Molecular basis for the evolution of xylem lignification
Gary Peter¹ and David Neale²

The lignification of xylem is an adaptive trait of great significance. Gymnosperms and angiosperms share an ancient, conserved set of enzymes that are regulated by a conserved transcription factor and that are responsible for the formation of guaiacyl lignin. Angiosperms have evolved at least two enzymes that catalyze the production of syringyl lignin. Association genetics is now being used to explore the adaptive significance of sequence variation in the genes that encode these monolignol biosynthetic enzymes.

Addresses
¹ School of Forest Resources and Conservation, Genetics Institute, PO Box 1, 10410 University of Florida, Gainesville, Florida 32611-0410, USA
e-mail: gfpeter@ufl.edu
² Institute of Forest Genetics, Pacific Southwest Research Station, US Department of Agriculture Forest Service, Department of Environmental Horticulture, University of California Davis, Davis, California 95616, USA
e-mail: dbneale@ucdavis.edu

Corresponding author: Gary Peter

Abbreviations
CAD (hydroxycinnamyl alcohol dehydrogenase)
CCoAOMT (caffeoyl CoA O-methyltransferase)
CCR (hydroxy)caffeoyl CoA reductase
C3H (p-coumaroyl shikimate/quinate 3-hydroxylase
4CL (cinnamate 4-hydroxylase)
COMT (caffeic acid/5-hydroxyferulic acid O-methyltransferase)
F5H/Cald5H (ferulate 5-hydroxylase/coniferylaldehyde 5-hydroxylase)
G lignin (guaiacyl lignin)
HCT/CST (hydroxycinnamoyl CoA:shikimate/quinate hydroxycinnamoyltransferase)
PAL (phenylalanine ammonia-lyase)
S lignin (sinapyl lignin)
SNP (single nucleotide polymorphism)

Introduction
Lignin is a complex three-dimensional, branched heteropolymer of bonded hydroxycinnamyl alcohols that crosslink carbohydrate polymers together, thereby rigidifying plant cell walls [1–3]. The importance of lignin ranges from its fundamental roles in the evolution of land plants, global carbon cycling, and plant growth and development, and its role in the biotic and abiotic stress resistance of plants, to the practical importance of lignin in agriculture and the utilization of plant materials. The adaptive significance of the lignification of xylem cells during the evolution of land plants [4] is evidenced by the proliferation of land plants, such that the mass of lignin in the biosphere is second only to the mass of cellulose. The synthesis of lignin is one of the most intensively studied metabolic pathways in plants, with extensive analyses having been carried out in herbaceous plants, woody gymnosperms and angiosperms [1–3]. Here, we review the molecular basis for the evolution of lignification.

In plants, phenylpropanoid metabolism leads to a diverse array of important compounds, including flavonoids/isoflavonoids, phytoalexins, and lignin, all of which were important for the colonization of land by plants [4]. Lignin biosynthesis occurs in three stages: biosynthesis of the hydroxycinnamyl alcohols or monolignols, transport of the monolignols out of the cytoplasm, and polymerization of the monolignols to form lignin in the cell wall [1–3]. The transport of monolignols from the cytoplasm into the wall and their polymerization into lignin are still poorly understood. Thus, our knowledge about lignin biosynthesis comes principally from the discovery and characterization of the enzymes that are responsible for synthesis of the monolignols, namely p-coumaryl, coniferyl, and sinapyl alcohols. Until recently, this pathway was most often depicted as a complete matrix of reactions [1–3]. This complexity is related to the role that the phenylpropanoid pathway plays in secondary metabolism, and to the observation that enzymes in the monolignol pathway can utilize multiple substrates. However, recent detailed analyses of enzyme function now show that some of these reactions do not occur in lignifying xylem [1,3]. Figure 1 depicts the prominent monolignol biosynthetic pathway in xylem tissues, and several recent reviews thoroughly summarize the genetic and biochemical evidence for this view [1,3,5]. This molecular evidence shows that an ancient, predominant pathway for the synthesis of coniferyl alcohol is conserved among land plants and that a pathway for sinapyl alcohol synthesis evolved in angiosperm species [6–9].

Evolution of the lignification of xylem cells
The function of lignin in conductive xylem cells is twofold: the hydrophobic nature of lignin makes the secondary walls impermeable to water, thereby facilitating hydraulic conductance; and the lignin crosslinks rigidify the thickened wall, allowing it to resist collapse from the high negative hydrostatic pressures in xylem [10]. The crosslinking of cellulose in the xylem secondary wall is also crucial for the mechanical support of the whole plant.
In gymnosperms, xylem tracheids function in both mechanical support and water transport [11]. Tracheid walls are rich in guaiacyl lignin (G lignin) and do not contain syringyl (S) lignin, apparently because they lack the enzymes for sinapyl alcohol synthesis [7,9,12]. By contrast, the secondary xylem of angiosperm trees contains two more specialized cell types: the vessel elements, which conduct water, and the fiber cells, which provide mechanical support [11]. Separation of these functions provides a more efficient and economical architecture for the secondary xylem. This efficiency is evident in the fewer but larger vessel elements and the reduced lignin in the woody stems of angiosperm (20% of dry matter) compared with gymnosperm (30%) trees. A lower lignin and higher carbohydrate content requires significantly less energy and carbon for growth [13]. Interestingly, in angiosperms, the water-conductive xylem vessel elements of secondary xylem and the primary xylem cells are rich in or contain only G lignin, like the tracheids of the more ancient gymnosperms, whereas the nonconductive xylem fiber cells are rich in S lignin [7,14]. The fact that water-conducting cells in both gymnosperms and angiosperms are principally comprised of G lignin suggests a strong selective pressure to conserve the pathway for and the regulation of G lignin biosynthesis in the water conducting cells of xylem during land plant evolution.

The predominant pathway for monolignol biosynthesis in xylem cells is outlined in black, with the dark arrows showing the primary substrates and products and the gray arrows showing the minor substrates and products. The blue shading indicates the pathway that is conserved between angiosperms and gymnosperms, whereas the green shading indicates the angiosperm-specific pathway. The enzymes and their abbreviations are as follows: CAD, (hydroxy)cinnamyl alcohol dehydrogenase; CCoAOMT, caffeoyl CoA O-methyltransferase; CCR, (hydroxy)cinnamoyl CoA reductase; C3H, p-coumaroyl shikimate/quinate 3-hydroxylase; C4H, cinnamate 4-hydroxylase; 4CL, 4-coumarate CoA ligase; COMT, caffeic acid/5-hydroxyferulic acid O-methyltransferase; F5H/Cald5H, ferulate 5-hydroxylase/coniferylaldehyde 5-hydroxylase; PAL, phenylalanine ammonia-lyase; SAD, sinapyl alcohol dehydrogenase.

Evolution of monolignol biosynthesis in xylem

Regulation of monolignol biosynthetic genes

The molecular basis of the fiber-cell-specific expression of F5H/Cald5H, COMT, and SAD, which leads to the
The biosynthesis of S lignin, which is unknown. Functional analysis of promoter regions in phenylalanine ammonia-lyase (PAL) genes first identified AC elements as important cis-regulatory sequences that are required for the proper expression of lignin biosynthesis genes in lignifying xylem cells (e.g. [16]). A recent bioinformatic analysis identified AC cis-elements in ten of the thirty-four Arabidopsis genes that are potentially involved in monolignol biosynthesis (i.e. PAL1, PAL2, 4-coumarate CoA ligase [4CL]1, 4CL2, hydroxycinnamoyl CoA:shikimate/quinate hydroxycinnamoyl transferase [HCT/CST], p-coumaroyl shikimate/quinate 3-hydroxylase [C3H], caffeoyl CoA O-methyltransferase [CCoAOMT], (hydroxy)cinnamyl alcohol dehydrogenase [CAD]5 and CAD6) [8]. Interestingly, these ten genes are highly expressed in tissues undergoing xylem lignification, cluster with gymnosperm homologues and represent all of the genes, except for cinnamate 4-hydroxylase [CAD]5 and CAD6) [8]. By contrast, AC cis-elements were not found in the F5H/Cald5H and COMT genes, which are involved in the biosynthesis of S lignin in the fiber cells of angiosperms [8]. This analysis suggests a conserved, possibly ancient, regulatory network for G lignin biosynthesis in xylem cells [8]. Support for such a conserved ancient pathway comes from the isolation and characterization of PtMYB4 from the xylem of the gymnosperm Pinus taeda. PtMYB4 binds to AC cis-elements in vitro and is a positive activator of transcription in yeast and transgenic tobacco, in which its misexpression is sufficient for ectopic lignin formation [17].

Conservation and divergence of monolignol biosynthetic enzymes

A conserved regulatory network of monolignol genes that are involved in the synthesis of G lignin in xylem suggests that these genes were highly conserved during evolution, and that homologues for each of the genes should be readily identifiable through molecular phylogenetic approaches [8]. Such an identification is complicated because the genes involved in monolignol biosynthesis are highly conserved and that homologues for each of the genes should be readily identifiable through molecular phylogenetic approaches [8]. Such an identification is complicated because the genes involved in monolignol biosynthesis are most often part of multigene families (Table 1) whose members are also involved in other aspects of phenylpropanoid metabolism, such as the biosynthesis of sinapate esters and flavonoid [1,18]. Thus, identifying individual members of gene families that are involved in monolignol biosynthesis is highly conserved during evolution, and that homologues for each of the genes should be readily identifiable through molecular phylogenetic approaches [8]. Such an identification is complicated because the genes involved in monolignol biosynthesis are highly conserved and that homologues for each of the genes should be readily identifiable through molecular phylogenetic approaches [8]. Such an identification is complicated because the genes involved in monolignol biosynthesis are highly conserved and that homologues for each of the genes should be readily identifiable through molecular phylogenetic approaches [8]. Such an identification is complicated because the genes involved in monolignol biosynthesis are highly conserved and that homologues for each of the genes should be readily identifiable through molecular phylogenetic approaches [8]. Such an identification is complicated because the genes involved in monolignol biosynthesis are highly conserved.
Table 2

<table>
<thead>
<tr>
<th>Gene</th>
<th>Decreased expression</th>
<th>Increased expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAL</td>
<td>10–70% reduction in lignin level; G units reduced to a greater extent than S units</td>
<td>Small increase in lignin content; slight decrease in S units</td>
</tr>
<tr>
<td>4CL</td>
<td>40–60% reduction in S and G units</td>
<td>No effect</td>
</tr>
<tr>
<td>C4H</td>
<td>30–60% reduction in S and G units</td>
<td></td>
</tr>
<tr>
<td>C3H</td>
<td>40–80% reduction in S and G units; mostly H lignin remaining</td>
<td></td>
</tr>
<tr>
<td>CCoAOMT</td>
<td>20–45% reduction in S and G units; mainly G units reduced</td>
<td></td>
</tr>
<tr>
<td>CCR</td>
<td>15–50% reduction in S and G units; mainly G units reduced</td>
<td></td>
</tr>
<tr>
<td>F5H/Cald5H</td>
<td>0–10% reduction in S lignin units</td>
<td>Increased S/G ratio</td>
</tr>
<tr>
<td>COMT</td>
<td>0–20% reduction in lignin levels; S units reduced; elevated levels of 5-hydroxyconiferyl alcohol (5OHG)</td>
<td></td>
</tr>
<tr>
<td>CAD</td>
<td>Altered lignin composition</td>
<td></td>
</tr>
<tr>
<td>4CL and F5H/Cald5H</td>
<td>Decreased expression of 4CL results in the reduction of lignin contents by 50%</td>
<td>Increased expression of F5H/Cald5H increases S/G ratio</td>
</tr>
</tbody>
</table>

* Condensed summary from [2,5,15].

encode suggest that this gene family has undergone both ancient duplications before the split of the angiosperm and gymnosperm lineages and more recent gene duplications in eudicots [8,18,20–22]. There are a large number of 4CL genes in plants because of the involvement of these genes in numerous biosynthetic pathways [20], and functional evidence shows that class I 4CL enzymes are involved in the biosynthesis of phenylpropanoids and monolignols (Table 1; [8,20,23]). HCT/CST is an acyl transferase that was identified in tobacco only very recently and that is encoded by a single gene in Arabidopsis [8,24]. HCT/CST has not yet been characterized in other plants, but is predicted to be involved in monolignol biosynthesis in gymnosperms. CCR is highly conserved from gymnosperms to angiosperms, and is considered to be the first enzyme that is specific to the monolignol pathway (Figure 1; [1,8]).

The COMT and CCoAOMT gene families share a very limited sequence similarity, but both are part of the 5'-adenosyl methionine-dependent methyl transferase superfamily. The CCoAOMT genes are highly conserved in gymnosperms and angiosperms (Figure 1; Table 1; [8,18,22]). The COMT and COMT-like genes form a diverse set of methyltransferases in plants [8]. Functional evidence from Arabidopsis shows that one COMT gene is involved in S lignin formation [22] and supports the prediction that COMT genes are specific to the S lignin pathway [6]. Interestingly, genes that are very similar in sequence to angiosperm COMTs have been isolated from three different pine species; however, we have no biochemical or genetic data to describe their role in lignin biosynthesis. A highly diverged multifunctional O-methyltransferase (hydroxycinnamic acids/hydroxycinnamoyl CoA esters o-methyltransferase [AEOMT]) that utilizes common COMT and CCoAOMT substrates has also been isolated from the xylem of P. taeda [25].

CAD is a member of the alcohol dehydrogenase superfamily and has been identified in many plant species. In Arabidopsis there are nine CAD genes [8,18,26,27]. The Arabidopsis class I CADs are most similar in sequence to the CADs that are involved in coniferyl alcohol synthesis in other angiosperm and in gymnosperm species (Table 1). Class II CADs from Arabidopsis are more similar to alcohol dehydrogenases that have diverse substrate preferences, including the aspen CAD [8]. Aspen CAD is a diverged alcohol dehydrogenase expressed in lignifying xylem fibers that shares limited sequence identity (53%) with aspen CAD, and preferentially converts sinapyl alcohol to sinapyl alcohol in vitro [7].

Naturally occurring variation in monolignol pathway genes and association with phenotype in trees

Knowledge of the function and developmental significance of genes in the monolignol biosynthesis pathway has largely been gained in a few well-studied plant species through forward-genetic isolation or reverse-genetic creation of null or nearly null mutant lines with large phenotypic effects (Table 2; [1,2,6]). These approaches do not, however, tell us about how these genes evolved or about the natural variation that exists within these genes, or whether this variation is adaptively significant. Association and population-genetic analyses of monolignol genes can now be used to analyze the natural variation of alleles whose selective differences lead to adaptive differences among populations [28]. For example, association genetics not only identifies the functional role of a genetic locus, but can also provide estimates of the effects of individual alleles at that locus on the phenotype. One of the first examples of the success of this natural mutant approach in trees was the discovery of a null allele of the CAD gene in P. taeda and the association of abnormal lignin with the presence of this null allele [29,30]. Recently, the molecular basis of this CAD null allele was identified as a two-base AA insertion.
in the fifth exon, which shifts the reading frame to one that has a premature stop codon [31].

A genomic approach to allele discovery and identifying relationships with desired lignin phenotypes is now being used on a large scale in *P. taeda* [http://dendrome.ucdavis.edu/adept](http://dendrome.ucdavis.edu/adept). This association genetics approach relies on linkage disequilibrium between single nucleotide polymorphisms (SNPs) and the mutations that cause the phenotypes [32–34]. SNP discovery and estimation of linkage disequilibrium has recently been completed in *P. taeda* for nineteen candidate genes for wood formation, including most of the genes in the monolignol pathway (PAL, C4H-1, C4H-2, 4CL, C3H-2, CCOAMT, COMT-2, CCR, and CAD [G Brown et al., unpublished]). High levels of haplotype diversity were found in these genes, together with linkage disequilibrium decays in the order of the physical length of a gene (approximately 2500 bp). These findings suggest that associations between SNPs found in candidate genes and phenotypes can be used to imply that a given candidate gene causes a certain phenotype. Although the exact causative basis of a phenotype cannot be precisely determined without further study, the advantage that this approach offers is in the large number of genes that can be studied simultaneously and the lack of a requirement for transformation. The association genetics approach should yield dozens, if not hundreds, of alleles that have modest but desirable effects on lignin and other wood properties.

Conclusions

Molecular evolution and comparative functional studies support the hypothesis that a core set of genes that are involved in monolignol biosynthesis in xylem, leading to the formation of G lignin, have been conserved during land plant evolution; and that genes leading to the biosynthesis of S lignin in the xylem fiber cells of angiosperms evolved more recently. This divergence occurred within the basal angiosperms before the separation of the grasses from the eudicots.

Acknowledgements

The authors’ research is supported by the National Science Foundation (grant no. 9975806) and the US Department of Agriculture (USDA) Initiative for Future Agriculture and Food Systems (IFAFS) (grant no. 2001-32100-11313), Florida Agricultural Experiment Station Journal Series R-10272.

References


