



## Genes encoding enzymes of the lignin biosynthesis pathway in *Eucalyptus*

Ricardo Harakava

*Instituto Biológico, Laboratório de Bioquímica Fitopatológica, São Paulo, SP, Brazil*

### Abstract

*Eucalyptus* ESTs libraries were screened for genes involved in lignin biosynthesis. This search was performed under the perspective of recent revisions on the monolignols biosynthetic pathway. *Eucalyptus* orthologues of all genes of the phenylpropanoid pathway leading to lignin biosynthesis reported in other plant species were identified. A library made with mRNAs extracted from wood was enriched for genes involved in lignin biosynthesis and allowed to infer the isoforms of each gene family that play a major role in wood lignin formation. Analysis of the wood library suggests that, besides the enzymes of the phenylpropanoids pathway, chitinases, laccases, and dirigent proteins are also important for lignification. Colocalization of several enzymes on the endoplasmic reticulum membrane, as predicted by amino acid sequence analysis, supports the existence of metabolic channeling in the phenylpropanoid pathway. This study establishes a framework for future investigations on gene expression level, protein expression and enzymatic assays, sequence polymorphisms, and genetic engineering.

**Key words:** *Eucalyptus*, lignin, expressed sequenced tag, phenylpropanoids pathway, monolignols.

Accepted: May 31, 2004; Received: March 17, 2005.

### Introduction

Lignin, after cellulose, is the second most abundant terrestrial organic polymer, accounting for up to 30% of all vascular plant tissue. Deposition of lignins reinforces plant cell walls, providing rigidity, impermeability to water, and protection against pathogens. Lignins are complex racemic aromatic heteropolymers that, in Gymnosperms, derive mainly from coniferyl alcohol and a small proportion of *p*-coumaryl alcohol, and in Angiosperms, from approximately equal parts of coniferyl and sinapyl alcohols. These monolignols are products of the phenylpropanoid metabolism, which is initiated by deamination of phenylalanine by the enzyme phenylalanine ammonia-lyase (PAL). A series of hydroxylation and *O*-methylation reactions, and conversion of side-chain carboxyl to an alcohol results in the building blocks of lignins (Figure 1; Humphreys and Chapple, 2002). In the traditional view, this series of reactions occurred at the level of free hydroxycinnamic acids, but recent discoveries led to a reformulation of the pathway where hydroxycinnamic acid esters play a central role (Humphreys and Chapple, 2002; Boerjan *et al.*, 2003).

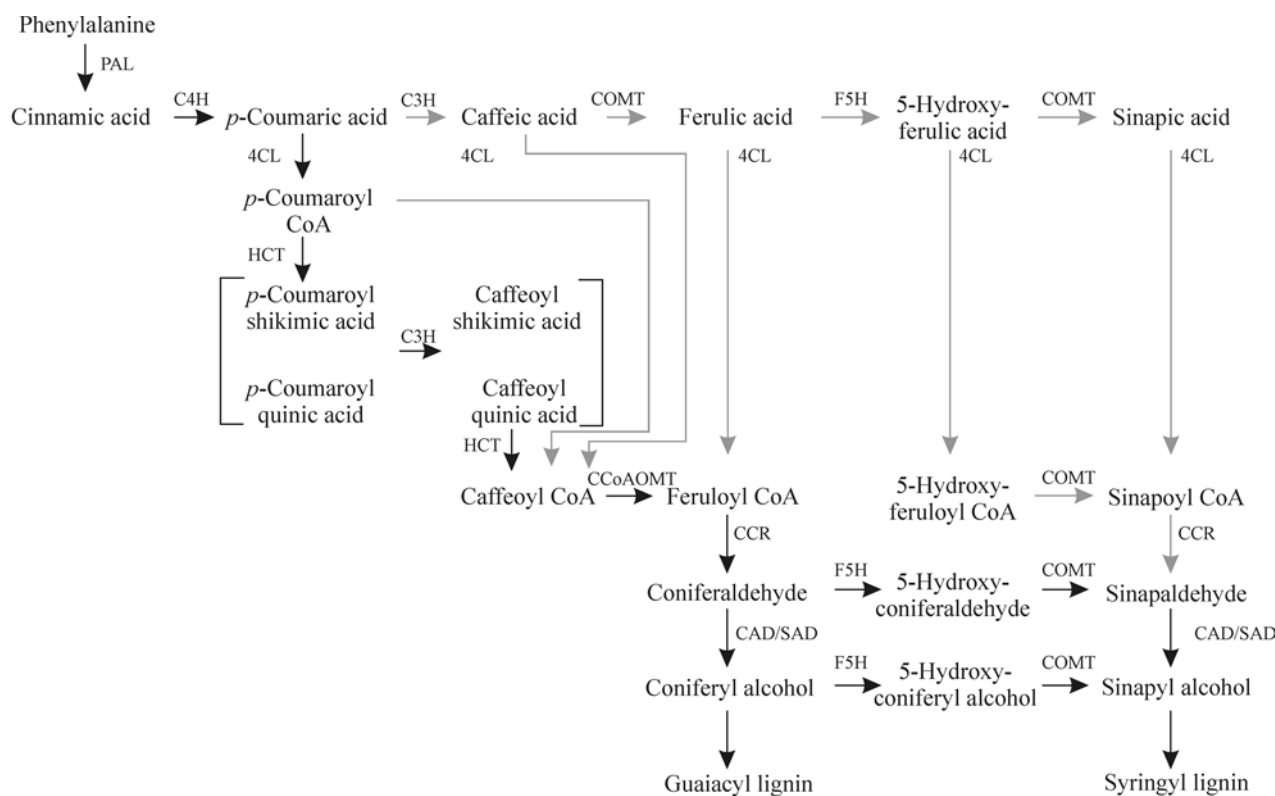
Removal of lignin for cellulose production requires large amounts of chemical and energy input, resulting in financial and ecological costs. Efforts are being made worldwide towards the development of tree varieties with less

and/or modified lignin which would enhance wood-pulp production efficiency. Transgenic trees with enhanced or reduced expression of different gene encoding enzymes of the phenylpropanoid metabolism show that it is possible to develop varieties with less and/or more easily extractable lignin (Pilate *et al.*, 2002; Li *et al.*, 2003; Baucher *et al.*, 2003).

In Brazil, *Eucalyptus* is the main source of cellulose for the paper industry. Breeding programs resulted in fast growing and high-yield varieties and is one of the reasons Brazil is a world leader in cellulose and paper production from *Eucalyptus*. Further improvements will require better understanding of the factors that control wood properties and the recently completed ESTs sequencing project (FOR-ESTs) is a invaluable tool to achieving this goal.

### Material and Methods

Fifteen *Eucalyptus grandis* ESTs libraries from different tissues, developmental stages, and growth conditions were mined for genes encoding enzymes of the phenylpropanoid metabolism or involved in lignin biosynthesis. Four libraries made from dark grown seedlings of *E. globulus*, *E. saligna*, *E. urophylla*, and *E. camaldulensis* were also included. These libraries contained altogether 123,889 reads that produced 33,080 clusters (17,286 singlets + 15,794 clusters). Protein sequences of enzymes already described for *Arabidopsis thaliana*, *Populus* spp., *Eucalyptus* spp, *Nicotiana tabacum*, and a few other plant



**Figure 1** - Phenylpropanoid pathway leading to monolignol precursors of lignin. Black arrows indicate the current revised pathway, gray arrows indicate previous interpretation of the pathway (adapted from Humphreys and Chapple, 2002). 4CL, 4-hydroxycinnamoyl CoA ligase; C3H, *p*-coumarate 3-hydroxylase; C4H, cinnamate 4-hydroxylase; CAD, cinnamyl alcohol dehydrogenase; CCoAOMT, caffeoyl CoA *O*-methyltransferase; CCR, cinnamoyl CoA reductase; COMT, caffeic acid *O*-methyltransferase; F5H, ferulate 5-hydroxylase; HCT, hydroxycinnamoyl-CoA:shikimate/quinic acid hydroxycinnamoyltransferase; PAL, phenylalanine ammonia-lyase; SAD, sinapyl alcohol dehydrogenase.

species were used to query the ESTs clusters database using T-Blast-n (Altschul *et al.*, 1997). Sequences alignment and comparisons were made using Clustal-X (Thompson *et al.*, 1997) and BioEdit (Hall, 1999). Cell compartment localization of the encoded protein was determined using PSORT (Nakai and Kanehisa, 1992). Special attention was given to the WD2 library (10,224 reads, 1346 singlets + 4164 clusters), produced with mRNA extracted from wood, which could be enriched for genes encoding enzymes involved in lignin biosynthesis. For each cluster mentioned in the text, total number of reads followed by number of reads from WD2 library are provided.

## Results and Discussion

### Phenylalanine ammonia-lyase - PAL

PAL is the first enzyme of the phenylpropanoid pathway and catalyzes the deamination of phenylalanine to produce trans-cinnamic acid. Twelve clusters with high similarities to PAL genes were found. None of them comprehended a full-length gene. This can be explained by the fact that PAL enzymes are large proteins (> 70 kDa), encoded by genes of more than 2 kbp, and most of the ESTs corresponding to either the 5' or 3' ends did not reach each other. Four clusters comprehended only one read each and

may contain sequencing errors. Of the eight clusters remaining, six contained several reads (7 to 66) and were considered robust. When the amino acid sequences encoded by these six clusters were aligned with known PAL protein sequences from *A. thaliana*, *Populus* spp., and *Eucalyptus globulus*, three pairs of clusters, each corresponding to the N- and C-termini of the protein, could be formed. The pair EGEQRT3100H09.b (66 reads, 23 in WD2)-EGBMRT3131H09.g (18 reads, 12 in WD2), corresponding to the N- and C-termini respectively, is the more intensely expressed PAL gene. The high number of reads from the WD2 library suggests that it is the main PAL enzyme involved in wood lignin biosynthesis. This pair is also the most closely related (~97.7 % amino acid sequence identity) to an *E. globulus* partial PAL sequence available in the GenBank (AF167487). The pair EGUTRT3109B 10.g (11 reads, 4 in WD2)-EGMCRT3145G11.g (17 reads, 6 in WD2) is the second most expressed PAL gene. The pair EGABLV 2284G08.g (7 reads, 3 in WD2)-EGEZFB 1045G07.g (8 reads, 0 in WD2) corresponds to the least expressed PAL gene. Analysis of the N-terminus of each PAL protein using PSORT indicated that all are directed to the endoplasmic reticulum (ER).

Similarly to other plant species studied, in *Eucalyptus*, PAL is encoded by a small multigene family. Four PAL

isoenzymes were identified in parsley (*Petroselinum crispum*), *Populus kitakamiensis*, and *A. thaliana* (Appert *et al.*, 1994; Osakabe *et al.*, 1995; Cochrane *et al.*, 2004). The existence of multiple isoforms of PAL allows regulatory flexibility where each member is preferentially expressed in response to developmental, environmental, or metabolic needs (Kumar and Ellis, 2001; Kao *et al.*, 2002; Costa *et al.*, 2003, Raes *et al.*, 2003).

#### Cinnamate 4-hydroxylase - C4H

C4H belongs to the CYP73A group of cytochrome P450-dependent monooxygenases protein family. It hydroxylates cinnamic acid to generate *p*-coumaric acid. Two clusters encoding C4H enzymes were found in *Eucalyptus*. Both EGEQRT3101D09.b (72 reads, 36 in WD2) and EGJ MCL1327G07.g (15 reads, 11 in WD2) contained full-length genes encoding proteins that share amino acid sequence identity of 59.8%. PSORT predicts the former to be localized to the ER and the latter to the plasma membrane. Accordingly, a C4H protein from hybrid poplar (*Populus balsamifera* subsp. *trichocarpa* x *Populus deltoides*; AAG50231), highly similar to the protein encoded by cluster EGEQRT3101D09.b (90.8% identical), has been demonstrated to be localized to the ER (Ro *et al.*, 2001). EGEQRT3101D09.b encoded protein belongs to class I C4Hs while EGJMCL1327G07.g groups with class II C4Hs (Raes *et al.*, 2003). In citrus and French bean, class II C4Hs are induced by elicitor treatment (Nedelkina *et al.*, 1999; Betz *et al.*, 2001). In *Eucalyptus*, besides WD2 library, class II C4H reads were found only in CL1 and CL2 libraries, made from callus tissue grown under dark and light, respectively, while class I C4H reads were more widely distributed among the different libraries. This suggests that, resembling other plant species, *Eucalyptus* class II C4H is particularly involved in stress responses but it is also involved in wood lignin biosynthesis and class I is constitutively expressed in any tissue requiring phenylpropanoid metabolites.

#### 4-Hydroxycinnamoyl CoA ligase - 4CL

4CL is responsible for the CoA esterification of *p*-coumaric acid, caffeic acid, ferulic acid, 5-hydroxyferulic acid, and sinapic acid. In *Eucalyptus*, 3 clusters encoding 4CLs were found. EGEQRT3001E07.b (46 reads, 22 in WD2), appears to be the main isoenzyme involved in lignin biosynthesis and encodes a class I 4CL (Raes *et al.*, 2003). The encoded protein is predicted to be targeted to the ER. Clusters EGSBRT3310C08.g (2 reads, 0 in WD2) and EGJMCL1315H12.g (1 read, 0 in WD2) encode N- and C-termini, respectively, of a class II 4CL. At least other 7 clusters encode 4CL-like proteins, among which cluster EGBFFB1044B05.b (13 reads, 3 in WD2) is the only one that contains reads from WD2 library and may have a role in lignin biosynthesis.

#### Hydroxycinnamoyl-CoA:shikimate/quinate hydroxycinnamoyltransferase - HCT

The first HCT was recently purified and the corresponding gene cloned from tobacco (Hoffmann *et al.*, 2003). This enzyme converts *p*-coumaroyl-CoA and caffeoyl-CoA to their corresponding shikimate or quinate esters and catalyzes the reverse reaction as well. Shikimate and quinate esters of *p*-coumaroyl-CoA have been shown to be preferred substrates for *p*-coumarate 3-hydroxylase (C3H), which converts them into their corresponding caffeoyl esters (Schoch *et al.*, 2001). Six clusters encoding HCT proteins were found in *Eucalyptus*. Among these, only EGBMRT3128A11.g (6 reads, 2 in WD2) corresponded to a full-length gene while cluster EGABRT3012D12.g (19 reads, 10 in WD2), encoding 342 amino acids of the protein N-terminus, seems to be the main HCT isoenzyme involved in wood lignin formation. Sequence comparison shows that the protein encoded by EGBMRT3128A11.g has a slightly higher level of identity, 77.8%, to the functionally characterized tobacco HCT, than the one encoded by EGABRT 3012D12.g, which shows 72.3% identity. Hoffman *et al.* (2003) have shown that tobacco HCT has a pronounced preference for shikimic acid versus quinic acid as acyl acceptors, thus it would be interesting to investigate whether the more divergent *Eucalyptus* HCT has an opposite preference. Cluster EGEQRT3101E12.b (9 reads, 1 in WD2) may correspond to the C-terminus of EGABRT3012D12.g. The three clusters remaining have few reads each and may contain either sequencing or assembling errors.

#### *p*-Coumarate 3-hydroxylase - C3H

C3H belongs to the CYP98A3 group of cytochrome P450-dependent monooxygenases family. Although its name indicates *p*-coumaric acid as substrate, as commented before, the shikimate and quinate esters of this acid are the substrates instead (Schoch *et al.*, 2001). Two bona fide clusters encoding C3H proteins were found. EGSBRT3314B10.g (30 reads, 17 in WD2) is a full-length gene and EGEQCL2001E08.g (11 reads, 8 in WD2) encodes 253 N-terminal amino acids of the protein. The protein encoded by cluster EGSBRT3314B10.g shows 77.2% sequence identity to the functionally characterized *A. thaliana* C3H (Schoch *et al.*, 2001; Franke *et al.*, 2002; Nair *et al.*, 2002), while the peptide encoded by EGEQCL2001E08.g shows an identity of 70.7%. PSORT indicates ER localization for both proteins. Since both shikimate and quinate esters of *p*-coumaric acid are substrates for C3Hs in the lignin biosynthetic pathway, it is tempting to speculate that the two *Eucalyptus* isoenzymes have differential specificity towards these compounds.

#### Caffeoyl CoA *O*-methyltransferase - CCoAOMT

CCoAOMT catalyzes methylation of caffeoyl CoA to generate feruloyl CoA. Four clusters encoding CCoAOMT

proteins were found in *Eucalyptus*. EGBMRT3129H01.g (37 reads, 13 in WD2) and EGEQST2200H08.g (24 reads, 11 in WD2) are full-length genes and apparently are the main CCoAOMT genes in wood lignin formation. The encoded proteins share 88.2% amino acid sequence identity, and PSORT predicts a cytoplasmic localization for both. Both *Eucalyptus* CCoAOMT isoenzymes have high sequence identity (> 87%) to the experimentally characterized one from poplar (O65862; Meyermans *et al.*, 2000). EGB MRT3129H01.g is orthologous to *Eucalyptus globulus* CCoAOMT-2 (GenBank AAD50443, 99.1% identical) and EGEQST2200H08.g is orthologous to *E. globulus* CCoAOMT-1 (AAC26191, 99.5% identical) and *E. gunnii* CCoAOMT (CAA72911, 98.5% identical). Cluster EGRFST 6280E09.g (2 reads, 0 in WD2) shows a 99 nucleotide internal deletion when compared to EGEQST2200H08.g but are otherwise identical. Cluster EGRFRT3333F04.g (2 reads, 0 in WD2) encodes 143 N-terminal amino acids of a protein that is 52% identical to both EGBMRT3129H01.g and EGEQST2200H08.g.

#### Cinnamoyl CoA reductase - CCR

CCR converts hydroxycinnamoyl CoA esters to their corresponding aldehydes. Cluster EGEQRT5001C10.g (14 reads, 0 in WD2) encodes a protein 74.7% identical to *A. thaliana* CCR1 (AAG46037), which plays a major role in lignification (Jones *et al.*, 2001; Raes *et al.*, 2003). EGEQRT5001C10.g is orthologous to previously cloned CCRs from *E. gunnii* (CAA56103, 98.8% identical), *E. saligna* (AAG16242, 99.1% identical), and *E. globulus* (AAM34502, 98.5% identical). Cluster EGBMRT6276A06.g (7 reads, 0 in WD2) encodes a CCR-like protein 40.4% identical to EGEQRT5001C10.g. It is noteworthy that no CCR encoding reads were found in the WD2 library but given the homology of EGEQRT5001C10.g to the previously cloned CCR gene from *E. gunnii*, which has been functionally characterized (Lacombe *et al.*, 1997), this cluster is likely to encode the main or exclusive CCR involved in lignin biosynthesis. Most of the CCR-encoding reads were from a root cDNA library (7 out of 14).

#### Ferulate 5-hydroxylase - F5H

F5H is a cytochrome P450-dependent monooxygenase of the CYP84 group that converts ferulic acid to 5-hydroxyferulic acid but in vitro studies using recombinant F5Hs from *A. thaliana* and sweetgum (*Liquidambar styraciflua*) have shown that it preferentially converted coniferaldehyde and/or coniferyl alcohol to synapaldehyde and/or sinapyl alcohol, respectively (Humphreys *et al.*, 1999; Osakabe *et al.*, 1999). For this reason F5H is also designated coniferaldehyde 5-hydroxylase (CAld5H). Three clusters encoding F5H isoenzymes were found in *Eucalyptus*: EGBFFB1040D04.g (67 reads, 54 in WD2); EGE QRT3200F07.g (27 reads, 0 in WD2); and EGCCRT6008 B11.g (7 reads, 0 in WD2). The elevated

number of reads from the WD2 library in cluster EGBFFB1040D04.g indicates the importance of this isoenzyme for lignin biosynthesis. The encoded protein has sequence identity of 76.5% and 75.0% to functionally characterized F5Hs from sweetgum (AAD48912) and *A. thaliana* (Q42600) respectively. PSORT predicts ER localization for the protein encoded by EGBFFB1040D04.g.

#### Caffeic acid O-methyltransferase - COMT

COMT was previously thought to be a bi-functional enzyme, methylating caffeic and 5-hydroxyferulic acids. More recently, in vitro studies using recombinant alfalfa (*Medicago sativa*) and sweetgum COMTs have shown that preferential substrates are 5-hydroxyconiferaldehyde and/or 5-hydroxyconiferyl alcohol, resulting in sinapaldehyde and/or sinapyl alcohol, respectively (Osakabe *et al.*, 1999; Parvathi *et al.*, 2001). Cluster EGEQRT3104D08.g (127 reads, 69 in WD2) encodes a protein with high identity to functionally characterized COMTs from alfalfa (P28002 - 80.8%) and sweetgum (AAD48913 - 78.2%). Among clusters encoding lignin biosynthetic pathway enzymes, COMT has the highest number of reads from the WD2 library. It is orthologous (99.1% amino acid sequence identity) to the previously cloned COMT from *E. gunnii* (Poeydomenge *et al.*, 1994). At least 14 other clusters encoding COMT-like proteins were found in *Eucalyptus* but only EGRFST7274F09.g (8 reads, 4 in WD2) and EGSBRT6210D12.g (3 reads, 1 in WD2) contained reads from the WD2 library.

#### Cinnamyl alcohol dehydrogenase - CAD

CADs catalyze the conversion of cinnamyl aldehydes into their corresponding alcohols. Plants show a large variety of CADs that reduce a wide range of aldehydes, many of which are expressed in response to pathogen infection (Walter *et al.*, 1988). In *A. thaliana*, 9 putative CAD genes were detected and separated into 3 classes (Raes *et al.*, 2003). Several clusters encoding putative CAD proteins were found in *Eucalyptus*. Only two among these, EGEQ RT3001D12.b (65 reads, 18 in WD2) and EGEQFB1004E07.g (28 reads, 5 in WD2), contained reads from the WD2 library, thus are candidates for playing a role in wood lignin biosynthesis. Indeed, among all *Eucalyptus* CAD proteins, clusters EGEQRT3001D12.b and EGEQFB1004E07.g encoded proteins show the highest identities (between 73.9%-77.5%) to functionally characterized AtCAD5 and AtCAD4 from *A. thaliana* (Kim *et al.*, 2004). Cluster EGEQRT3001D12.b is orthologous to previously sequenced CAD genes from other *Eucalyptus* species (*E. gunnii* - 99.1% amino acid sequence identity; *E. saligna* - 99.4%; *E. globulus* - 99.4%; *E. botryooides* - 98.8%) (Feuillet *et al.*, 1995; Hibino *et al.*, 1994; Grima-Pettenati *et al.*, 1993; Endt *et al.*, 2000; De Melis *et al.*, 1999). Cluster EGEQFB1004E07.g encodes a more divergent protein, showing between 82.3% to 83.9% identity to the same set

of CAD proteins. Another 3 clusters appeared to be full-length genes: EGBFRT3114B12.g (9 reads, 0 in WD2); EGEQLV1201D07.g (13 reads, 0 in WD2); and EGEQST2002A03.g (20 reads, 0 in WD2). These 3 proteins show below 55% sequence identity to previously described *Eucalyptus* CADs. Cluster EGEQLV1201D07.g encodes a protein with the highest identity level (70.8%), among *Eucalyptus* CADs, to a sinapyl alcohol dehydrogenase (SAD) from *Populus tremuloides* (Li *et al.*, 2001). The encoded enzyme may show preference for sinapaldehyde, making it an S-branch specific enzyme, but the absence of reads from the WD2 library in this cluster argues against this hypothesis. Supporting the latter, Kim *et al.* (2004) did not detect any specific substrate preference among all *A. thaliana* CADs and genes with closest homology to *P. tremuloides* SAD displayed the poorest ability to use any of the 5 aldehyde substrates tested. At least another 13 clusters encoded CAD-like proteins in *Eucalyptus* but were not full-length and since none contained reads from the WD2 library, these are probably not involved in wood lignin biosynthesis.

#### Other enzymes related to lignin biosynthesis and wood formation

**Chitinase** - The cluster with highest number of reads from the WD2 library, EGEQRT3301F08.g, encodes a class I chitinase. Apparently this gene is highly and specifically expressed in wood since among all reads encoding this enzyme (110 reads), 93 were from the WD2 library. Zhong *et al.* (2002) have shown that a mutation in a class I chitinase gene (*elp1*) in *A. thaliana* causes ectopic deposition of lignin and aberrant shapes of cells with incomplete cell walls in the pith of inflorescence stems. The protein encoded by cluster EGEQRT3301F08.g is 70.3% identical to the ELP1 protein. Although plants are not known to produce chitin, it is likely that the substrates or products of the class I chitinases-mediated reaction are necessary for normal plant growth and development (Zhong *et al.*, 2002).

**Laccase** - Laccases can promote polymerization of monolignols in the absence of H<sub>2</sub>O<sub>2</sub>, resulting in either lignans or lignins. Multiple clusters encoding putative laccase proteins were found in *Eucalyptus*, most of them were not full-length. Among the five clusters encoding complete proteins, only EGEQRT3101C03.b (19 reads, 17 in WD2), EGEZRT3005B09.g (15 reads, 9 in WD2), and EGCCRT6011G09.g (11 reads, 1 in WD2) contained reads from the WD2 library. PSORT predicts that the proteins encoded by EGEQRT3101C03.b and EGCCRT6011G09.g are secreted while the one encoded by EGEZRT3005B09.g is localized to the ER membrane. An extracellular localization is in accordance to the proposed role of laccases as polymerization catalysts of monolignols. Interestingly, cluster EGEZRT3005B09.g encodes a laccase protein 67.2% identical to that encoded by the poplar *lac3* gene, which, when silenced, causes alterations in phenolics me-

tabolism and cell wall structure (Ranocha *et al.*, 2002). It is interesting to note that a large number of reads encoding peroxidase enzymes was found in all *Eucalyptus* ESTs libraries except WD2, which had only 3 reads encoding such enzymes. In *Eucalyptus*, apparently, laccases rather than peroxidases play a more important role in lignification. This is in agreement to what has been observed in developing xylem ESTs libraries from poplar (Sterky *et al.*, 1998) and *Pinus taeda* (Allona *et al.*, 1998).

**Dirigent protein** - Dirigent proteins can promote stereoselective coupling of monolignols and their role in the formation of (+)-pinoresinol lignan in *Forsythia* sp and western red cedar (*Thuja plicata*) has been well established (Davin *et al.*, 1997; Gang *et al.*, 1999; Kim *et al.*, 2002). Although their importance for polymerization of lignins remains to be demonstrated, their mode of action may provide clues on how plants control composition and structure of lignins. Five clusters encoding full-length dirigent-like proteins were found. Among these, cluster EGBFST 7259E05.g (7 reads, 5 in WD2) contained the highest number of reads from the WD2 library. The encoded protein has an identity level between 22.5% and 25.5% when compared to previously characterized dirigent proteins from *Forsythia* sp. and western red cedar (Gang *et al.*, 1999; Kim *et al.*, 2002). PSORT indicates that the protein is secreted, in accordance to its proposed role either as a lignan or lignin polymerization agent. Cluster EGBGWD2291B02.g (1 read, 1 in WD2) encodes a dirigent-like protein with the highest identity to those from *Forsythia* and *Thuja*, ranging from 50.5% to 60.7%. It is possible that these dirigent-like proteins are involved in lignans formation only, but even if it is the case, the importance of lignans as wood extractives that interfere with pulping makes investigation of those proteins worthwhile.

#### Metabolic channeling

Colocalization on the ER membrane of many enzymes of the phenylpropanoid pathway, as predicted by PSORT for the *Eucalyptus* enzymes PAL, C4H, 4CL, C3H, and F5H, suggests the existence of multienzyme complexes that could result in channeling of pathway intermediates without their release into the general metabolic pool (Hrazdina and Wagner, 1985). According to Winkel (2004), metabolic channeling can be envisioned as a means to attain high local substrate concentrations, regulate competition between branch pathways for common metabolites, coordinate the activities of pathways with shared enzymes or intermediates, and sequester reactive or toxic intermediates. Channeling of the product of PAL, trans-cinnamic acid, to C4H, the next enzyme on the pathway, has been demonstrated in vivo and in vitro, using tobacco stem tissue and cell suspension culture (Rasmussen and Dixon, 1999). Colocalization of PAL and C4H on the endoplasmic reticulum membrane of tobacco cells has been demonstrated using fluorescence labeled proteins and confocal microscopy

(Achnine *et al.*, 2004). It was also shown that PAL1 isoform has a higher affinity for its membrane localization than PAL2 (Achnine *et al.*, 2004). Association between PAL and C4H might reduce the cinnamate pool within the cell and, thereby, reduce feedback inhibition of PAL (Blount *et al.*, 2000). Although there is evidence that supports the existence of metabolic channeling at the entry point into the phenylpropanoid pathway, Ro and Douglas (2004) could not demonstrate this mechanism in yeast cells expressing poplar PAL and C4H.

Transgenic approaches to improve wood quality, by increasing or reducing expression of specific enzyme isoforms, may affect their subcellular localization and lead to unexpected results. Therefore, detailed studies of enzyme isoforms localization and whether or not they form multi-enzyme complexes are extremely important. The availability of sequences of all genes encoding phenylpropanoid pathway enzymes from *Eucalyptus* will facilitate studies on this subject.

## Conclusions

Clusters encoding enzymes of all steps in lignin biosynthesis could be found in the *Eucalyptus* EST libraries. Comparison of the number of reads of each cluster hints at the expression level of the different genes and in many cases points at the isoform(s) most likely to play a role in wood lignin biosynthesis. Several potential targets for gene silencing aiming reduction and/or modification of lignin through plant transformation were identified. All genes found in this study are good candidates for detailed investigation of single nucleotide polymorphisms (SNPs) and comparison of expression levels between *Eucalyptus* lineages varying in lignin quantity and quality.

## References

- Achnine L, Blancaflor EB, Rasmussen S and Dixon RA (2004) Colocalization of L-phenylalanine ammonia-lyase and cinnamate 4-hydroxylase for metabolic channeling in phenylpropanoid pathway. *Plant Cell* 16:309-3109.
- Allona I, Quinn M, Shoop E, Swope K, St Cyr S, Carlis J, Riedl J, Retzel E, Campbell MM, Sederoff R and Whetten RW (1998) Analysis of xylem formation in pine by cDNA sequencing. *Proc Natl Acad Sci USA* 95:9693-9698.
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W and Lipman DJ (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucl Acids Res* 25:3389-3402.
- Appert C, Logemann E, Hahlbrock K, Schmid J and Armhein N (1994). Structural and catalytic properties of the four phenylalanine ammonia-lyase isoenzymes from parsley (*Petroselinum crispum* Nym.) *Eur J Biochem* 225:491-499.
- Baucher M, Halpin C, Petit-Conil M and Boerjan W (2003) Lignin: Genetic engineering and impact on pulping. *Crit Rev Biochem Mol Biol* 38:305-350.
- Betz C, McCollum TG and Mayer RT (2001) Differential expression of two cinnamate 4-hydroxylase genes in 'Valencia' orange (*Citrus sinensis* Osbeck). *Plant Mol Biol* 46:741-748.
- Blount JW, Korth KL, Masoud SA, Rasmussen S, Lamb C and Dixon RA (2000) Altering expression of cinnamic acid 4-hydroxylase in transgenic plants provides evidence for a feedback loop at the entry point into the phenylpropanoid pathway. *Plant Physiol* 122:107-116.
- Boerjan W, Ralph J and Baucher M (2003) Lignin biosynthesis. *Annu Rev Plant Biol* 54:519-546.
- Cochrane FC, Davin LB and Lewis NG (2004) The *Arabidopsis* phenylalanine ammonia lyase gene family: Kinetic characterization of the four PAL isoforms. *Phytochem* 65:1557-1564.
- Costa MA, Collins RE, Anterola AM, Cochrane FC, Davin LB and Lewis NG (2003) An in silico assessment of gene function and organization of the phenylpropanoid pathway metabolic networks in *Arabidopsis thaliana* and limitations thereof. *Phytochem* 64:1097-1112.
- Davin LB, Wang HB, Crowell AL, Bedgar DL, Martin DM, Sarkanen S and Lewis NG (1997) Stereoselective bimolecular phenoxy radical coupling by an auxiliary (dirigent) protein without an active center. *Science* 275:362-366.
- De Melis LE, Whiteman PH and Stevenson TW (1999) Isolation and characterization of a cDNA clone encoding cinnamyl alcohol dehydrogenase in *Eucalyptus globulus* Labill. *Plant Sci* 143:173-182.
- Endt DV, Costa P, Zago MK, Zanettini MHB and Pasquali G (2000) Genes de lignificação: Em busca do eucalipto transgênico com reduzidos teores de ligninas. *Biotechnol Cienc Desenvol* 15:152-159.
- Feuillet C, Lauvergeat V, Deswarte C, Pilate G, Boudet A and Grima-Pettenati J (1995) Tissue- and cell-specific expression of a cinnamyl alcohol dehydrogenase promoter in transgenic poplar plants. *Plant Mol Biol* 27:651-667.
- Franke R, Humphreys JM, Hemm MR, Denault JW, Ruegger MO, Cusumano JC and Chapple C (2002) The *Arabidopsis* REF8 gene encodes the 3-hydroxylase of phenylpropanoid metabolism. *Plant J* 30:33-45.
- Gang DR, Costa MA, Fujita M, Dinkova-Kostova AT, Wang HB, Burlat V, Martin W, Sarkanen S, Davin LB and Lewis NG (1999) Regiochemical control of monolignol radical coupling: A new paradigm for lignin and lignan biosynthesis. *Chem Biol* 6:143-151.
- Grima-Pettenati J, Feuillet C, Goffner D, Borderies G and Boudet AM (1993) Molecular cloning and expression of a *Eucalyptus gunnii* cDNA clone encoding cinnamyl alcohol dehydrogenase. *Plant Mol Biol* 21:1085-1095.
- Hall TA (1999) BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* 41:95-98.
- Hibino T, Chen JQ, Shibata D and Higuchi T (1994) Nucleotide sequence of a *Eucalyptus botryoides* gene encoding cinnamyl alcohol dehydrogenase. *Plant Physiol* 104:305-306.
- Hoffmann L, Maury S, Martz F, Geoffroy P and Legrand M (2003) Purification, cloning, and properties of an acyltransferase controlling shikimate and quinate ester intermediates in phenylpropanoid metabolism. *J Biol Chem* 278:95-103.
- Hrazdina G and Wagner G (1985) Metabolic pathways as enzyme complexes: Evidence for the synthesis of phenylpropanoids and flavonoids on membrane-associated enzyme complexes. *Arch Biochem Biophys* 237:88-100.
- Humphreys JM, Hemm MR and Chapple C (1999) New routes for lignin biosynthesis defined by biochemical characterization of recombinant ferulate 5-hydroxylase, a multifunctional

- cytochrome P450-dependent monooxygenase. *Proc Natl Acad Sci USA* 96:10045-10050.
- Humphreys JM and Chapple C (2002) Rewriting the lignin roadmap. *Curr Opin Plant Biol* 5:224-229.
- Jones L, Ennos AR and Turner SR (2001) Cloning and characterization of irregular xylem4 (irx4): A severely lignin-deficient mutant of *Arabidopsis*. *Plant J* 26:205-216.
- Kao YY, Harding SA and Tsai CJ (2002) Differential expression of two distinct phenylalanine ammonia-lyase genes in condensed tannin-accumulating and lignifying cells of quaking aspen. *Plant Physiol* 130:796-807.
- Kim MK, Jeon JH, Fujita M, Davin LB and Lewis NG (2002) The western red cedar (*Thuja plicata*) 8-8' dirigent family displays diverse expression patterns and conserved monolignol coupling specificity. *Plant Mol Biol* 49:199-214.
- Kim SJ, Kim MR, Bedgar DL, Moinuddin SG, Cardenas CL, Davin LB, Kang C and Lewis NG (2004) Functional reclassification of the putative cinnamyl alcohol dehydrogenase multigene family in *Arabidopsis*. *Proc Natl Acad Sci USA* 101:1455-1460.
- Kumar A and Ellis BE (2001) The phenylalanine ammonia-lyase gene family in raspberry. Structure, expression, and evolution. *Plant Physiol* 127:230-239.
- Lacombe E, Hawkins S, Van Doorselaere J, Piquemal J, Goffner D, Poeydomenge O, Boudet AM and Grima-Pettenati J (1997) Cinnamoyl CoA reductase, the first committed enzyme of the lignin branch biosynthetic pathway: Cloning, expression and phylogenetic relationships. *Plant J* 11:429-441.
- Li L, Cheng XF, Leshkevich J, Umezawa T, Harding SA and Chiang VL (2001) The last step of syringyl monolignol biosynthesis in angiosperms is regulated by a novel gene encoding sinapyl alcohol dehydrogenase. *Plant Cell* 13:1567-1586.
- Li L, Zhou Y, Cheng X, Sun J, Marita JM, Ralph J and Chiang VL (2003) Combinatorial modification of multiple lignin traits in trees through multigene cotransformation. *Proc Natl Acad Sci USA* 100:4939-4944.
- Meyermans H, Morreel K, Lapierre C, Pollet B, De Bruyn A, Busson R, Herdewijn P, Devreese B, Van Beeumen J, Marita JM, Ralph J, Chen C, Burggraeve B, Van Montagu M, Messens E and Boerjan W (2000) Modifications in lignin and accumulation of phenolic glucosides in poplar xylem upon down-regulation of caffeoyl-coenzyme A *O*-methyltransferase, an enzyme involved in lignin biosynthesis. *J Biol Chem* 275:36899-36909.
- Nair RB, Xia Q, Kartha CJ, Kurylo E, Hirji RN, Datla R and Selvaraj G (2002) *Arabidopsis* CYP98A3 mediating aromatic 3-hydroxylation. Developmental regulation of the gene, and expression in yeast. *Plant Physiol* 130:210-220.
- Nakai K and Kanehisa M (1992) A knowledge base for predicting protein localization sites in eukaryotic cells. *Genomics* 14:897-911.
- Nedelkina S, Jupe SC, Blee KA, Schalk M, Werck-Reichhart D and Bolwell GP (1999) Novel characteristics and regulation of a divergent cinnamate 4-hydroxylase (CYP73A15) from French bean: engineering expression in yeast. *Plant Mol Biol* 39:1079-1090.
- Osakabe Y, Osakabe K, Kawai S, Katayama Y and Morohoshi N (1995) Characterization of the structure and determination of mRNA levels of the phenylalanine ammonia-lyase gene family from *Populus kitakamiensis*. *Plant Mol Biol* 28:1133-1141.
- Osakabe K, Tsao CC, Li L, Popko JL, Umezawa T, Carraway DT, Smeltzer RH, Joshi CP and Chiang VL (1999) Coniferyl aldehyde 5-hydroxylation and methylation direct syringyl lignin biosynthesis in angiosperms. *Proc Natl Acad Sci USA* 96:8955-8960.
- Parvathi K, Chen F, Guo D, Blount JW and Dixon RA (2001) Substrate preferences of *O*-methyltransferases in alfalfa suggest new pathways for 3-*O*-methylation of monolignols. *Plant J* 25:193-202.
- Pilate G, Guiney E, Holt K, Petit-Conil M, Lapierre C, Leple JC, Pollet B, Mila I, Webster EA, Marstorp HG, Hopkins DW, Jouanin L, Boerjan W, Schuch W, Cornu D and Halpin C (2002) Field and pulping performances of transgenic trees with altered lignification. *Nat Biotechnol* 20:607-612.
- Poeydomenge O, Boudet AM and Grima-Pettenati J (1994) A cDNA encoding S-adenosyl-L-methionine:caffeic acid 3-*O*-methyltransferase from *Eucalyptus*. *Plant Physiol* 105:749-750.
- Raes J, Rohde A, Christensen JH, Van de Peer Y and Boerjan W (2003) Genome-wide characterization of the lignification toolbox in *Arabidopsis*. *Plant Physiol* 133:1051-1071.
- Ranocha P, Chabannes M, Chamayou S, Danoun S, Jauneau A, Boudet AM and Goffner D (2002) Laccase down-regulation causes alterations in phenolic metabolism and cell wall structure in poplar. *Plant Physiol* 129:145-155.
- Rasmussen S and Dixon RA (1999) Transgene-mediated and elicitor-induced perturbation of metabolic channeling at the entry point into the phenylpropanoid pathway. *Plant Cell* 11:1537-1551.
- Ro DK and Douglas CJ (2004) Reconstitution of the entry point of plant phenylpropanoid metabolism in yeast (*Saccharomyces cerevisiae*): Implications for control of metabolic flux into the phenylpropanoid pathway. *J Biol Chem* 279:2600-2607.
- Ro DK, Mah N, Ellis BE and Douglas CJ (2001) Functional characterization and subcellular localization of poplar (*Populus trichocarpa* x *Populus deltoides*) cinnamate 4-hydroxylase. *Plant Physiol* 126:317-329.
- Schoch G, Goepfert S, Morant M, Hehn A, Meyer D, Ullmann P and Werck-Reichhart D (2001) CYP98A3 from *Arabidopsis thaliana* is a 3'-hydroxylase of phenolic esters, a missing link in the phenylpropanoid pathway. *J Biol Chem* 276:36566-36574.
- Sterky F, Regan S, Karlsson J, Hertzberg M, Rohde A, Holmberg A, Amini B, Bhalerao R, Larsson M, Villarroel R, Van Montagu M, Sandberg G, Olsson O, Teeri TT, Boerjan W, Gustafsson P, Uhlén M, Sundberg B and Lundeberg J (1998) Gene discovery in the wood-forming tissues of poplar: Analysis of 5,692 expressed sequence tags. *Proc Natl Acad Sci USA* 95:13330-13335.
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F and Higgins DG (1997) The ClustalX windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucl Acids Res* 25:4876-4882.
- Walter MH, Grima-Pettenati J, Grand C, Boudet AM and Lamb CJ (1988) Cinnamyl-alcohol dehydrogenase, a molecular marker specific for lignin synthesis: cDNA cloning and mRNA induction by fungal elicitor. *Proc Natl Acad Sci USA* 85:5546-5550.
- Winkel BJS (2004) Metabolic channeling in plants. *Annu Rev Plant Biol* 55:85-107.
- Zhong R, Kays SJ, Schroeder BP and Ye ZH (2002) Mutation of a chitinase-like gene causes ectopic deposition of lignin, aberrant cell shapes, and overproduction of ethylene. *Plant Cell* 14:165-179.