acid level were newly identified. Gene specific primers were designed and relative transcript levels of 11 loblolly pine AGP and AGP-like genes were examined in various conditions using real-time RT-PCR analysis. Varying expression patterns for different AGPs and AGP-like genes under the different conditions observed in this study suggest different functions for each loblolly pine AGP. In addition, significant sequence similarities among putative ptx14A9 orthologs from other plant species and an expression study for ptx14A9 orthologs from poplar and Arabidopsis suggest that other systems suitable for reverse genetics such as arabidopsis and poplar could be useful for functional verification of pine AGPs in the future.

The expression of most of the genes examined in this study was not previously studied for tissue specificity, different growing seasons, geographical source, and other conditions examined. Based on the results from this study, candidate genes may be further studied for association with significant traits, used for genetic modification of wood properties, or included in future studies to further examine the molecular mechanisms of wood formation.

## REFERENCES

**Allona I, Quinn M, Shoop E, Swope K, Carlis J, Riedl J, Retzel E, Campbell MM, Sederoff RR, Whetten RW** (1998) Analysis of xylem formation in pine by cDNA sequencing. Proc Natl Acad Sci USA **95**: 9693-9698

**Bailey IW** (1952) Biological processes in the formation of wood. Science **115**: 255-259

**Baker JB, Langdon OG** (1990) *Pinus taeda L*. Loblolly pine. *In* Silvics of North America, Vol. 1. Conifers. Agriculture Handbook 654. Forest Service, United States Department of Agriculture, Washington, DC, pp. 497-512

**Bao W, O'Malley DM, Whetten R, Sederoff RR** (1993) A laccase associated with lignification in loblolly pine xylem. Science **260**: 672-674

**Baskin TI, Wilson JE, Cork A, Williamson RE** (1994) Morphology and microtubule organization in *Arabidopsis* roots exposed to oryzalin or taxol. Plant Cell Physiol **35**: 935-942

**Beaty JS, West KA Nepom GT** (1995) Functional effects of a natural polymorphism in the transcriptional regulatory sequence of HLA-DQB1. Mol Cell Biol **15**: 4771-4782

**Benting JH, Rietveld AG, Simons K** (1999) N-Glycans mediate the apical sorting of a GPI-anchored, raft-associated protein in Madin-Darby canine kidney cells. J Cell Biol **146**: 313-320

**Borner GHH, Sherrier DJ, Stevens TJ, Arkin IT, Dupree P** (2002) Prediction of Glycosylphosphatidylinositol-anchored proteins in Arabisopsis. A genomic analysis. Plant Physiol **129**: 486-499

**Bousquet J, Strauss SH, Doerksen AH, Price RA** (1992) Extensive variation in evolutionary rate of rbcL gene sequences among seed plants. Proc Natl Acad Sci USA **89**: 7844-7848

**Brown CL, Lawrence RH** (1968) Culture of pine callus on a defined medium. Forest Sci **14**: 62–64

**Buckner B, Janick-Buckner D, Gray J, Johal GS** (1998) Cell-death mechanisms in maize. Trends Plant Sci **3**: 218-223

**Byram TD, Lowe WJ** (1988) Specific gravity variation in a loblolly pine seed source study in the western gulf region. For Sci **34**: 798-803

**Cassab GI** (1998) Plant cell wall proteins. Annu Rev Plant Physiol Plant Mol Biol **49**: 281-309

**Chaffey N** (2002) Why there is so little research into the cell biology of the secondary vascular system of trees? New Phytologist **153**: 213-223

**Champman A, Blervacq AS, Vasseur J, Hilbert JL** (2000) Arabinogalactan-proteins in *Cichorium* somatic embryogenesis: Effects of β-glucosyl Yariv reagent and epitope localization during embryo development. Planta **211**: 305-314

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

**Cheung AY, Wang H, Wu HM** (1995) A floral transmitting tissue-specific glycoprotein attracts pollen tubes and stimulates their growth. Cell **82**: 383-393

**Cosgrove DJ** (1997) Assembly and enlargement of the primary cell wall in plants. Annu Rev Cell Dev Biol **13**: 171-201

**Costa P, Pionneau C, Bauw G, Dubos C, Bahrmann N, Kremer A, Frigerio JM, Plomion C** (1999) Separation and characterization of needle and xylem maritime pine proteins. Electrophoresis **20**: 1098-1108

**Crawford DL, Segal JA, Barnett JL** (1999) Evolutionary analysis of TATA-less proximal promoter function. Mol Biol Evol **16**: 194-207

**Delmer DP, Amor Y** (1995) Cellulose biosynthesis. Plant Cell **7**: 987-1000

**DiGiovanni F, Kevan PG, Arnold J** (1996) Lower planetary boundary layer profiles of atmospheric conifer pollen above a seed orchard in northern Ontario, Canada. For Ecol Manage **83**: 87-97

**Ding L, Zhu JK** (1997) A role for arabinogalactan-proteins in root epidermal cell expansion. Planta **203**: 239-294

**Dodd RS, Fox P** (1990) Kinetics of tracheid differentiation in Douglas-fir. Ann Bot **65**: 649-657

**Du H, Simpson RJ, Mortiz RL, Clarke AE, Bacic A** (1994) Isolation of the protein backbone of an arabinogalactan-protein from the styles of *Nicotiana alata* and characterization of a corresponding cDNA. Plant Cell **6**: 1643-1653

**Eberhardt TL, Bernards MA, He L, Davin LB, Wooten JB, Lewis NG** (1993) Lignification in cell suspension cultures of *Pinus taeda*. J. Biol. Chem **268**: 21088–21096

**Eisenhaber B, Wildpaner M, Schultz CJ, Borner GHH, Dupree P, Eisenhaber F** (2003) Glycosylphosphatidylinositol lipid anchoring of plant proteins: Sensitive prediction from sequence- and genome-wide studies for Arabidopsis and rice. Plant Physiol **133**: 1691-701

**Elkins T, Zinn K, McAllister L, Hoffmann FM, Goodman CS** (1990) Genetic analysis of a Drosophila neural cell adhesion molecule: Interaction of fasciclin 1 and abelson tyrosine kinase mutations. Cell **60**: 565-575

**Fincher GB, Stone BA, Clarke AE** (1983) Arabinogalctan-proteins: Structure, biosynthesis, and function. Annu Rev Plant Physiol **34**: 47-70

**Gao M, Kieliszewski MJ, Lamport DTA, Showalter AM** (1999) Isolation, characterization and immunolocalization of a novel, modular tomato arabinogalactan-protein corresponding to the LeAGP-1 gene. Plant J **18**: 43-55

**Giddings TH Jr, Staehelin LA** (1988) Spatial relationship between microtubules and plasma-membrane rosettes during the deposition of primary wall microfibrils in *Closterium sp*. Planta **173**: 22-30

**Gilbeaut DM, Capita NC** (1991) Tracing cell wall biogenesis in intact cells and plants, selective turnover and alteration of soluble and cell wall polysaccharides in grasses. Plant Physiol **97**: 551-561

**Gilmore AR, Boyce SG, Ryker RA** (1966) The relationship of specific gravity of loblolly pine to environmental factors in southern Illinois. For Sci **12**: 399-405

**Goggans JF, Lynch KF, Garin GI** (1972) Early results of a loblolly pine seed source study in Alabama. Auburn Univ Agric Exp Stn Res Pap SO-89 Auburn AL 10p

**Greenberg JT** (1996) Programmed cell death: A way of life for plants. Proc Natl Acad Sci **93**: 12094-12097

**Griffith OH, Ryan M** (1999) Bacterial phosphatidyl-inositol-specific phospholipase C: Structure, function, and interaction with lipids. Biochim Biophys Acta **1441**: 237-254

**Hauri H, Appenzeller C, Kuhn F, Nufer O** (2000) Lectins and traffic in the secretory pathway. FEBS Lett **476**: 32-37

**Haygreen JG, Bowyer JL** (1996) Specific gravity and density. *In* Forest products and wood science. 3rd edition. Iowa State University Press, Ames. pp191-212

**Heath LS, Ramakrishnan N, Sederoff RR, Whetten R, Chevone BI, Struble CA, Jouenne VY, Chen DW, van Zyl L, Grene R** (2002) Studying the functional genomics of stress responses in loblolly pine with the expresso microarray experiment management system. Comparative and Functional Genomics **3**: 226-243

**Heid CA, Stevens J, Livak KJ, Williams PM** (1996) Real time quantitative PCR Genome Res **6**: 986-94

**Hertzberg M, Aspeborg H, Schrader J, Andersson A, Erlandsson R, Blomqvist K, Bhalerao R, Uhlen M, Teeri TT, Lundeberg J, Sunberg B, Nilsson P, Sandberg G** (2001) A transcriptional road map to wood formation. Proc Natl Acad Sci USA **98**: 14732-14737

**Hussey PJ, Haas N, Hunsperger J, Larkin J, Snustad DP, Silflow CD** (1990) The beta-tubulin gene family in *Zea mays*: Two differentially expressed beta-tubulin genes. Plant Mol Biol **15**: 957-72

**Jauh GY, Lord EM** (1996) Localization of pectins and arabinogalactan-proteins in lily (Lilium longiflum L.) pollen tube and style, and their possible roles in pollination. Planta **199**: 251-261

**Jayawickrama KJS, McKeand SE, Jett JB, Wheeler EA** (1997) Date of earlywood-latewood transition in provenances and families of loblolly pine, and its relationship to growth phenology and juvenile wood specific gravity. Can. J. For. Res. **27**:1245-1253

**Jenkins PA** (1974) Influence of applied indoleacetic acid on xylem cell dimension in *Pinus radiata D*. Bull R Soc New Zealand **12**: 737-742

**Johnson KL, Jones BJ, Bacic A, Schultz CJ** (2003) The fasciclin-like arabinogalactan proteins of Arabidopsis. A multigene family of putative cell adhesion molecules. Plant Physiol **133**: 1911-1925

**Jones A, Dangle J** (1996) Logjam at the Styx. Trends in Plant Science **1**: 114-119

**Kamm A, Doudrick RL, Heslop-Harrison JS, Schmidt T** (1996) The genomic and physical organization of Ty-copia-like sequences as a component of large genomes in *Pinus elliotti var. elliotti* and other gymnosperms. Proc Natl Acad Sci USA **93**: 2708-2713

**Kawamoto T, Noshiro M, Shen M, Nakamasu K, Hashimoto K, Kawashima-Ohya Y, Gotoh O, Kato Y** (1998) Structural and phylogenetic analyses of RGD-CAP/βig-h3, a fasciclin-like adhesion protein expressed in chick chondrocytes. Biochim Biophys Acta **1395**: 288-292

**Keller B** (1993) Structural cell wall proteins. Plant Physiol **101**: 1127-1130

**Kerr MK, Churchill GA** (2001a) Statistical design and the analysis of gene expression microarray data. Genetical Research **77**: 123-128

**Kerr MK, Churchill GA** (2001b) Experimental design for gene expression microarrays. Biostatistics **2**: 183-201

**Kieliszewski MJ** (2001) The latest hype on Hyp-O-glycosylation codes. Phytochemistry **57**: 319-323

**Kieliszewski MJ, Lamport DTA** (1994) Extensin: Repetitive motifs, functional sites, post-translational codes, and phylogeny. Plant J **5**: 157-172

**Kinlaw C, Neale D** (1997) Complex gene families in pine genomes. Trends Plant Sci **2**: 356-359

**Kinoshita T, Inoue N** (2000) Dissecting and manipulating the pathway for glycosylphosphatidylinositol-anchor biosynthesis. Curr Opin Chem Biol **4**: 632-638

**Kirst M, Johnson AF, Bausom C, Ulrich E, Hubbard K, Staggs R, Paule C, Retzel E, Whetten R, Sederoff RR** (2003) Apparent homology of expressed genes from wood-forming tissues of loblolly pine (*Pinus taeda L.*) with *Arabidopsis thaliana*. Proc Natl Acad Sci USA **100**: 7383-7388

**Knox JP, Linstead PJ, Peart J, Cooper C, Roberts K** (1991) Developmentally regulated epitopes of cell surface arabinogalactan proteins and their relation to root tissue pattern formation. Plant J **1**: 317-326

**Kohorn BD** (2000) Plasma membrane-cell wall contacts. Plant Physiol **124**: 31-38

**Koivisto UM, Palvimo JJ, Janne OA, Kontula K** (1994) A single-base substitution in the proximal Sp1 site of the human low density lipoprotein receptor promoter as a cause of heterozygous familial hypercholesterolemia. Proc Natl Acad Sci USA **91**: 10526-10530

**Kossack DS, Kinlaw CS** (1999) IFG, a gypsi-like retrotransposon in Pinus (Pinaceae), has an extensive history in pines. Plant Mol Biol **39**: 417-426

**Kozlowski TT, Kramer PJ, Pallardy SG** (1991) The Physiological Ecology of Woody Plants.Academic Press, Inc. San Diego, CA pp.258, 268-269

**Kraus JF, Squillace AE** (1964) Selfing vs. outcrossing under artificial conditions in *Pinus elliotti Engelm*. Silvae Genet **13**: 72-76

**Kreuger M, van Holst GJ** (1993) Arabinogalactan proteins are essential in somatic embryogenesis of *Daucus carota L*. Planta **189**: 243-248

**Kreuger M, van Holst GJ** (1996) Arabinogalactan proteins and plant differentiation. Plant Mol Biol **30**: 1077-1086

**Kriebel HB** (1985) DNA sequence components of Pinus strobes nuclear genome. Can J For Res **15**: 1-4

**Lambeth CC, Dougherty PM, Gladstone WT, McCullough RB, Wells OO** (1984) Large-scale planting of North Carolina loblolly pine in Arkansas and Oklahoma: A case of gain versus risk. J Forestry **82**: 736-741

**Larson PR** (1962) Auxin gradients and the regulation of cambial activity. In Tree Growth. Ed. T.T. Kozlowski. New York: Ronald Press, pp. 97-117

**Larson PR** (1969) Wood formation and the concept of wood quality. Bull. No. 74, Yale University School of Forestry, New Haven, Conn

**Lee MLT, Kuo FC, Whitmore GA, Sklar J** (2000) Importance of replication in microarray gene expression studies: Statistical methods and evidence from repetitive cDNA hybridizations. Proc Natl Acad Sci USA **97**: 9834-9839

**Li B, McKeand SE, Weir RJ** (1999) Tree improvement and sustainable forestry impact of two cycles of loblolly pine breeding in the U.S.A. For Gen **6**: 229-234

**Li X, Gu W, Mohan S, Baylink DJ** (2002) DNA microarrays: Their use and misuse. Microcirculation **9**: 13-22

**Loopstra CA, Puryear JD, No EG** (2000) Purification and cloning of an arabinogalactan-protein from xylem of loblolly pine. Planta **210**: 686-689

**Loopstra CA, Sederoff RR** (1995) Xylem-specific gene expression in loblolly pine. Plant Mol Biol **27**: 277-291

**Lorenz WW, Dean JFD** (2002) SAGE profiling and demonstration of differential gene expression along the axial developmental gradient of lignifying xylem in loblolly pine (*Pinus taeda*). Tree Physiol **22**: 301-310

**Ludwig SR, Oppenheimer DG, Silflow CD, Snustad DP** (1987) Characterization of the alpha-tubulin gene family of *Arabidopsis thaliana*. Proc Natl Acad Sci USA **84**: 5833–5837

**MacKay JJ, O'Malley DM, Presnell T, Booker FL, Campbell MM, Whetten RW, Sederoff RR** (1997) Inheritance, gene expression, and lignin characterization in a mutant pine deficient in cinnamyl alcohol dehydrogenase. Proc Natl Acad Sci USA **95**: 13330–13335

**Majewska-Sawka A, Nothnagel EA** (2000) The multiple roles of arabinogalactan proteins in plant development. Plant Physiol **122**: 3-9

**McAdams HH, Arkins A** (1999) It's a noisy business! Genetic regulation at the nanomolar scale. Trends Genet **15**: 65-69

**McKeever DB, Howard J L** (1996) Value of timber and agricultural products in the United States. Forest Products Journal. **46**: 45-50

**Meagher RB, McKinney EC, Kandasamy MK** (1999) Isovariant dynamics expand and buffer the responses of complex systems: The diverse plant actin gene family. Plant Cell **11**: 995-1005

**Megraw R** (1985) Wood Quality factors in loblolly pine. Tappi Press Atlanta, GA

**Mellorowicz EJ, Baucher M, Sundberg B, Boerjan W** (2001) Unraveling cell wall formation in the woody dicot stem. Plant Mol Biol **47**: 239-274

**Mijnsbrugge KV, Meyermans H, Montagu MV, Bauw G, Boerjan W** (2000). Wood formation in poplar: Identification, characterization, and seasonal variation of xylem proteins. Planta **210**: 589-598

**Moulton RJ, Hernandez G** (2000) Tree Planting in the United States-1998. Tree Planters' Notes **49**: 23-36

**Nakai K, Horton P** (1999) PSORT: A program for detecting sorting signals in proteins and predicting their subcellular localization. Trends Biochem Sci **24**: 34-6

**Nam J, Mysore KS, Zheng C, Knue MK, Mattysse AG, Gelvin SB** (1999) Identification of T-DNA tagged Arabidopsis mutants that are resistant to transformation by *Agrobacterium*. Mol Gen Genet **261**: 429-438

**National Research Council** (1990) *Forestry Research: A Mandate for Change*. National Academy Press, Washington, DC

**Nick P** ed. (2000) Plant Microtubules. Nick P, Springer, Berlin, pp. 1-24

**Nielsen H, Engelbrecht J, Brunak S, von Heijne G** (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Engineering **10**: 1-6

**No EG, Loopstra CA** (2000) Hormonal and developmental regulation of two loblolly pine xylem arabinogalactan-proteins. Physiologia Plantarum **110**: 524 - 529

**Nothnagel EA** (1997) Proteoglycans and related components in plant cells. Int Rev Cytol **174**: 195-291

**Oleksiak MF, Churchill GA, Crawford DL** (2002) Variation in gene expression within and among natural populations. Nature Genet. **32**: 261-266

**Pennell RI** (1992) Cell surface arabinogalactan proteins, arabinogalactans and plant development. *I*n AJ Callow, JR Green, eds, Prospectives in Plant Cell Recognition, Vol 48 Cambridge University Press, Cambridge, UK, pp 105-121

**Pennell RI, Roberts K** (1990) Sexual development in the pea is presaged by altered expression of arabinogalactan protein. Nature **344**: 547-549

**Phillips K** (1998) Drought chops forests by $342 million officials say. AgNews, Texas A&M University Agriculture Program. September 15, 1998. Available at: http://agnews.tamu.edu/stories/DRGHT/Sep1598a.htm

**Pickett FB, Meeks-Wagner DR** (1995) Seeing double: Appreciating genetic redundancy. Plant Cell **7**: 1347-1356

**Roberts DR, Dumbroff EB** (1986) Relationships among drought resistance, transpiration rates, and abscisic acid levels in three northern conifers. Tree Physiol **1**: 161-167

**Roberts K, McCann MC** (2000) Xylogenesis: The birth of a corpse. Current Opinion in Plant Biology **3**: 517-522

**Roy S, Jauh GY, Helper PK, Lord EM** (1998) Effects of Yariv phenylglycoside on cell wall assembly in the lily pollen tube. Planta **204**: 450-458

**Saka S, Goring DAI** (1985) Localization of lignins in wood cell walls. *In* Higuchi T (ed) Biosynthesis and biodegradation of wood components. Academic Press, Orlando, FL, pp 51-62

**Schindler T, Bergfeld R, Schopfer P** (1995) Arabinogalactan proteins in maize coleoptiles: Developmental relationship to cell death during xylem differentiation but not to extension growth. Plant J **7**: 25-36

**Schultz CJ, Gilson P, Oxley D, Youl J, Bacic A** (1998) GPI-anchors on arabinogalactan-proteins: Implications for signaling in plants. Trends Plant Sci **3**: 426-431

**Schultz CJ, Rumsewicz MP, Johnson KL, Jones BJ, Gaspar YM, Bacic A** (2002) Using genomic resources to guide research direction. The arabinogalactan protein gene family as a test case. Plant Physiol **129**: 1448-1463

**Sederoff RR, Campbell M, O'Malley D, Whetten R** (1994) Genetic regulation of lignin biosynthesis and the potential modification of wood by genetic engineering in loblolly pine. Recent Advances in Phytochemistry **28**: 313-355

**Segal JA, Barnett JL, Crawford DL** (1999) Functional analyses of natural variation in Sp1 binding sites of a TATA-less promoter. J Mol Evol **49**: 736-749

**Serpe MD, Nothnagel EA** (1994) Effects of Yariv phenylglycosides on Rosa cell suspensions: Evidence for the involvement of arabinogalactan-proteins in cell proliferation. Planta **193**: 541-550

**Sewell MM, Davis MF, Tuskan GA, Wheeler NC, Elam CC, Bassoni DL, Neale DB** (2002) Identification of QTLs influencing wood property traits in loblolly pine (*Pinus taeda* L.) II. Chemical wood properties. Theor Appl Genet **104**: 214-222

**Sewell MM, Neale DB** (2000) Mapping quantitative traits in forest trees. Jain, S. M. and Minocha, S. C. eds, *In* Molecular Biology of Woody Plants, Forestry Sciences, Volume 64, Kluwer Academic Publishers, Dordrecht, The Netherlands, pp 407-423

**Shi H, Kim Y, Guo Y, Stevenson B, Zhu JK** (2003) The Arabidopsis SOS5 locus encodes a putative cell surface adhesion protein and is required for normal cell expansion. Plant Cell **15**: 19-32

**Showalter AM** (1993) Structure and function of plant cell wall proteins. Plant Cell **5**: 9-23

**Showalter AM, Varner JE** (1987) Plant hydroxyproline-rich glycoproteins. *In* Marcus A (ed) The Biochemistry of Plants, vol 15, pp. 385-519. Academic Press, New York

**Shpak E, Barber E, Leykam JF, Kieliszewski MJ** (2001) Contiguous hydroxyproline residues direct hydroxyproline arabinosylation in *Nicotiana tabacum*. J Biol Chem **276**: 11272-11278

**Shpak E, Leykam JF, Kieliszewski MJ** (1999) Synthetic genes for glycoprotein design and the elucidation of hydroxyproline-O-glycosylation codes. Proc Natl Acad Sci USA **96**: 14736-14741

**Slunder ER** (1980) A study of geographic variation in loblolly pine in Georgia – 20$^{th}$ year results. USDA Forest Serv Res Paper SE-213 Southeastern Forest Exp Stn Asheville NC 26p

**Smith RC, Fry SC** (1991) Endotransglycosylation of xyloglucans in plant cell suspension cultures. Biochem J **279**: 529-535

**Smook GA, Kocurek MJ** (1988) Handbook for pulp and paper technology. TAPPI Press, Atlanta, GA

**Snustad DP, Haas NA, Kopczak SD, Silflow CD** (1992) The small genome of *arabidopsis* contains at least nine expressed β-tubulin genes. Plant Cell **4**: 549-556

**Somerville C, Somerville S** (1999) Plant functional genomics. Science **285**: 380-383

**Sommer-Knudsen J, Bacic A, Clarke AE** (1998) Hydroxyproline-rich plant glycoproteins. Phytochemistry **47**: 483-497

**Stasolla C, Egerstdotter U, Scott J, Kadla J, Sederoff RR, van Zyl L** (2003) Lignin composition in wild type and cinnamyl alcohol dehydrogenase deficient mutant cultured cells of *Pinus taeda*. Plant Physiology and Biochemsitry **41**: 439-445

**Stephens JC, Schneider JA, Tanguay DA, Choi J, Acharya T, Stanley SE, Jiang R, Messer CJ, Chew A, Han JH, Duan J, Carr JL, Lee MS, Koshy B, Kumar AM, Zhang G, Newell WR, Windemuth A, Xu C, Kalbfleisch TS, Shaner SL, Arnold K, Schulz V, Drysdale CM, Nandabalan K, Judson RS, Ruano G, Vovis GF** (2001) Haplotype variation and linkage disequilibrium in 313 human genes. Science **293**: 489-93

**Sterky F, Regan S, Karlsson J, Hertzberg M, Rohde A, Holmberg A, Amini B, Bhalerao R, Larsson M, Villaroel R, Montagu MV, Sanberg G, Olsson O, Teeri TT, Boerjan W, Gustafsson P, Uhlen M, Sundberg B, Lundeberg J** (1998) Gene discovery in the wood-forming tissues of poplar: analysis of 5692 expressed sequence tags. Proc Natl Acad Sci USA **95**: 13330–13335

**Takahashi T, Gasch A, Nishizawa N, Chua NH** (1995) The *DIMMUTO* gene of Arabidopsis is involved in regulating cell elongation. Genes Dev **9**: 97-107

**Thompson HJM, Knox JP** (1998) Stage-specific responses of embryogenic carrot cell suspension cultures to arabinogalactan protein-binding β-glucosyl Yariv reagent. Planta **205**: 32-38

**Uggla C, Magel E, Moritz T, Sundberg B** (2001) Function and dynamics of auxin and carbohydrates during earlywood/latewood transition in scots pine. Plant Physiol **125**: 2029–2039

**van Buijtenen JP** (1964) Anatomical factors influencing wood specific gravity of slash pines and the implications for the development of high-quality pulpwood. TAPPI **47**: 401-404

**van Zyl L, Bozhkov P, Clapham D, Sederoff R, von Arnold S** (2002) Up, down and up again is a signature global gene expression pattern at the beginning of gymnosperm embryogenesis. Gene Expression Patterns **3**: 83-91

**Wakamiya I, Price HJ, Messina MG, Newton RJ** (1996) Pine genome size diversity and water relations. Physiologia Plantarum **96**: 13-20

**Watts WA** (1983) Vegetational history of the eastern United States 25,000 to 10,000 years ago. *In* S.C. Porter ed. Late-Quaternary environments of the United States. Volume 1. The late Pleistocene. Univ. Minnesota Press, Minneapolis, pp. 294-310

**Wells OO** (1985) Use of Livingston Parish, Louisiana loblolly pine by forest products industries in the southeast. South J Appl For **9**: 180-185

**Wells OO, Switzer GL, Schmidtling RC** (1991) Geographic variation in Mississippi loblolly pine and sweetgum. Silvae Genet **40**: 105-118

**Wells OO, Wakeley PC** (1966) Geographic variation in survival, growth and fusiform rust infection of planted loblolly pine. For Sci Monog **11**: 40

**Westfall PH, Young SS** (1993) Resampling-based multiple testing: Examples and methods for p-value adjustment. Wiley, New York

**Whetten R, Sun YH, Zhang Y, Sederoff RR** (2001) Functional genomics and cell wall biosynthesis in loblolly pine. Plant Mol Biol **47**: 275- 291

**Whitmore FW Zahner R** (1966) Development of the xylem ring in stems of young red pine trees. Forestry Sci **12**:198-210

**Willats WGT, Knox JP** (1996) A role for arabinogalactan-proteins in plant cell expansion: Evidence from studies on the interaction of β-glucosyl Yariv reagent with seedlings of *Arabidopsis thaliana* Plant J **9**: 919-925

**Wu HM, Wong E, Ogdahl J, Cheung AY** (2000) A pollen tube growth-promoting arabinogalactan protein from *Nicotiana alata* is similar to the tobacco TTS protein. Plant J **22**: 165-176

**Yang SH, van Zyl L, No EG, Loopstra CA** (2004) Microarray analysis of genes preferentially expressed in differentiating xylem of loblolly pine (*Pinus taeda*). Plant Sci **166**: 1185-1195

**Yang YH, Dudoit S, Luu P, Speed TP** January 2001. Normalization for cDNA microarray data SPIE. BiOS. San Jose, CA

**Yariv J, Lis H, Katchalski E** (1967) Precipitation of arabic acid and some seed polysaccharides by glycosylphenylazo dyes. Biochem J **105**: 1c-2c

**Youl JJ, Bacic A, Oxley D** (1998) Arabinogalactan-proteins from *Nicotiana alata* and *Pyrus communis* contain glycosylphosphatidylinositol membrane anchors. Proc Natl Sci **95**: 7921-7926

**Yue H, Eastman PS, Wang BB, Minor J, Doctolero MH, Nuttall RL, Stack R, Becker JW, Montgomery JR, Vainer M, Johnston R** (2001) An evaluation of the performance of cDNA microarrays for detecting changes in global mRNA expression. Nucl Acids Res **29**: e41

**Zhang Y, Sederoff RR, Allona I** (2000) Differential expression of genes encoding cell wall proteins in vascular tissues from vertical and bent loblolly pine trees. Tree Physiol **20:** 457-466

**Zhu JK, Bressan RA, Hasegawa PM** (1993) Loss of arabinogalactan-proteins from the plasma membrane of NaCl-adapted tobacco cells. Planta **190**: 221-226

**Zimmermann MH, Brown CL** (1971) Trees: Structure and Function. Springer-Verlag, New York

**Zobel BJ, Jett JB** (1995) Genetics of wood production. Springer-Verlag, New York

# VITA

**SUK HWAN YANG**
3625 Wellborn Rd. #1815
Bryan, TX 77801
Home: (979) 260-1608
Work: (979) 862-4803
Fax: (979) 862-4790
yangsh@neo.tamu.edu

## Education

| | |
|---|---|
| August 2004 | Ph.D. in Molecular and Environmental Plant Sciences<br>Texas A&M University, College Station, TX<br>Dissertation: Transcript Profiling of Differentiating Xylem of Loblolly Pine *(Pinus taeda L.*)<br>Advisor: Dr. Carol A. Loopstra |
| December 1998 | M.S. in Genetics<br>Iowa State University, Ames, IA<br>Thesis: Expression of a Synthetic Porcine α-Lactalbumin Gene in the Kernels of Transgenic Maize<br>Advisor: Dr. Paul Scott |
| August 1995 | B.S. in Horticulture<br>Korea University, Seoul, South Korea |

## Permanent Address

407 Seogui Dong
Seoguipo-City, Jeju-Do, 697-010
South Korea