**SAGE transcript profiling of the juvenile cambial region of *Eucalyptus grandis***

MAYRA COSTA DA CRUZ GALLO DE CARVALHO,1,2 DANIELLE GREGORIO GOMES CALDAS,1,2 RAPHAEL TOZELLI CARNEIRO,1,2 DAVID HENRY MOON,1,2 GUILLERMO RAFAEL SALVATIERRA,1 LÍVIA MARIA FRANCESCHINI,1 ALEXANDER DE ANDRADE,1 PAOLA ALEJANDRA FIORANI CELEDON,1 SHINITIRO ODA3 and CARLOS ALBERTO LABATE1,4

1 Laboratório Max Feffer de Genética de Plantas, Departamento de Genética, Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba-SP, Brazil
2 These authors contributed equally to this work as first authors
3 Suzano Papel e Celulose, Av. Brigadeiro Faria Lima No. 1355, 8º andar, CEP 01452-919, São Paulo-SP, Brazil
4 Corresponding author (calabate@esalq.usp.br)

Received August 20, 2007; accepted November 28, 2007; published online April 1, 2008

**Summary** Despite the importance of *Eucalyptus* spp. in the pulp and paper industry, functional genomic approaches have only recently been applied to understand wood formation in this genus. We attempted to establish a global view of gene expression in the juvenile cambial region of *Eucalyptus grandis* Hill ex Maiden. The expression profile was obtained from serial analysis of gene expression (SAGE) library data produced from 3- and 6-year-old trees. Fourteen-base expressed sequence tags (ESTs) were searched against public *Eucalyptus* ESTs and annotated with GenBank. Altogether 43,304 tags were generated producing 3066 unigenes with three or more copies each, 445 with a putative identity, 215 with unknown function and 2406 without an EST match. The expression profile of the juvenile cambial region revealed the presence of highly frequent transcripts related to general metabolism and energy metabolism, cellular processes, transport, structural components and information pathways. We made a quantitative analysis of a large number of genes involved in the biosynthesis of cellulose, pectin, hemicellulose and lignin. Our findings provide insight into the expression of functionally related genes involved in juvenile wood formation in young fast-growing *E. grandis* trees.

**Keywords:** cell wall, cellulose, gene expression, lignin, nucleotide-sugars interconversions, transcriptome, wood formation.

**Introduction**

Among plantation tree species, *Eucalyptus* spp. have the highest growth rates. In Brazil, *Eucalyptus* wood is widely used as raw material for the pulp and paper industry. Despite the high productivity (45–60 m³ ha⁻¹ year⁻¹) of *Eucalyptus* plantations, increasing demand for cellulose has resulted in wood shortages in recent years. Hence, many efforts are being made to improve forest productivity.

Wood is formed by the differentiation of cells produced by the vascular cambium. During differentiation, the xylem mother cells undergo an ordered series of developmental steps that include cell division, cell expansion, deposition of the secondary cell wall, lignification and programmed cell death (Larson 1994). The genetic factors controlling wood formation in *Eucalyptus* are not fully understood, although many genes involved in wood formation have been identified by large-scale genomic approaches in *Populus*, *Pinus* and *Eucalyptus* (Hertzberg et al. 2001, Lorenz and Dean 2002, Paux et al. 2004). Hertzberg et al. (2001) established a hierarchical pattern of gene expression through different zones of developing xylem in *Populus* by isolating cells at different stages of xylogenesis. Based on microarray analysis, these authors showed that genes encoding enzymes involved in cellulose and lignin biosynthesis, as well as a large number of transcription factors and potential xylogenesis regulators, are under strict control at each xylem differentiation stage. Paux et al. (2004) developed a targeted approach to functional genomics by constructing a xylem versus leaves subtractive library to identify genes involved in the control of *Eucalyptus* wood formation. The two main classes of expressed sequence tags (ESTs) preferentially expressed in xylem were related to auxin signaling through ubiquitin proteolysis and cell wall biosynthesis and remodeling. More recently, the induction of tension wood has been used as a model of wood formation because of the high cellulose content and limited lignification of tension wood (Paux et al. 2005, Andersson-Gunneras et al. 2006).

To investigate gene expression during juvenile wood formation, when tree growth is maximal, we produced two SAGE li-
Plant material

Tissue samples were collected from a half-sibling population of E. grandis derived from a single mother tree, second generation from a clonal seed orchard, introduced from Coff’s Harbour, Australia, and kindly provided by Suzano Papel e Celulose. Forty 3-year-old and forty 6-year-old trees were sampled from stands located in Itapetininga, State of São Paulo, Brazil (23°35′20″ S, 48°03′11″ W) at an altitude of 656 m. The 3- and 6-year-old trees were spaced at 3 × 1.5 m and had a mean height of 18 and 25 m, and a diameter at breast height (DBH) of 10 and 15 cm, respectively. On the same morning in the summer of 2003, the cambial region from each tree was collected by opening a window (20 × 15 cm) in the bark at breast height, scraping the exposed stem tissue and immediately freezing the sample in liquid nitrogen (Figures 1a and 1b). The stem tissue was scraped until the fibrous material below the differentiating cells was reached. We also scraped the inner side of the bark to ensure that all the meristematic material would be represented in the SAGE libraries. This procedure was first used by Foucart et al. (2006) who showed that cambial cells adhere to the bark during removal. In the two transverse sections of Eucalyptus wood (xylem, cambial region and phloem) shown in Figures 1c and 1d, it can be seen that the meristematic cells tend to stay with the phloem when the bark is removed. Two bulked samples, one representing the 3-year-old trees and the other the 6-year-old trees, were made by grinding and mixing all the sampled material.

Young fully expanded leaves from 3-year-old E. grandis trees were collected and immediately frozen in liquid nitrogen and stored at –80 °C until processed for RNA extraction. E. grandis seeds were surface sterilized with hypochlorite, germinated on agar medium (Murashige and Skoog 1962, 0.6% agar) and the hypocotyls harvested after 30 days, frozen and stored at –80 °C. Transverse sections of the hypocotyl showed that the vascular bundles had not fused to form a continuous cambium (Figure 1e).

Extraction of RNA

Total RNA was extracted from 2 g of each bulked sample by the phenol-based protocol described by Salzman et al. (1999) and quantified spectrophotometrically at 260 nm. Because the samples oxidized rapidly and contained polyphenol compounds, we doubled the extraction buffer volume and increased the PVP concentration to 2% (w/v). Before starting the construction of each serial analysis of gene expression (SAGE) library, the RNA quality was checked spectrophotometrically and by RT-PCR analysis using specific primers for the Eucalyptus gene encoding isocitrate dehydrogenase. The same protocol was used to extract total RNA from the leaves and hypocotyls.

SAGE library construction and sequencing

Total RNA (75 µg) was used to construct each library using the 1-SAGE Kit (Invitrogen). NlaIII was used as the anchoring enzyme and BsmFI as the tagging enzyme (Invitrogen T5000-01) to produce 10- to 14-base SAGE tags. The SAGE technique involved: (1) ligation of the polyadenylated mRNA to oligo dT magnetic beads, (2) cDNA synthesis, (3) digestion with NlaIII which cuts at the CATG recognition site, (4) ligation of adapters containing the recognition site for BsmFI and annealing sites for specific primers, (5) digestion with BsmFI cleaving 10 bases downstream of the recognition site, (6) ligation of digested fragments leaving adapters at the extremities, (7) amplification and digestion of the products with NlaIII releasing the ditags, (8) concatamerization of the ditags and purification of the concatamers (400–800 bp), (9) cloning into pZeroO and transformation of Escherichia coli by electroporation (Micro PulserTM da Bio-Rad). Plasmid DNA extraction from the SAGE clones was carried out with the PureLink Plasmid Puriﬁcation System (Invitrogen) and inserts were sequenced with the BigDye Terminator v.3.1 system and the ABI PRISM 3100 Genetic Analyzer (Applied Biosystems).

Figure 1. Sampling of the Eucalyptus grandis stem. (a) Removal of the bark. (b) Harvesting the cambial region by scraping the tissue. (c) Transverse section of E. grandis wood before removal of the bark. (d) Transverse section of E. grandis wood after removal of the bark. (e) Transverse section of the hypocotyl showing the four vascular bundles (arrow). Bars = 100 µm. Abbreviations: X = xylem; and C = cambial region.
Bioinformatics

Because interpretation of the SAGE data is primarily dependent on the correct tag-gene association, we adopted five criteria for data analysis: (1) all reads from the sequencing were analyzed with phred (www.phrap.org) and only those with $q = 20$ were selected for tag extraction using SAGE 2000 v4.5 (Invitrogen) software; (2) tags with three or more copies were selected for identification; (3) the CATG site and the tag sequence were used to search a locally constructed database containing publicly available Eucalyptus ESTs and complete cDNAs using BLASTn (www.ncbi.nlm.nih.gov); (4) only sequences where the 10-base tag was immediately after the last NlaIII recognition site (CATG) before the poly-A tail were used for annotation; (5) the ESTs representing the tags were used in BLASTx to find the most similar protein in the Genbank database ($E$ value = e-05). BLASTn was carried out locally using a database constructed specifically for SAGE tag annotation containing 15,351 sequences. This local database contained complete cDNA sequences from some important genes and ESTs from four libraries: cDNA library of differentiating xylem from Eucalyptus gunnii Hook. (gi 10348194–103473644 and 51555734–515555511); gravity versus microgravity subtracted library from E. globulus Labill. (gi 103484112–103482902 and 103473643–103472139); E. globulus ESTs from juvenile versus mature wood and mature wood versus juvenile wood subtractive libraries (gi 103482899–103481625); and E. gunnii xylem versus phloem subtractive library (gi 84490924–84490662) (see Supplementary Tables S1 and S2). An explanation of how the tags could be associated with eucalypt sequences using the remote BLASTn service at NCBI and the NCBI gi numbers of the Eucalyptus sequences used to construct the local database can be obtained from the corresponding author.

Because some important enzymes had no associated tag with three or more copies, we carried out a reverse search using a public protein sequence and tBLASTn to identify an EST. Where the 10-base tag was immediately after the last NlaIII recognition site (CATG) before the poly-A tail were used for annotation in the Genbank database ($E$ value = e-05). BLASTn was carried out locally using a database constructed specifically for SAGE tag annotation containing 15,351 sequences. This local database contained complete cDNA sequences from some important genes and ESTs from four libraries: cDNA library of differentiating xylem from Eucalyptus gunnii Hook. (gi 10348194–103473644 and 51555734–515555511); gravity versus microgravity subtracted library from E. globulus Labill. (gi 103484112–103482902 and 103473643–103472139); E. globulus ESTs from juvenile versus mature wood and mature wood versus juvenile wood subtractive libraries (gi 103482899–103481625); and E. gunnii xylem versus phloem subtractive library (gi 84490924–84490662) (see Supplementary Tables S1 and S2). An explanation of how the tags could be associated with eucalypt sequences using the remote BLASTn service at NCBI and the NCBI gi numbers of the Eucalyptus sequences used to construct the local database can be obtained from the corresponding author.

Because some important enzymes had no associated tag with three or more copies, we carried out a reverse search using a public protein sequence and tBLASTn to identify an EST in our local database, extract the tag and confirm its presence in the SAGE database. When a cell-wall-related gene (Figures 2a and 2b) was represented by more than one tag, the respective ESTs were aligned using tbl2seq (www.ncbi.nlm.nih.gov) to look for potential alternative transcripts.

Functional analysis of the annotated genes expressed in the cambial region was carried out based on the categories described by Rison et al. (2000) with a few modifications. The subcategories vesicle transport and transport factors were included in the transport category, and the subcategory cytoskeleton was included in the category structure and organization of cellular structure. Genes associated with the biosynthesis of cell wall components were also grouped into specific metabolic pathways.

Validation by qRT-PCR

To validate the correlation between tag number and gene expression, mRNA was purified from an aliquot (20 µg) of the same total RNA used to construct the 3-year-old SAGE library (Dynabeads mRNA purification kit, Dynal Biotech). First strand cDNA synthesis was carried out with gene-specific primers and reverse transcriptase (Superscript RT-PCR Invitrogen 10928-034) (Table 1). The real-time reverse-transcriptase polymerase chain reaction (qRT-PCR) mix contained 400 nM primer, the cDNA equivalent to 240, 24 or 2.4 ng total RNA and SYBR green (SYBR Green real-time RT-PCR SuperMix-UDG 11733-038) in a reaction volume of 25 µl. Three replications of each dilution for each gene were carried out in an q5 instrument (BioRad) to determine the threshold cycle ($C_T$) values and to construct a standard curve to calculate the amplification efficiencies ($E$). Melting curves for each amplified gene were determined between 55 and 95 °C. An arbitrary expression value ($V$) was calculated for each gene based on $E$ exponential to the inverse power of $C_T$ (Calsa and Figueira 2007). Estimated $V$ was standardized to the reference gene UDP-glucose pyrophosphorylase (UGPase) and the absolute and standardized $V$ were correlated (Pearson’s correlation) to the relative tag frequencies (TPT) from the SAGE 3-year-old cambial region library (Calsa and Figueira 2007).

To confirm the preferential expression of some of the genes identified by SAGE, 10 genes were selected for qRT-PCR analysis (see Table 2) using NADP isocitrate dehydrogenase (IDH) as the reference gene. The expression levels of mRNA from three tissues (cambial region of 3-year-old E. grandis trees, leaves and hypocotyls) were calculated as described above and compared.

Results and discussion

SAGE libraries and Eucalyptus juvenile cambial region expression profile

The 3- and 6-year-old SAGE libraries were produced by sequencing 737 and 703 clones and generating 22,660 and 22,024 tags, respectively. Based on physico-chemical analyses carried out on seven Eucalyptus species, Oliveira et al. (2005) demonstrated that E. grandis trees with a diameter at breast height (DBH) of 30 cm still had juvenile wood characteristics. Because our SAGE libraries were produced from trees at the same developmental stage (DBH less than 30 cm) and the number of tags produced by the two libraries was similar, we combined the sequences from each library and analyzed the data as a single library.

The SAGE 2000 software extracted 43,304 tags from the sequencing data and produced 26,958 tags with three or more copies representing 3066 unique tags or genes. Of these, we assigned ESTs to 659 unique tags. The remaining 2406 tags could not be associated with any sequence because of the small number of publicly available ESTs and complete cDNAs (Table 3). However, because the majority of the available ESTs were sequenced from differentiating xylem, we were able to identify the probable function of 445 genes and produce a transcriptional profile for the cambial region of E. grandis that included the majority of the genes considered to be important during growth and wood formation in Eucalyptus species.

Table 4 shows the 50 most highly expressed genes in the E. grandis juvenile cambial region library to which a putative
identity could be attributed. The most highly expressed gene was an isoflavone reductase belonging to the phenylcoumaran benzylc ether group of reductases that are involved in the biosynthesis of isoflavonoids and lignans (Shoji et al. 2002) and are positively correlated to increased secondary growth in transgenic poplar trees (Israelsson et al. 2003). An ADP-ribosylation factor, involved in vesicle transport and an alpha-tubulin were the second and third most highly expressed genes indicating the importance of cytoskeleton organization and vesicular transport in the cambium. Also included among the 50 most highly expressed genes were cdc2 protein kinase and CHK1 checkpoint protein, both involved in cell division, indicating that our samples contained meristematic cambial cells at the beginning of the differentiation process. Fourteen genes associated with cell wall biosynthesis were among the 50 most expressed genes, two caffeoyl-CoA O-methyltransferases (CCoAOMT), *E. grandis* cellulose synthase 1 gene (EgCesA1), sucrose synthase, coumarate 3-hydroxylase (C3H), endo-1,4-β-glucanase (KOR), caffeic acid 3-O-methyltransferase (COMT), dirigent-like protein (pDIR10), cinnamyl alcohol dehydrogenase (CAD), chitinase-like protein, cinnamate 4-hydroxylase (C4H), pectinesterase and two arabinogalactan proteins (Table 4).

Among the most highly expressed genes was an adenosine kinase whose importance in wood formation has recently been highlighted. Pereira et al. (2006) demonstrated that Arabidopsis mutants with a deficiency in this enzyme produced leaves and hypocotyls with abnormal morphology, probably because of the large quantity of low methyl-esterified pectins in the primary cell wall. These results indicate that adenosine kinases function in the recycling of S-adenosyl-L-methionine-dependent methyltransferases during the process of pectin methyl-esterification and in the determination of cell adhesion properties.

The 445 identified genes were assigned to functional categories as proposed by Rison et al. (2000) (see Supplementary Tables S1 and S2). Because the principal advantage of SAGE is the quantification of the expression level of a large number of genes simultaneously, it was possible to compare expression of groups of functionally related genes by analyzing the number of transcripts per gene in each category or subcategory (Table 5). Based on the transcripts per gene ratio, the most expressed categories were Metabolism and energy and Structure and organization, followed by Transport, information pathways and Cellular processes (Table 5). The analysis also showed that the differences in expression levels within a category were higher than the differences between categories.

Most of the transcripts in Metabolism and energy were related to the subcategories Secondary metabolism, Nitrogen metabolism and Small molecule metabolism (Table 5). Four genes included in the Secondary metabolism subcategory represented 86.8% of all transcripts and showed high similarity to isoflavone reductases. In the subcategory Nitrogen metabolism, two glutamine synthetase (GS1) genes accounted for 93.75% of the transcripts. Similarly, about 81% of the transcripts in the subcategory Small molecule metabolism were represented by a single adenosine kinase gene. Although these genes were included in Metabolism and energy, they also have important roles in the production of precursors required in the process of wood formation.

Cells of the cambial region require energy for maintenance and development; accordingly, among the genes in the subcategory Autotrophic and energy (ATP) metabolism we found genes associated with glycolysis, TCA cycle, alcoholic fermentation and ATP synthesis. In non-photosynthetic organs, carbohydrates are consumed through respiration to produce energy and carbon skeletons for cellular metabolism and biosynthesis of structural molecules, including cell wall polymers. It is possible that the quantity of free O2 in the cambial region is limited by the physical barrier imposed by the bark and by respiratory O2 consumption. If so, a proportion of the energy necessary for secondary growth could be provided by alcoholic fermentation (Kimmerer and Stringer 1988). The possible role of alcoholic fermentation as an alternative or principal supplier of energy in juvenile *E. grandis* trees is indicated by the presence of alcohol dehydrogenase and pyruvate decarboxylase transcripts. Other recent studies have reported the presence of transcripts or proteins associated with anaerobic respiration, particularly alcohol dehydrogenase and pyruvate decarboxylase, during xylem formation and secondary growth (Gion et al. 2005, Juan et al. 2006, Ranik et al. 2006, Celedon et al. 2007).

We found genes representing light-induced proteins and components of photosystems I and II in our library. The presence of functional chloroplasts with active photosystems in the outer peridermal layers (chlorenchyma) and in deeper stem tissues such as ray cells and pith, has been reported in many woody species (Pfanz et al. 2002), which could explain the presence of transcripts related to photosynthesis in our material. Supporting this finding, Celedon et al. (2007) identified Rubisco proteins by LC-MS/MS in the same biological samples used in our study. The outer bark of the stems has a rather

---

Figure 2. Transcript profiling of genes related to the main biosynthetic routes for cell wall components. For each enzyme, the corresponding tag(s) and its expression level are presented. Tags preceeded by an asterisk (*) have been identified as potential alternative transcripts. (a) Pathways related to sugar-nucleotide interconversions, pectin, hemicellulose and cellulose biosynthesis. (b) Pathways related to lignin biosynthesis. Abbreviations: GT8D, glycosyl transferase family 8D; UX5, UDP-xyllose synthase; UGDH, UDP-glucose dehydrogenase; PAL, phenylalanine ammonia-lyase; C4H, cinnamate 4-hydroxylase; 4CL, 4-coumarate: coenzyme A ligase; CCR, cinnamoyl-coenzyme A reductase; CAD, cinnamyl alcohol dehydrogenase; C3H, coumarate 3-hydroxylase; HCT, hydroxycinnamoyltransferase; COMT, caffeic acid/5-hydroxyconiferaldehyde 4-O-methyltransferase; CCoAOMT, caffeoyl-coenzyme A O-methyltransferase; F5H, ferulic acid/coniferaldehyde 5-hydroxylase; SAMS, S-adenosyl-methionine-synthetase and GS1, cytosolic glutamine synthetase.
Table 1. Gene-specific primers used in qRT-PCR validation and their respective amplification efficiencies ($E$). For each gene, the representative tag, the copy number observed in the 3-year-old library and the gi of the template sequence are indicated.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Tag (copy no.)</th>
<th>gi (NCBI)</th>
<th>Primer pairs</th>
<th>$E$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NADP-Isocitrate dehydrogenase ($IDH$)</td>
<td>TACTCAGATG</td>
<td>1750379</td>
<td>Left: CTGTTGAGTCTGGCAAGATGAC Right: CATTTAATTCTCTCCCCAACAAA</td>
<td>1.98</td>
</tr>
<tr>
<td>5-Adenosylmethionine synthetase ($SAMS$)</td>
<td>GTTCGCGCCT</td>
<td>103482634</td>
<td>Left: TGTCCTGTCTCGTGATCTTCA Right: GCCGACTATACCCCAAGTAG</td>
<td>1.93</td>
</tr>
<tr>
<td>Cinnamyl alcohol dehydrogenase ($CADI$)</td>
<td>CCCCCTGTGG</td>
<td>2984652</td>
<td>Left: GCTGATGCTCTCTCCACACAT Right: ATGACATTTCTCCACACAC</td>
<td>1.91</td>
</tr>
<tr>
<td>UDP-Glucose pyrophosphorylase ($UGPase$)</td>
<td>TCGGTCTTAT</td>
<td>103475065</td>
<td>Left: AGATGCTTCAACCGAAATTGT Right: ACCGAGTCTTCGCAAAATTC</td>
<td>1.95</td>
</tr>
<tr>
<td>Pectinesterase</td>
<td>GGAAGGCAG</td>
<td>103474979</td>
<td>Left: ATCCGTTAAATTGTTATTTCTCAG Right: ACACCAACTCAAGAATGG</td>
<td>1.94</td>
</tr>
<tr>
<td>4-Coumarate: coenzyme A ligase ($4CL$)</td>
<td>CTATGTAAT</td>
<td>73665525</td>
<td>Left: ATGGCTGATTGCAGTCTCAG Right: AGAACCAATTAGGATAG</td>
<td>1.88</td>
</tr>
<tr>
<td>Caffeoyl-CoA O-methyltransferase ($CCoAOMT$)</td>
<td>AACCTAGAAA</td>
<td>3319277</td>
<td>Left: TCTTCTGATGACGGAAGATGAT Right: ATGACATTTCTCCACACAC</td>
<td>1.87</td>
</tr>
<tr>
<td>Xylan 1,4-beta-xylosidase</td>
<td>TAGCCCAAAT</td>
<td>103474759</td>
<td>Left: GAGAAACCCCGATATACACCA Right: ATGCCCTGTACTGAAATTGT</td>
<td>1.63</td>
</tr>
<tr>
<td>UDP-Xylose synthase 4 ($UXS4$)</td>
<td>TGACTAACCT</td>
<td>80973755</td>
<td>Left: GAAATCGCTACCTATTTCAT Right: TCTCAGTTGACGTCTTCA</td>
<td>2.36</td>
</tr>
<tr>
<td>Sucrose synthase ($SuSy$)</td>
<td>TGACTAACCT</td>
<td>80973755</td>
<td>Left: GAAATCGCTACCTATTTCAT Right: TCTCAGTTGACGTCTTCA</td>
<td>2.36</td>
</tr>
<tr>
<td>UDP-Arabino 4-epimerase</td>
<td>TATTGTAAC</td>
<td>103480449</td>
<td>Left: ATGGCTGATTGCAGTCTCAG Right: AGAACCAATTAGGATAG</td>
<td>1.87</td>
</tr>
<tr>
<td>Alcohol dehydrogenase ($ADH$)</td>
<td>TATTTCTCGG</td>
<td>103474114</td>
<td>Left: ATGCCCTGTACTGAAATTGT Right: ATGCCCTGTACTGAAATTGT</td>
<td>2</td>
</tr>
<tr>
<td>Pyruvate kinase ($PK$)</td>
<td>TCCTTTTGTG</td>
<td>103478670</td>
<td>Left: GAAATCGCTACCTATTTCAT Right: TCTCAGTTGACGTCTTCA</td>
<td>2.36</td>
</tr>
<tr>
<td>Glutamine synthetase ($GS1$)</td>
<td>TATATCGTCG</td>
<td>103473664</td>
<td>Left: CTGTTGAGTCTGGCAAGATGAC Right: CATTTAATTCTCTCCCCAACAAA</td>
<td>1.98</td>
</tr>
</tbody>
</table>

Table 2. Gene-specific primers used in qRT-PCR validation in different tissues. For each gene, the expressed sequence tag (EST) gi of the template sequence and the amplification efficiencies ($E$) in each tissue are indicated.

<table>
<thead>
<tr>
<th>Gene</th>
<th>EST gi</th>
<th>Primer pairs</th>
<th>Cambial</th>
<th>Leaves</th>
<th>Hypocotyl</th>
</tr>
</thead>
<tbody>
<tr>
<td>NADP-Isocitrate dehydrogenase ($IDH$)</td>
<td>1750379</td>
<td>Left: CTGTTGAGTCTGGCAAGATGAC Right: CATTTAATTCTCTCCCCAACAAA</td>
<td>1.75</td>
<td>1.78</td>
<td>1.80</td>
</tr>
<tr>
<td>Caffeoyl-CoA O-methyltransferase ($CCoAOMT$)</td>
<td>3319277</td>
<td>Left: TCTTCTGATGACGGAAGATGAT Right: ATGACATTTCTCCACACAC</td>
<td>1.86</td>
<td>1.91</td>
<td>1.87</td>
</tr>
<tr>
<td>Caffeic acid/5-hydroxyferulic acid 3/5-O-</td>
<td>437776</td>
<td>Left: AAATCTCATGGAAAGCTGTTA Right: GTCGAAGTTGATCCCTTCAT</td>
<td>1.69</td>
<td>1.70</td>
<td>1.79</td>
</tr>
<tr>
<td>methyletherase ($COMT$)</td>
<td>103482634</td>
<td>Left: TGGCCCTGTCTTGATTTTCA Right: GTCGAAGTTGATCCCTTCAT</td>
<td>1.78</td>
<td>1.81</td>
<td>1.79</td>
</tr>
<tr>
<td>Cellulose synthase 1</td>
<td>67003906</td>
<td>Left: CGATGCAGTATGTGATAG Right: TCCACGCTGAGATGCGGCAT</td>
<td>1.71</td>
<td>1.77</td>
<td>1.80</td>
</tr>
<tr>
<td>Cellulose synthase 4</td>
<td>67003912</td>
<td>Left: AGCCAAAAGCAGAGAAAGTCA Right: ATACGCTGAGATGCGGCAT</td>
<td>1.67</td>
<td>1.75</td>
<td>1.78</td>
</tr>
<tr>
<td>Cellulose synthase 5</td>
<td>67003914</td>
<td>Left: GGGAAAGTTGCGAATAAAGA Right: AACAGGAGACTGACCCGAC</td>
<td>1.75</td>
<td>1.76</td>
<td>1.77</td>
</tr>
<tr>
<td>Endo 1.4-β-glucanase ($KOR$)</td>
<td>70779690</td>
<td>Left: CATCTCCAATCCAAAGGAT Right: TGGACGCTGAGATGCGGCAT</td>
<td>1.69</td>
<td>1.66</td>
<td>1.71</td>
</tr>
<tr>
<td>Sucrose synthase ($SuSy$)</td>
<td>80973755</td>
<td>Left: GAAATCTCTGCTTCCTTTGAG Right: TGGACGCTGAGATGCGGCAT</td>
<td>1.70</td>
<td>1.97</td>
<td>1.95</td>
</tr>
<tr>
<td>Alcohol dehydrogenase ($ADH$)</td>
<td>103474114</td>
<td>Left: TGGCCCTGTCTTGATTTTCA Right: GTCGAAGTTGATCCCTTCAT</td>
<td>1.64</td>
<td>1.75</td>
<td>1.71</td>
</tr>
<tr>
<td>LIM</td>
<td>62087116</td>
<td>Left: GAATCATGAAAAAGCTGTTGAT Right: TGGACGCTGAGATGCGGCAT</td>
<td>1.698</td>
<td>1.736</td>
<td>1.747</td>
</tr>
</tbody>
</table>
low permeability to gaseous diffusion, leading to accumulation of CO₂ in the intercellular air spaces, which may reach concentrations 500–800 times those of ambient air (Pfanz et al. 2002). Thus, stem photosynthesis may reduce tissue anaerobiosis thereby promoting dark respiration and ATP production (Pfanz et al. 2002).

The Cellular processes category was mainly represented by the subcategory Cell division which showed the highest ratio of transcripts per gene (Table 5). Among the genes in this class, were some with similarity to the CDC2 protein kinase genes, CHK1 checkpoint homolog and translationally controlled tumor proteins (TCTP), indicating that our study sample contained the zone of cell division. Although TCTP is believed to be important in cell growth and division in mammalian systems (Bommer and Thiele 2004), its exact role in the plant system remains unknown. In pea, TCTP expression is localized in dividing cells within the root cap, suggesting that these genes, like their mammalian counterparts, are involved in cell proliferation (Woo and Hawes 1997). An elevated number of genes involved in the cell cycle were also found in dividing cambial cells in *Populus* (Hertzberg et al. 2001).

Other genes classified in the Cellular processes category, included those coding for 14-3-3-like proteins and a germin-like protein (GLP). The 14-3-3 proteins are multigenic families present in most organisms and involved in the regulation of multiplex signal transduction events during the activation of a wide variety of target proteins (Ferl 2004). Four genes, with a total of 46 transcripts, were similar to those of 14-3-3-like proteins suggesting a possible role for these proteins in the regulation of *Eucalyptus* juvenile xylem differentiation. The 14-3-3-like proteins were also found in a proteomic study on the same material indicating a direct link between transcript level and protein expression (Celedon et al. 2007). Germins and GLPs are part of the cupin super family that is characterized by two enzymatic functions, oxalate oxidase and superoxide dismutase. Both enzymes catalyze the production of H₂O₂ which could be diverted to the oxidative polymerization of monolignols during lignification (Mathieu et al. 2006). Thus, there may be a link between the expression of GLP in the *E. grandis* cambial region and a role in lignification.

Some genes that are important during PCD and autolysis of vessel elements, such as apoptosis regulator Bcl-2 binding factor (Chintiharlapalli et al. 2005), Radical-Induced Cell Death 1 protein (RCD1) (Ahlfors et al. 2004), cysteine proteases and subtilisin serine protease (Moreau et al. 2005), were represented by a large number of transcripts in our library. This finding indicates that all developmental phases of xylogenesis were included in the material we sampled from the cambial region.

The high transcripts to gene ratios observed in the subcategories Vesicular Transport and Cytoskeleton (Table 5) were expected because cell-wall polysaccharides synthesized in the Golgi complex are exported to the cell matrix in Golgi-derived vesicles through the cytoskeletal network. As well, membrane protein complexes involved in cellulose synthesis are incorporated into Golgi vesicles, transported through the cytoskeleton network and incorporated into the cell membrane by vesicle fusion (Oda et al. 2005). Hertzberg et al. (2001) observed high expression of tubulins in poplar cambial regions C and D, where secondary wall synthesis occurs. Among the genes associated with intracellular vesicle transport, we found members of the small GTPase families, like Ras, Rab and Arf (ADP-rybosylation factors), and annexins (Carrol et al. 1998). We also observed a large number of transcripts for alpha-tubulins, actins, actin depolymerizing factors and profilins.

Two remorins genes were expressed in the cambial material indicating the probable importance of these genes in vascular development. Bariola et al. (2004) used monospecific antibodies to study tissue localization of remorins and observed that the tomato remorin 1 protein is enriched in vascular, meristematic and embryonic-type cells. These authors suggested that remorins are associated with the cytoskeleton or membrane skeleton and help to determine cell integrity or act as scaffolds for signaling in defense or development.

No large differences in transcripts per gene ratios were observed among the subcategories within the Information pathways. However, in the *Eucalyptus* cambial region we observed: class III HD-Zip members similar to the poplar *PtAHB1*, which is developmentally and seasonally regulated, and closely associated with wood formation in *Populus* (Ko et al. 2006); a bHLH transcription factor of *Arabidopsis*; two ethylene transcriptional factors containing the AP2 domain, important in the transition from primary to secondary growth in phloem cells (van Raemdonck et al. 2005); two LIM proteins analogous to the *E. globulus* Lim1; and one *E. grandis* EgMYB1 linked to lignin biosynthesis regulation (Paux et al. 2005). The *Eucalyptus* ESTs corresponding to the LIM sequences are similar to the *Ntlim1* of *Nicotiana tabacum* L., which is positively associated with lignin concentration (Kawaoka and Ebinuma 2001).

The Information pathways category included many genes involved in the proteasome-ubiquitin pathway, the main proteolytic system in plants that is directly or indirectly associated with plant hormones (Hellmann and Estelle 2002). The Ub/26S proteasome system has been identified as playing an important role in the pathways that respond to auxins (Smalle...
The Cell wall subcategory had the second highest transcripts per gene ratio of the Structure and organization category. We identified genes similar to those encoding arabinogalactan proteins, like the poplar FLA13 gene, preferentially expressed in differentiating xylem tissue undergoing differentiation in the xylem (Berleth and Sachs 2001, Paux et al. 2004). The Cell wall subcategory had the second highest transcripts per gene ratio of the Structure and organization category. We identified genes similar to those encoding arabinogalactan proteins, like the poplar FLA13 gene, preferentially expressed in differentiating xylem tissue undergoing differentiation in the xylem (Berleth and Sachs 2001, Paux et al. 2004).
secondary wall thickening (Lafarguette et al. 2004). Although plants contain no endogenous chitin, a gene encoding a chitinase-like protein was expressed with high frequency in our library. Zhang et al. (2004) characterized a group of chitinase-like proteins (CLP) in cotton and showed that they are preferentially expressed during secondary cell wall deposition. Similarly, Aspeborg et al. (2005) observed the expression of a GH19 chitinase-like protein during poplar secondary wall synthesis and suggested its importance in secondary cell wall formation. Thus, the high expression of a chitinase-like protein in

table 5. transcription profile of the identified genes based on functional categories and subcategories.

<table>
<thead>
<tr>
<th>Functional categories</th>
<th>Genes</th>
<th>Transcript no.</th>
<th>Transcripts per gene</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Metabolism and energy</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Autotrophic and energy (ATP) metabolism</td>
<td>61</td>
<td>849</td>
<td>13.91</td>
</tr>
<tr>
<td>Carbon metabolism</td>
<td>1</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Carbohydrate metabolism</td>
<td>23</td>
<td>248</td>
<td>10.78</td>
</tr>
<tr>
<td>Nitrogen metabolism</td>
<td>3</td>
<td>112</td>
<td>37.33</td>
</tr>
<tr>
<td>Amino acid metabolism</td>
<td>18</td>
<td>226</td>
<td>12.55</td>
</tr>
<tr>
<td>Lipid metabolism</td>
<td>4</td>
<td>63</td>
<td>15.75</td>
</tr>
<tr>
<td>Small molecules metabolism</td>
<td>6</td>
<td>118</td>
<td>19.66</td>
</tr>
<tr>
<td>Secondary metabolism</td>
<td>8</td>
<td>560</td>
<td>70</td>
</tr>
<tr>
<td>Total</td>
<td>124</td>
<td>2179</td>
<td>17.57</td>
</tr>
<tr>
<td><strong>Cellular processes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cell division</td>
<td>9</td>
<td>258</td>
<td>28.6</td>
</tr>
<tr>
<td>Signal transduction</td>
<td>26</td>
<td>266</td>
<td>10.23</td>
</tr>
<tr>
<td>Cell regulation</td>
<td>26</td>
<td>235</td>
<td>9.04</td>
</tr>
<tr>
<td>Protection responses/detoxification</td>
<td>22</td>
<td>272</td>
<td>12.36</td>
</tr>
<tr>
<td>Responses to stimuli</td>
<td>3</td>
<td>23</td>
<td>7.66</td>
</tr>
<tr>
<td>Total</td>
<td>86</td>
<td>1054</td>
<td>12.25</td>
</tr>
<tr>
<td><strong>Transport</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Large molecules</td>
<td>5</td>
<td>25</td>
<td>5</td>
</tr>
<tr>
<td>Small molecules</td>
<td>8</td>
<td>86</td>
<td>10.75</td>
</tr>
<tr>
<td>Vesicular</td>
<td>20</td>
<td>439</td>
<td>21.95</td>
</tr>
<tr>
<td>Kinesins/transport factors</td>
<td>3</td>
<td>20</td>
<td>6.66</td>
</tr>
<tr>
<td>Total</td>
<td>36</td>
<td>570</td>
<td>15.83</td>
</tr>
<tr>
<td><strong>Structure and organization</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cell envelope/membrane</td>
<td>7</td>
<td>67</td>
<td>9.57</td>
</tr>
<tr>
<td>Cell wall</td>
<td>41</td>
<td>901</td>
<td>21.97</td>
</tr>
<tr>
<td>Cytoskeleton</td>
<td>15</td>
<td>385</td>
<td>25.66</td>
</tr>
<tr>
<td>Ribosomal RNAs/proteins</td>
<td>41</td>
<td>416</td>
<td>10.14</td>
</tr>
<tr>
<td>Chromosome-related</td>
<td>3</td>
<td>12</td>
<td>4</td>
</tr>
<tr>
<td>Total</td>
<td>107</td>
<td>1781</td>
<td>16.64</td>
</tr>
<tr>
<td><strong>Information pathways</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNA synthesis/degradation/modification</td>
<td>3</td>
<td>32</td>
<td>10.66</td>
</tr>
<tr>
<td>RNA synthesis/degradation/modification</td>
<td>19</td>
<td>230</td>
<td>12.10</td>
</tr>
<tr>
<td>Protein synthesis/degradation/modification</td>
<td>70</td>
<td>931</td>
<td>13.3</td>
</tr>
<tr>
<td>Total</td>
<td>92</td>
<td>1193</td>
<td>12.96</td>
</tr>
<tr>
<td>Total identified</td>
<td>445</td>
<td>6777</td>
<td>–</td>
</tr>
</tbody>
</table>
the local *Eucalyptus* EST database and no significant similarity was found. If MIOX were a key enzyme in the production of UDP-glucuronate in *Eucalyptus* cambium, it would be represented in the xylem-specific EST libraries and, as such, be represented in our SAGE library. Seitz et al. (2000) demonstrated a predominance of UDP-glucose dehydrogenase in *Arabidopsis* meristems, growing roots and vascular tissues, whereas the inositol oxidation pathway was predominant in young hypocotyls and cotyledons.

UDP-Glucose dehydrogenase (UGDH) and UDP-glucur-

![Graph](https://via.placeholder.com/150)

**Table 6. Correlation analysis between SAGE tags frequencies (tags per thousand, TPT) and the expression values obtained from qRT-PCR (amplification efficiency exponential to the inverse power of CT, \( V \)). Correlation coefficients \((r)\) were estimated based on absolute TPT and \( V \) values and on the normalized TPT and \( V \) values (adopted reference gene was *UGPase*).**

<table>
<thead>
<tr>
<th>Gene</th>
<th>( C_T \pm SD )</th>
<th>Absolute</th>
<th>Normalized</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>TPT ( V (10^{-5}) )</td>
<td>TPT ( V (10^{-5}) )</td>
</tr>
<tr>
<td>UDP-Glucose pyrophosphorylase (<em>UGPase</em>)</td>
<td>13.65 ± 0.09</td>
<td>0.6619</td>
<td>10.9908</td>
</tr>
<tr>
<td>Pectinesterase</td>
<td>15.52 ± 0.04</td>
<td>0.3972</td>
<td>3.4076</td>
</tr>
<tr>
<td>Caffeoyl-CoA O-methyltransferase (CCoAOMT)</td>
<td>12.87 ± 0.18</td>
<td>2.1183</td>
<td>16.2593</td>
</tr>
<tr>
<td>( \beta )-Xylo oxidase</td>
<td>20.87 ± 0.63</td>
<td>0.0883</td>
<td>0.1226</td>
</tr>
<tr>
<td>UDP-Xylose synthase 4 (<em>UXS4</em>)</td>
<td>25.89 ± 0.34</td>
<td>0.2206</td>
<td>0.3205</td>
</tr>
<tr>
<td>Sucrose synthase (<em>SuSy</em>)</td>
<td>9.77 ± 0.01</td>
<td>1.2798</td>
<td>22.6352</td>
</tr>
<tr>
<td>UDP-Arabinose 4 epimerase</td>
<td>16.06 ± 0.05</td>
<td>0.2206</td>
<td>6.6686</td>
</tr>
<tr>
<td>Alcohol dehydrogenase (<em>ADH</em>)</td>
<td>14.39 ± 0.15</td>
<td>0.1324</td>
<td>12.2748</td>
</tr>
<tr>
<td>Pyruvate kinase (<em>PK</em>)</td>
<td>14.26 ± 0.07</td>
<td>0.5296</td>
<td>5.0864</td>
</tr>
<tr>
<td>Glutamine synthetase (<em>GS</em>)</td>
<td>12.81 ± 0.10</td>
<td>2.9126</td>
<td>50.0193</td>
</tr>
</tbody>
</table>

\[ r (12) = 0.85, P < 0.01 \]

Figure 3. Comparison of the expression levels of 10 genes in three tissues by qRT-PCR. The y axes represent the expression values \((V)\) \((E – C_T)\) and the x axes represent tissues, where \( C = \) cambial region, \( L = \) leaves and \( H = \) hypocotyls. Abbreviations: CCoAOMT, caffeoyl-CoA \( O \)-methyltransferase; COMT, caffeic acid/5-hydroxyferulic acid 3/5-\( O \)-methyltransferase; SAMS, \( S \)-adenosylmethionine synthase; CesA1, cellulose synthase 1; CesA4, cellulose synthase 4; CesA5, cellulose synthase 5; KOR, endo-1,4-\( \beta \)-glucanase; SuSy, sucrose synthase; ADH, alcohol dehydrogenase; and LIM, Lim gene.
onate decarboxylase (UXS), responsible for UDP-glucuronate and UDP-xylene production, respectively, were represented by highly frequent tags (Figure 2a). However, we observed only one UGDH tag with 22 copies, and the total expression of the three UXS tags totaled 47 copies. This expression pattern is consistent with previous findings of low UGDH activity compared with the activities of other enzymes in subsequent reaction steps in the same pathway, suggesting its rate-limiting function in the synthesis of matrix polysaccharides (Dalessandro and Northcote 1977). Among the three UGDH isoforms present in poplar, only one is up-regulated during secondary cell wall formation (Hertzberg et al. 2001, Johansson et al. 2002) and our tag is similar to the poplar isoform (gi 61645910).

Among the UXS tags (Figure 2a), one (TACTCGGTGG) with 27 copies is associated with the Arabidopsis AtUXS3 soluble form. The other two, occurred at a frequency of 20 copies and are associated with the Arabidopsis AtUX4 Golgi membrane form (Harper and Bar-Peled 2002). A more detailed analysis of the ESTs associated with the AtUXS4 tags revealed that the TCATTATCAA tag was present in both ESTs sequences, 27 and 146 bp from the poly-A, suggesting that two alternative transcripts of AtUXS4-like gene are expressed in E. grandis wood-forming tissues although further results are needed to confirm this.

Besides UDP-xylene production, another important UDP-glucuronate-derived branch leads to pectin metabolism. Pectic polymers are important in primary cell walls where they comprise about 47% of the polysaccharides. Despite the importance of pectin in cell wall structure, only a few genes responsible for pectin biosynthesis have been identified. We found only one tag with low expression associated with the poplar PtTGTD8 gene (Figure 2a), a glycosyl transferase belonging to the same subgroup as Arabidopsis QUASIMODO 1 (QUA1). QUA1 may be involved in pectin biosynthesis because the homogalacturonan-deficient phenotype is shown by QUA1 mutants (Bouton et al. 2002). Another indication that the GTSD gene is involved in pectin synthesis was demonstrated by the positive correlation in the reduction of PtTGTD8 expression and reduced hemicellulose concentration in poplar tension wood (Andersson-Gunneras et al. 2006). In addition, Aspeborg et al. (2005) showed co-expression of GTSD with the secondary-cell-wall-related CesA genes suggesting a potential role in cell wall synthesis.

In dicotyledonous plants, the primary cell wall consists basically of a cellulose microfibril framework embedded in a polysaccharide matrix of pectin and cross-linked glycans (Carpita and Gibeaut 1993). During cell extension, modifications in the structure and composition of the cross-linked pectin xyloglucans occur (Bourquin et al. 2002). For example, xyloglucan endotransglycosylases (XETs) are responsible for cell wall remodeling during primary cell wall biosynthesis by cutting and rejoining the xyloglucan chains. During secondary cell wall deposition, XETs create and reinforce the connections between primary and secondary wall layers (Bourquin et al. 2002). Despite the importance of XETs in cell wall remodeling, we found only one XET tag with 19 copies. An increased number of transcripts for pectinesterases and pectate lyases were observed during increased secondary growth in poplar tension wood (Andersson-Gunneras et al. 2006). Under the normal growth conditions, we observed two pectinesterases tags at a frequency of 35 and 11 copies and one pectate lyase tag with four copies in the cambial region library. In contrast to the 11 highly expressed genes identified in active poplar cambium by Geisler-Lee et al. (2006), we found only one tag representing polygalacturonase with low expression in our library (Figure 2a).

**Cellulose biosynthesis**

Current models of cellulose biosynthesis involve both CesA proteins and membrane-associated proteins like CORRIGAN (endo-1,4-β-glucanase) and SuSy (sucrose synthase) (Joshi et al. 2004). Five tags corresponding to E. grandis EgCesA genes were identified, one showing high similarity to the primary-cell-wall-related cellulose synthase gene E. grandis E. CesA4, and four showing high similarity to the secondary-cell-wall-related genes EgCesA1, EgCesA2 and EgCesA3 (Figure 2a). Two tags (AATTGATATG and GAATCAAAAT) represented the EgCesA1 gene, the first tag occurred in both EST sequences at a distance of 151 and 30 bp from the poly-A tail, indicating possible alternative transcripts.

According to Ranik and Myburg (2006), genes implicated in Eucalyptus secondary wall formation (EgCesA1, EgCesA2 and EgCesA3) have higher expressions than genes involved in primary cell wall formation (EgCesA4, EgCesA5 and EgCesA6) in xylem. We observed similar transcriptional profiles for primary cell wall EgCesA genes. The EgCesA4 gene showed low expression (Figure 2a), and the EgCesA5 and EgCesA6 genes were represented as single-copy transcripts (data not shown). The higher expression of the secondary-cell-wall-associated CesA genes was expected because our sample contained more cells from the xylem-side of the cambial region. Ranik and Myburg (2006) reported that EgCesA3 gene was the most abundantly expressed CesA gene in secondary xylem. In contrast, our results indicated that EgCesA1 was the most abundantly expressed (a total of 71 copies) cellulose synthase gene in the E. grandis juvenile-wood-forming tissue (Figure 2a). This discrepancy is of interest because Ranik et al. (2006), using cDNA-AFLPs identified several transcript-derived fragments from immature and mature xylem of a highly productive E. grandis × E. nitens hybrid tree, reported that a secondary-cell-wall cellulose synthase gene, similar to Arabidopsis AtCesA8, Populus PtcCesA1 and Eucalyptus EgCesA1 (Ranik and Myburg 2006), was strongly up-regulated in xylem tissues.

The pool of UDP-glucose destined for cellulose synthesis can be produced either by UDP-glucose pyrophosphorylase or by sucrose synthase. In our study, the higher frequency of SuSy transcripts compared with UDP-glucose pyrophosphorylase transcripts indicates that SuSy activity is probably the main source of UDP-glucose for cellulose synthesis in differentiating xylem of E. grandis (Figure 2a). SuSy transcripts were the most abundant CAZyme transcripts in poplar and have been shown to be highly expressed during secondary cell wall remodeling.
The membrane-associated SuSy was detected in developing cotton fibers giving rise to a functional model for cellulose biosynthesis where this enzyme directly channels UDP-glucose to the membrane-bound cellulose synthesis complex, thereby avoiding competition from the cellular metabolic pool of UDP-glucose, and facilitating more efficient cellulose synthesis (Amor et al. 1995), which is important during active secondary growth.

The primer-mediated model for cellulose biosynthesis considers that an endo-1,4-β-glucanase (KORRIGAN) cleaves the growing β-1,4-glucan chain from the primer and transfers the chain to another CesA protein allowing further glucan chain elongation (Doblin et al. 2002). Supporting this idea, an endo-1,4-β-glucanase gene was shown to be preferentially expressed in Eucalyptus secondary xylem (Paux et al. 2004).

Our results indicate a role for endo-1,4-β-glucanase in E. grandis wood-forming tissue because of the large number of transcripts observed (Figure 2a). The more frequent KORRIGAN tag (41 copies) is similar to the E. globulus endo-1,4-β-glucanase gene (EG2), whereas the other tag (three copies) is similar to EG1. In general, the similar expression levels of the genes encoding KORRIGAN, EgCesA1 and SuSy (Figure 2a) fit the proposed model for cellulose biosynthesis (Joshi et al. 2004).

**Lignin biosynthesis**

The phenylpropanoid pathway starts with the deamination of phenylalanine to cinnamic acid by phenylalanine ammonia-lyase (PAL). Cinnamic acid is then converted to coumaric acid by cinnamate 4-hydroxylase (C4H) and diverted to monolignol synthesis (Figure 2b). Both PAL and C4H were represented by highly frequent tags (Figure 2b). Two PAL genes, one with 30 copies, showed high similarity to the Arabidopsis AtPAL1 gene, and one with only three copies, was similar to both AtPAL1 and AtPAL2. The alignment of the two ESTs indicates that they represent different genes. AtPAL1 and AtPAL2 are believed to be the most important genes in lignin synthesis during vascular lignification among the four Arabidopsis PAL genes (Raes et al. 2003). An EST representing C4H (with 36 copies) was similar to the poplar PtreC4H2 gene, which is more xylem specific than the PtreC4H1, and more weakly expressed in phloem (Lu et al. 2006).

Only one tag representing the 4-coumarate:coenzyme A ligase (4CL) gene (Figure 2b) was identified with high similarity to the 4CL1 gene from E. camaldulensis Dehn. The 4CL enzyme produces CoA thioesters of the hydroxycinnamic acids p-coumaryl alcohol, coniferyl alcohol and sinapyl alcohol which are precursors of p-hydroxyphenyl (H), guaiacyl (G) and syringyl (S) lignin subunit synthesis, respectively. A poplar 4CL gene was reported to be up-regulated in the cambial zone undergoing lignification (Hertzberg et al. 2001) confirming its role in the initial steps of lignin synthesis. Lignin polymers in angiosperm wood are composed of large amounts of G and S units, whereas only small amounts of H units are added. This is because H units are predominantly deposited in the middle lamella and cell corners followed by G units, which are mainly laid down in the secondary wall, and S units that are deposited during the late stages of lignification (Lewis 1999). Thus, it is expected that the key enzymes in the biosynthesis of the monolignols G and S will show higher expression levels during secondary growth. Genes for caffeic acid/5-hydroxyferulic acid 3/5-O methyltransferase (CCoAOMT), caffeoyl coenzyme A O-methyltransferase (CCoAOMT) and S-adenosylmethionine synthase (SAMs) showed higher expression levels than those for 4CL and cinnamoyl-CoA reductase (CCR) (Figure 2b). These results agree with Paux et al. (2005), who suggested that the expression of 4CL, CCR and CAD is under a common transcriptional control while COMT and CCoAOMT form another co-regulated transcriptional cluster. In support of this hypothesis, similar expression profiles for the COMT and CCoAOMT genes were also observed by Hertzberg et al. (2001).

The determining step for the diversion of p-coumaryl-CoA for G and S monolignols synthesis is its conversion to p-coumaroyl shikimic acid/quinic acid by hydroxycinnamoyltransferase (HCT), because p-coumarate 3-hydroxylase (C3H) cannot use p-coumaryl-CoA as substrate (Schoch et al. 2001). It has been shown that p-coumaroyl shikimate and p-coumaroyl quinate are important intermediates in the phenylpropanoid pathway with HCT acting both upstream and downstream of C3H in the production of caffeoyl CoA (Hoffmann et al. 2004). The EST representing the C3H gene was similar to the Arabidopsis CYP98A3 gene whose expression is more evident in lignifying vascular cells (Nair et al. 2002), and its expression was almost four times greater than that of HCT (Figure 2b). Alternatively, C3H can act on p-coumaric acid precursors producing caffeic acid which in turn can be diverted to ferulate, by COMT, or to caffeoyl CoA, by 4CL (Figure 2b).

Although COMT was first believed to convert caffeic acid into ferulate (Dixon 2001), it was subsequently shown that COMT preferentially catalyzes the conversion of 5-hydroxyferulate, 5-hydroxyconiferaldehyde and 5-hydroxyconiferyl alcohol into sinapic acid, sinapaldehyde and sinapyl alcohol, respectively, and thus acts preferentially on ferulic acid and coniferaldehyde 5-hydroxylase (F5H) derived products (Parvati et al. 2001). A differential regulation for F5H and COMT genes is supported by recent findings obtained through a proteomics approach where F5H proteins were not found among the expressed proteins during poplar cambial regeneration, whereas COMT isoforms were detected at all stages (Juan et al. 2006). Consistent with this result, our data showed that COMT transcript abundance was almost twice that of the F5H transcripts (Figure 2b).

Based on our results, CCoAOMT has the highest expression level of all the genes involved in lignin biosynthesis in E. grandis wood-forming tissue (Figure 2b). This finding is in agreement with the results of Paux et al. (2004) who reported a preferential expression of CCoAOMT during Eucalyptus wood formation. CCoAOMT adds a methyl radical to caffeoyl CoA, producing feruloyl CoA in an alternative route for monolignol production (Zhong et al. 1998). The importance of
this alternative route was demonstrated by the down regulation of CCoAOMT in transgenic tobacco and poplar plants leading to an altered S/G ratio and significantly decreased lignin content (Zhong et al. 1998, Zhong et al. 2000). Because the number of CCoAoMT transcripts is higher than the number of COMT transcripts, the S and G units are probably synthesized preferentially from caffeoyl CoA in *E. grandis*.

Expression of both the CCoAoMT and SAMS genes was high in our library (Figure 2b), although a common transcriptional regulatory system for these genes is unknown. Although SAMS is a housekeeping enzyme involved in methionine metabolism, its activity has been detected in xylem tissue undergoing secondary growth (Juan et al. 2006). According to Cantón et al. (2005), SAMS may provide methyl groups consumed locally by CCoAoMT and COMT, thus ensuring high rates of lignification. Confirming the importance of SAMS in xylem formation/lignification, three spots representing this protein were identified in a proteomics study with the same 3- and 6-year-old *Eucalyptus* juvenile cambial region (Celedon et al. 2007).

During the first step in phenylpropanoid biosynthesis, catalyzed by PAL, extensive amounts of ammonium are liberated by phenylalanine deamination, and an efficient system to recycle nitrogen is needed to prevent a severe N deficiency in plants during active lignification (Cantón et al. 2005). Cantón et al. (2005) proposed a mechanism in which the liberated nitrogen is locally recycled and re-incorporated into glutamine by GS1 (glutamine synthetase). Our results indicate a role for the GS1 gene during lignification because GS1 showed high expression (101 copies), the fifth highest among the 50 most expressed genes in the *E. grandis* juvenile cambial region library (Figure 2b), which is in agreement with the results of Celedon et al. (2007) showing the presence of GS1 in the *Eucalyptus* cambial region.

The final step in lignin biosynthesis is the polymerization of monolignols catalyzed by peroxidases and laccases (Baucher et al. 2003). No peroxidases were represented in our library, even though two *E. globulus* peroxidases (gi 88659655 and 88659653) are present in the *Eucalyptus* EST database. This finding suggests low expression of these isoforms or expression of other peroxidase isoforms unrepresented in public databases and does not rule out the participation of peroxidases during monolignol polymerization. Although the role of laccases in lignin polymerization remains a matter of debate, their importance during wood formation has been reported (Paux et al. 2004), and our results corroborate the potential role of laccases in lignin polymerization during *Eucalyptus* wood formation (Figure 2b). Supporting this idea, no significant decrease in peroxidase expression was observed in poplar tension wood (characterized by decreased lignin content), whereas the laccase gene *lac3* was co-regulated with the lignin biosynthesis genes (Andersson-Gunneras et al. 2006).

Hertzberg et al. (2001) demonstrated the induction of a dirigent-like protein during lignification and secondary xylem cell wall formation. Dirigent proteins can bind to sites localized mainly in the S(1) sublayer and mediate monolignol coupling during the assembly of the lignin biopolymer (Burlat et al. 2001). We observed two tags representing dirigent protein genes from *Picea glauca*, *pDIR18* with low expression and *pDIR10* with higher expression (Figure 2b). Thus, in accordance with their probable functions and their expression levels in our study, it seems likely that both laccases and dirigent proteins are involved in lignification in *E. grandis* juvenile wood-forming tissues.

In conclusion, we used SAGE to investigate the expression of genes involved in wood formation in the cambial region of juvenile *E. grandis* trees. We identified 445 genes that represented all functional categories necessary for the maintenance of an actively growing tissue. Our data offer insight into the expression of functionally related genes directly and indirectly involved in cell wall biosynthesis during the first years of development of fast-growing *Eucalyptus* trees.

**Acknowledgments**

This work was supported by FAPESP Innovation Technology (proc. 01/11080-8), CNPq and CAPES. We thank Suzano Papel e Celulose for the biological material and financial support. We also thank Juliano Bragatto and Maria A.R. Chavez Bermudez for the anatomical illustrations, and Dr. João Lúcio de Azevedo (Laboratório de Genética de Microrganismos-ESALQ/USP) for the use of the iQ5 instrument (BioRad).

**References**


Supplementary material

Table S1. List of the 445 annotated sequences divided into functional categories based on their putative assigned functions. The table shows the original tag sequence, tag count and the BLASTx result. Available at:

http://www.heronpublishing.com/tree/supplementary/28-905/28-905_TableS1.pdf

Tables S2. The NCBI Accession numbers of all the publicly available Eucalyptus EST and cDNA sequences used to construct our local database. Available at:

http://www.heronpublishing.com/tree/supplementary/28-905/28-905_TableS2.pdf