Transcriptome Profiling To Identify Genetic Determinants Of The Juvenile To Mature Wood Transition

W. Walter Lorenz and Jeffrey F.D. Dean

Daniel B. Warnell School of Forest Resources and the Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA, 30602-2152, USA

jeffdean@uga.edu

ABSTRACT

Wood formation is a process that has been extensively studied at the cellular and biochemical levels, but which remains as yet poorly understood with respect to gene expression and regulation. As a first step towards understanding how changes in the gene expression patterns in wood-forming tissues lead to variations in wood quality, serial analysis of gene expression (SAGE) was used to quantitate gene expression in the lignifying xylem from a single juvenile loblolly pine. Two SAGE libraries were generated using lignifying xylem isolated from either the upper (crown) or lower (base) portions of the trunk. The crown juvenile wood library contained ~ 85,000 tags representing a maximum of 27,398 expressed genes, while the base juvenile wood library contained ~65,000 tags representing a maximum of 25,983 expressed genes. Combining these two data sets to reflect the sum of genes expressed in lignifying xylem, 150,855 tags were catalogued representing a maximum of 42,000 different genes. These results are all the more remarkable given recent reports suggesting that the Drosophila genome only harbors about 13,000 genes while the human and Arabidopsis genomes may only contain 25,000-30,000 genes. Thus, SAGE analysis suggests either that loblolly pine expresses many more genes than would be expected from estimates made in other organisms, or that post-transcriptional processing (e.g. alternative splicing) is very common in pine. To date, this study represents the most extensive analysis of its kind in a higher plant and provides a quantitative description of the transcriptome representing the lignifying xylem of a juvenile loblolly pine.

INTRODUCTION

In addition to the radial developmental gradients that give rise to growth rings in trees, the vascular cambium has an axial developmental gradient (Zobel and Sprague 1998). When the cambium is young and in close proximity to photosynthetic tissues, xylem cells tend to be relatively short with thin walls and their long axes are oriented away from the axial axis of the stem. (The latter characteristic leads to a phenomenon known as spiral grain.) Additional characteristics of xylem cells formed in cambium near the tree crown include cellulose microfibrils that tend to be oriented less toward the cell's longitudinal axis, as well as cell wall composition that tends toward lower cellulose and higher lignin content. The wood comprising these cells is generally of lower density and is referred to as being *juvenile*, in contrast to the denser, *mature* wood formed by older cambial tissues distant from the photosynthetic machinery. Many of the commercial products manufactured from wood were developed and optimized on the basis of using mature wood as a feedstock, but shifting land-use patterns and changing silvicultural practices are leading the forest products industry to use younger trees that have a

higher content of juvenile wood. Consequently, there is substantial interest in understanding, and manipulating, the transition from juvenile to mature wood formation.

A variety of mechanisms have been suggested for the induction of juvenile wood formation (Gartner 1996, Lindstrom 1996). Most of these are based on acropetal auxin gradients in which cambial tissues close to the photosynthetic tissues of the tree crown contain higher levels of auxin than do cambial tissues well below the crown. Such a role for auxin would be in keeping with results suggesting that auxin gradients also control the radial growth and development of cambial tissues (Tuominen et al. 1997, Uggla et al. 1998). Regardless of what mechanism regulates formation of the juvenile wood developmental gradient, the cellular phenotype is a reflection of the genes expressed under the influence of this gradient. Thus, the genes critical for the juvenile and mature xylem cell phenotypes should be differentially expressed along the axial gradient of the tree. Identification of these genes should enhance our understanding of wood formation and suggest ways in which their expression might be manipulated to provide superior wood for commercial products.

For our effort to profile secondary xylem gene expression and to identify genes that are differentially expressed along the axial developmental gradient in loblolly pine xylem, we employed "serial analysis of gene expression" (SAGE), a technique that so far has not been much exploited outside the biomedical community (Velculescu, et al. 1995, Velculescu et al. 1997). SAGE is essentially a method for rapidly generating and quantitating "microESTs" representing the entire transcriptome of sampled tissues. A short nucleotide sequence (9-10 bp) contains sufficient information to uniquely identify transcripts from individual genes, provided the sequence (tag) is derived from a defined position along the transcript (a general outline of the SAGE technique can be found at http://www.sagenet.org/). These tags are subsequently strung together, like beads on a necklace, at which point their short length allows them to be rapidly and economically identified and tallied using automated DNA sequencers and specialized software. Identification of differentially expressed genes then comes about through *in silico* analysis of the tags isolated from different tissue samples.

RESULTS

SAGE Analysis

For this study, the trunk of a 10-year old loblolly pine was arbitrarily bisected at the point of attachment for the lowest living branch, and the non-lignified xylem harvested from the bole above this point was designated crown juvenile wood, while xylem from below this point was designated base juvenile wood. SAGE libraries, hereafter referred to as the crown and base libraries, were prepared using RNA isolated from each of these samples. The crown library was sequenced to an extent of ca. 85,000 tags, while the base library was sequenced to ca. 65,000 tags. The total number of SAGE tags sequenced from both libraries was 150,855 and the maximum total number of unique tags (potential genes) identified was 42,641. The datasets from these SAGE library analyses are available in spreadsheet format (MS Excel) at http://www.arches.uga.edu/~jeffdean/SAGE/SAGEData.html.

Table 1 summarizes the frequency of appearance (relative abundance) of groups of unique genes within each library. The SAGE data were similar between the two libraries with respect to the number of genes identified, the number of tags falling into each frequency class, and the relative number of tags each class represented as a percentage of the total tags sequenced. In either library, the majority of unique tags were detected only once and represented

Table 1. Summary of SAGE Data

	Frequency Distribution*				
Library	≥20	19 to 5	4 to 2	= 1	**Totals**
Crown wood					
Unique tags	592 (2.2)	2,786 (10.1)	6,405 (23.4)	17,615 (64.3)	27,398
Tags sequenced	27,260 (32.0)	23,814 (28.0)	16,513 (19.4)	17,615 (20.6)	85,202
Bole wood					
Unique tags	388 (1.5)	2,034 (7.8)	5,810 (22.4)	17,752 (68.3)	25,984
Tags Sequenced	16,243 (24.7)	16,930 (25.8)	14,758 (22.5)	17,752 (27.0)	65,683
Total Xylem					
Unique tags	1,150 (2.7)	4,487 (10.5)	9,726 (22.8)	27,279 (64.0)	42,641
Tags Sequenced	59,745 (39.6)	38,794 (25.7)	25,067 (16.6)	27,279 (18.1)	150,855

*Frequency distributions were calculated based on the total number of unique or sequenced tags in each library shown in the totals column. The percentages of tags in each frequency group are shown parenthetically.

64% and 68% of the tags identified in the crown and base samples, respectively. However, these represented only 20.6% and 27% of the total tags sequenced in the respective libraries. The combined crown and base data sets yielded similar values with 64% of the unique tags being detected only once, while representing 18% of the total tags counted. The high number of low abundance tags seen in these libraries was consistent with similar data from other studies. For instance, single tag entries accounted for 69% of the tags identified in a human skeletal muscle SAGE library (Welle et al. 1999). Similarly, they accounted for 65% and 69% of the tags, respectively, in SAGE libraries generated from human resting versus activated mast cells (Chen et al. 1998). These values are also in accordance with observations from various EST projects (see for example Adjaye et al. 1997, Lee et al. 1999).

Looking at the distribution of tags over all the abundance classes, the values remained relatively consistent between the crown, base and combined libraries for moderately expressed tags, i.e. those counted 2-4 times and those counted 5-19 times. In contrast, the tags observed most frequently (≥20), represented 1.5% and 2.2% of the genes found in crown and base libraries, respectively, and they accounted for 32% and 24.7% of the total tags sequenced. For the combined data set, the percentage of abundant genes increased to 2.7% of genes identified, representing 39.6% of the total tags sequenced. Looking at tags only seen once, 4% more unique genes were identified in the base library than in the crown library.

A BLAST search of the Pine Gene Discovery Database (http://www.cbc.umn.edu/ResearchProjects/Pine/DOE.pine/index.html) was performed using the entire 14 bp tag sequence (10 bp tag + 4 bp anchoring enzyme recognition site) for the 500 most abundant SAGE tags. Only when there was an exact match of all 14 bp in the correct orientation was the tag annotated with the EST information. In instances where the pine EST sequence was not annotated, a BLAST search was performed against Genbank using the EST sequence and the most probable match was added to the SAGE tag annotation. Table 2 shows EST matches made the 10 most

Table 2. The Ten Most Abundant Transcripts in Loblolly Pine Xylem as Identified by SAGE

Rank	Abundance (%)	Tag Sequence	Accession #	Identification
1	0.32	GGTTCAAGAC	NM*	NM
2	0.29	TATGCCAAGC	AW056690	α-Tubulin
3	0.29	CATTCATTTT	U09554	AGP
4	0.28	AGCAATGGGG	AA556993	AGP
5	0.28	ACATTTTTCT	AA739536 AA739984	NM
6	0.26	CAAATCTTTG	AA557054	Ubiquitin
7	0.25	GCGGCCCTGG	U09556	AGP
8	0.20	GCATTAAAGG	NM	NM
9	0.19	TGTATTGTTG	AA556884 AA556664	Aquaporin
10	0.18	GGATACTCTG	AW042892	Cyclophilin

*NM = no match

abundant tags in the combined pine xylem data set. EST or cDNA matches were made for 160 of the 200 most abundant tags (data not shown). Surprisingly, a match for the most highly expressed tag (0.32% of total tags counted) was not found among the pine EST sequences; the closest *Pinus* sequence in the database only matched 11 of the 14 nucleotides in the tag. Not surprisingly, four of the seven most highly expressed tags corresponded to structural proteins, such as α-tubulin and arabinogalactan-rich proteins (AGPs). Several housekeeping genes, included ubiquitin, aquaporin, cyclophilin and thioredoxin, were also represented amongst the most highly expressed tags. Other highly expressed tags were associated with genes whose products are involved in C1 metabolism (methionine synthetase and S-adenosylmethionine [SAM] synthetase), as well as metal metabolism (metallothionein-like protein).

Differential Gene Expression

The primary focus of this study was to generate an expression profile for lignifying xylem isolated from a 10 year old loblolly pine. However, a secondary focus was to determine what, if any, differences in gene expression patterns existed between the crown and base regions of the tree. As the crown and base SAGE libraries were sequenced to slightly different degrees, the tag values for the base library were normalized to allow direct comparison with values for the larger crown library. As shown in Figure 1, comparison of the differential expression patterns of the two libraries revealed that, as expected, the majority of tags were expressed at similar levels.

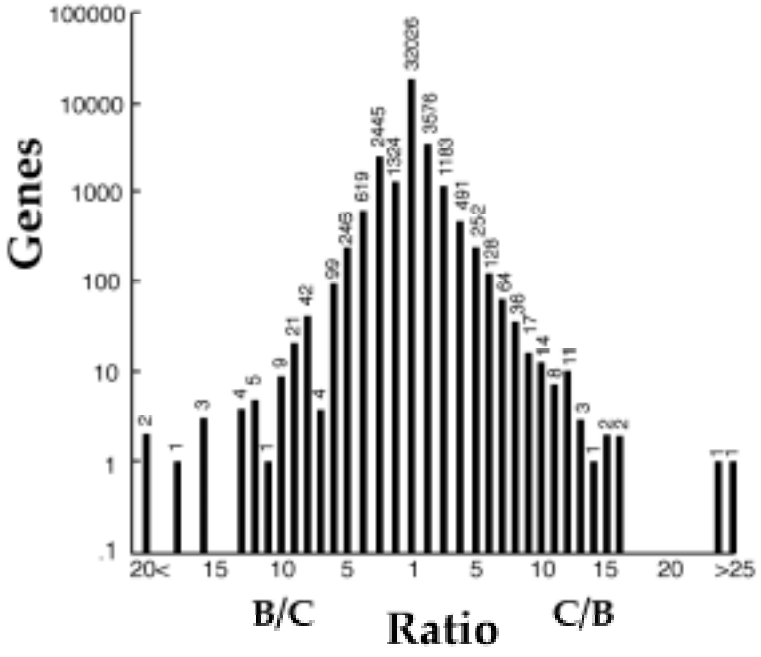

Figure 1. Differential gene expression patterns along the axial gradient of lignifying xylem are displayed as the ratios between the crown and base library counts for each tag. The number of genes displaying each ratio is plotted against a log scale on the ordinate with the exact number displayed above each bar.

With respect to genes having the greatest expression differences in the lignifying xylem from the base of the tree, two tags had a 21-fold greater abundance in the base library versus the crown library. Twenty-three other tags were expressed at least 10 times more frequently in the base library than in the crown library. In the converse situation, i.e. genes more highly expressed in the crown wood, one tag was 61-fold more abundant and another 24-fold, while another forty tags having at least 10-fold greater expression in the crown wood were identified. In all cases where the difference in expression was at least 10-fold, $P \leq 0.01$.

DISCUSSION

There is an axial developmental gradient in the xylem of trees which results in the production of wood having different quality in the crown versus the base of the tree. To identify those genes that are differentially expressed along this developmental gradient in loblolly pine, as well as to catalog total gene expression in the non-lignified xylem of an individual tree, SAGE libraries were generated using tissues isolated from either the crown or base regions of the trunk. This is only the second reported instance in which SAGE has been used to study plant gene expression, and its quantitation of 150,855 tags leading to the identification of 42,641 expressed genes (maximum) put it on a par with some of the most detailed SAGE studies performed using human tissues (for example, St. Croix et al. 2000). Comparison of the two libraries showed that while the vast majority of genes did not show significant differences in expression levels between the two tissues, more than 60 genes appeared to have greater than 10-fold differences in their levels of expression. It is hoped that many of the tags identified as being differentially expressed in crown and base wood xylem will represent genes that are important determinants of wood quality during juvenile and mature wood formation. The power of SAGE as a tool to simultaneously identify and evaluate quantitatively the expression patterns of thousands of genes in pine, the majority of which have not yet been characterized, is juxtaposed to the limitations on the number of libraries made and the present difficultly in assigning tags to their respective genes, given the limited number of cDNA and EST sequences currently in the pine database. The latter will no doubt be remedied as more sequence data is generated and compiled.

Acknowledgements

The authors wish to thank Drs. Kenneth Kinzler, Bert Vogelstein and the members of their groups at the John's Hopkins Oncology Center.

REFERENCES

Adjaye, J., Daniels, R., Bolton, V., Monk, M. 1997. cDNA libraries from single human preimplantation embryos. Genomics 46: 337-344.

Chen, B.H., Centola, M., Altschul, S.F., Metzger, H. 1998. Characterization of gene expression in resting and activated mast cells. J. Exp. Med. 188: 1657-1668.

Gartner, B.L. 1996. Does photosynthetic bark have a role in the production of core vs outer wood? Wood Fib. Sci. 28: 53-61.

Lee, Y.H., Huang, G.M., Cameron, R.A., Graham, G., Davidson, E.H., Hood, L., Britten, R.J. 1999. EST analysis of gene expression in early cleavage-stage sea urchin embryos. Development 126: 3857-3867.

Lindstrom, H. 1996. Basic density in Norway spruce .1. A literature review. Wood Fib. Sci. 28: 15-27.

St. Croix, B., Rago, C., Velculescu, V., Traverso, G., Romans, K.E., Montgomery, E., Lal, A., Riggins, G.J., Lengauer, C., Vogelstein, B., Kinzler, K.W. 2000. Genes expressed in human tumor endothelium. Science 289: 1197-202.

Tuominen, H., Puech, L., Fink, S., Sundberg, B. 1997. A radial concentration gradient of indole-3-acetic acid is related to secondary xylem development in hybrid aspen. Plant Physiol. 115: 577-585.

Uggla, C., Mellerowicz, E.J., Sundberg, B. 1998. Indole-3-acetic acid controls cambial growth in Scots pine by positional signaling. Plant Physiol. 117: 113-121.

Velculescu, V.E., Zhang, L., Vogelstein, B., Kinzler, K.W. 1995. Serial analysis of gene expression. Science 270: 484-487.

Velculescu, V.E., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M.A., Bassett Jr., D.E., Hieter, P., Vogelstein, B., Kinzler, K.W. 1997b. Characterization of the yeast transcriptome. Cell 88: 243-251.

Welle, S., Bhatt, K., Thornton, C.A. 1999. Inventory of high abundance mRNA's in skeletal muscle of normal men. Genome Res. 9: 506-513.

Zobel, B.J., Sprague, J.R. 1998. Juvenile Wood in Forest Trees. Springer-Verlag, Berlin.