Distinct Roles of Cinnamate 4-hydroxylase Genes in *Populus*

Shanfa Lu *, Yihua Zhou 1, Laigeng Li and Vincent L. Chiang

Forest Biotechnology Group, Department of Forestry and Environmental Resources, College of Natural Resources, North Carolina State University, Raleigh, NC 27695-7247, USA

Cinnamate 4-hydroxylase (C4H) catalyzes the conversion of cinnamate into 4-hydroxy-cinnamate, a key reaction of the phenylpropanoid pathway which leads to the biosynthesis of several secondary metabolites. C4H genes exist as a multigene family in various plant species. In order to understand the roles of individual C4H members, four C4H cDNAs (PtrC4H) were isolated from *Populus tremuloides* and three C4H loci (PtrC4H) were identified in the *P. trichocarpa* genome. The ability of *Populus* C4H isoforms to convert trans-cinnamate into p-coumaric acid was verified by the examination of yeast recombinant PtreC4H proteins. *Populus* C4H genes were expressed in various tissues, including developing xylem, phloem and epidermis; however, the expression patterns of individual members were different from each other. Sequential analysis of C4H promoters showed that the differential expression of C4H genes was associated with cis-acting regulatory elements such as box L, box P and H box, suggesting that the divergent C4H isoforms played distinct roles in the production of secondary metabolites. The involvement of specific C4H isoforms in the biosynthesis of guaiacyl and syringyl monolignols is discussed.

**Keywords:** Cinnamate 4-hydroxylase (C4H) — Lignin — Phenylpropanoid — *Populus tremuloides* — *Populus trichocarpa*.

Abbreviations: C4H, cinnamate 4-hydroxylase; 4CL, 4-coumarate coenzyme A:ligase; CPR, NADPH-cytochrome P450 reductase; Ct, threshold cycle; CTAB, cetyltrimethylammonium bromide; DIG, digoxigenin; G, guaiacyl; MSD, mass spectrometer detector; ORF, open reading frame; PAL, phenylalanine ammonia-lyase; RT–PCR, reverse transcription–PCR; S, syringyl; UTR, untranslated region.

The nucleotide sequences reported in this paper have been submitted to the GenBank database under accession numbers DQ522292 (PtrC4H1-1), DQ522293 (PtrC4H1-2), DQ522294 (PtrC4H2-1), DQ522295 (PtrC4H2-2), DQ522296 (PtrC4H1p1), DQ522297 (PtrC4H1p2) and DQ522298 (PtrC4H2p).

Introduction

The core reactions of the phenylpropanoid metabolism that produces important secondary metabolites participating in plant development and defense responses (Hahlbrock and Scheel 1989) involve three enzymes, phenylalanine ammonia-lyase (PAL; EC 4.3.1.5), cinnamate 4-hydroxylase (C4H, EC 1.14.13.11) and 4-coumarate coenzyme A ligase (4CL; EC 6.2.1.12) (Mizutani et al. 1997). PAL catalyzes the deamination of phenylalanine to produce trans-cinnamic acid, which is converted to p-coumaric acid by an oxidative reaction catalyzed by a cytochrome P450 enzyme, C4H. A thioesterification is then activated by 4CL and the product, p-coumaroyl-CoA, is funneled into one of the branched pathways leading to the production of cell wall constituents (lignins), pigments (flavonoids), UV protectants (coumarins) and plant defense compounds (isoflavonoids, furanocoumarins and norlignans) (Whitbred and Schuler 2000, Suzuki et al. 2002, Suzuki et al. 2004).

Genes encoding PAL, C4H and 4CL have been cloned from various plant species (for a recent review, see Li et al. 2006). *PAL* genes present as a small gene family in plants, such as Arabidopsis (Raes et al. 2003) and raspberry (Kumar and Ellis 2001). Divergent *PAL* genes, exhibiting tissue-specific expression in plants, have been suggested to function in specific metabolic milieus (Kao et al. 2002). For example, one *PAL* gene isolated from *Populus tremuloides* is associated with condensed tannin metabolism, whereas another is involved in monolignol biosynthesis (Kao et al. 2002). Similar to *PAL*, multiple 4CL gene members have been found in Arabidopsis (Raes et al. 2003), quaking aspen (Hu et al. 1998), tobacco (Lee and Douglas 1996), *P. trichocarpa* (http://genome.jgi-psf.org/Potr1/Potr1.home.html), and others. The divergent 4CL genes have also been found to play a role in the biosynthesis of specific secondary metabolites. In Arabidopsis, *At4CL1* and *At4CL2* function in monolignol biosynthesis (Ehlting et al. 1999, Raes et al. 2003), *At4CL3* activates the biosynthesis of flavonoid (Ehlting et al. 1999), whereas *At4CL4* is most probably related to sinapate activation (Hamberger and Hahlbrock 2004). Additionally, two 4CL genes isolated from quaking aspen display distinct expression patterns. One presents primarily in developing xylem, while the other

1 Present address: Institute of Genetics, Chinese Academy of Sciences, Beijing 100101, PR China.

*Corresponding author: E-mail, slu@unity.ncsu.edu; Fax, +1-919-515-7801.
is expressed in epidermal and leaf tissues (Hu et al. 1998). The results from PAL and 4CL studies indicate that C4H genes could also exist as a multigene family in various plant species and the divergent isoforms could play distinct roles in secondary metabolisms.

So far, the C4H gene has been studied in several plant species. In Arabidopsis, only one C4H gene (AtC4H) has been described (Raes et al. 2003). AtC4H is expressed in all tissues and responds to light, wounding and fungal infection (Bell-Leong et al. 1997, Mizutani et al. 1997, Nair et al. 2002, Raes et al. 2003), indicating that it plays diverse functions in phenylpropanoid metabolism. Conversely, C4Hs are encoded by a small gene family in alfalfa (Fahrendorf and Dixon 1993), maize (Potter et al. 1995), orange (Betz et al. 2001), pea (Whitbred and Schuler 2000), periwinkle (Hotze et al. 1995) and hybrid aspen (Kawai et al. 1996). However, little is known about the roles of individual C4H isoforms in the production of secondary metabolites (Betz et al. 2001). Here we report the characterization of four C4Hs in P. tremuloides and three in P. trichocarpa. We verified their biochemical functions using yeast recombinant proteins and analyzed their differential expression by in situ hybridization, reverse transcription–PCR (RT–PCR) and real-time PCR. Interestingly, we found that the differential expression of C4H genes is associated with cis-acting regulatory elements, suggesting that the divergent C4H isoforms play distinct roles in phenylpropanoid metabolism in Populus.

Results

Isolation of PtreC4H cDNAs in aspen developing xylem

In order to isolate multiple C4H cDNAs, we screened an aspen developing xylem cDNA library under low stringency using 32P-labeled aspen C4H cDNA (Ge and Chiang 1996) as a probe. A total of 30 positive clones were identified. Sequence analysis revealed four different C4H cDNA clones, which were named PtreC4H1-1 (accession No. DQ522292), PtreC4H1-2 (accession No. DQ522293), PtreC4H2-1 (accession No. DQ522294) and PtreC4H2-2 (accession No. DQ522295), respectively. Each of the four PtreC4H genes has an open reading frame (ORF) of 1,515 bp coding for a 505 amino acid protein, which contains an N-terminal hydrophobic domain consisting of 21 amino acids, four conserved domains identified in eukaryotic P450s and a heme-binding motif (Kalb and Loper 1988). The putative PtreC4H proteins have a range of 85–99% amino acid sequence identities to the C4Hs in Arabidopsis (Bell-Leong et al. 1997, Mizutani et al. 1997), Helianthus tuberosus (Teutsch et al. 1993), mung bean (Mizutani et al. 1993), alfalfa (Fahrendorf and Dixon 1993), Catharanthus roseus (Hotze et al. 1995), parsley (Logemann et al. 1995), hybrid poplar (Ro et al. 2001) and Populus kitakamiensis (Kawai et al. 1996).

The four PtreC4H cDNAs could be divided into two groups according to sequence comparison. PtreC4H1-1 and PtreC4H1-2 showing 99% sequence identity were grouped and named PtreC4H1 genes. Similarly, PtreC4H2-1 and PtreC4H2-2 were grouped and named PtreC4H2 genes. The homology between PtreC4H1 and PtreC4H2 genes is 91% at the nucleotide level and 97% in their deduced amino acid sequences in the coding region. However, the difference is substantial in the 5′- and 3′-untranslated regions (UTRs). In these regions, approximately 60–70% sequence homology was found between PtreC4H1 and PtreC4H2 genes. The divergent sequences in the non-coding regions indicated that the PtreC4H1 and PtreC4H2 genes could be regulated differently in various tissues.

Biochemical characterization of PtreC4H proteins

Co-expression of four highly similar PtreC4H genes in xylem made us wonder whether they are capable of converting trans-cinnamate into p-coumaric acid, since the substrate specificity of PtreC4H, as a P450, could be changed by a single amino acid substitution (Lindberg and Negishi 1989). To address this question, we analyzed the biochemical function of PtreC4H proteins. Protein-coding regions of PtreC4H1-1 and PtreC4H2-1 were cloned into a yeast expression vector pYES2.1 under the control of a galactose-inducible GAL1 promoter and expressed in a Saccharomyces cerevisiae strain INVSc2(CPR) which carried the Arabidopsis NADPH-P450 reductase gene and was previously constructed in our laboratory (Osakabe et al. 1999). A PtreC4H1-1 and PtreC4H2-1 activity assay was conducted by a modified in vivo method (Schoendorf et al. 2001, Shimada et al. 2001). Induced culture of the yeast transformants which carried PtreC4H1-1, PtreC4H2-1 or control vector pYES2.1 was incubated with the C4H substrate, trans-cinnamate. The ether extracts resulting from these incubations were analyzed by HPLC-UV/mass spectrometer detector (MSD) (Osakabe et al. 1999). As shown in Fig. 1, the substrate trans-cinnamate was converted into p-coumaric acid in yeast strains carrying PtreC4H1-1 or PtreC4H2-1. There was no such conversion in the yeast strain transformed with control vector, pYES2.1. From this result, we concluded that both PtreC4H1-1 and PtreC4H2-1 encode active C4Hs.

In situ expression analysis of PtreC4H genes

To investigate the expression profile of C4H genes in P. tremuloides, in situ hybridization was performed using digoxigenin (DIG)-labeled sense and antisense RNA probes corresponding to 0.7 kb of the 5′ end of PtreC4H1-1.
As *PtreC4H* genes shared a high level of homology, the hybridization signals detected by the *PtreC4H1-1* probe might reflect the collective expression of the four *PtreC4H* genes. As shown in Fig. 2, *PtreC4H* transcripts were not detected in young stems (the fourth internode, Fig. 2A) and young leaves (the third and sixth leaves, Fig. 2G, H) of *P. tremuloides* plants, suggesting that the expressional levels of *C4H* genes in these tissues are low; however, in the older tissues (Fig. 2B–D, F, I), the transcripts are abundant in developing xylem, phloem and epidermis tissues, suggesting that *C4H* isoforms have a broad range of physiological functions in plants.

Quantitative real-time PCR analysis of *PtreC4H* expression

Because of the high level of sequence homology among *PtreC4H* cDNAs, it is difficult to distinguish the transcripts of individual *PtreC4H* genes by in situ hybridization. To solve this problem, gene-specific RT–PCR was conducted using a gene-specific forward primer located in the 5' UTR and a reverse primer located in the coding region. The results showed that the expression of the *PtreC4H1* genes was the most abundant in developing xylem, followed by phloem and mature leaf tissues; however, *PtreC4H2* genes were expressed strongly in developing xylem, more weakly in phloem, and were undetectable in mature leaf tissues (Fig. 3).

To analyze quantitatively the differential expression of divergent *PtreC4H* genes further, real-time PCR using gene-specific primers located in the coding region was carried out. As shown in Fig. 4, the fold changes of expression of the *PtreC4H2* genes in xylem compared with those in young leaves or phloem were apparently higher than those of *PtreC4H1* genes, although both *PtreC4H1* and *PtreC4H2* genes were highly expressed in xylem, suggesting that *PtreC4H2* genes were more xylem specific.

Identification and real-time PCR analysis of *C4H* genes in *P. trichocarpa*

Although we applied a low stringency screening strategy when isolating *PtreC4H* cDNAs, it might not be capable of collecting all the *C4H* genes in *P. tremuloides*. With the availability of the *P. trichocarpa* genome sequence, which has been predicted at this stage to contain approximately 58,036 gene models (http://genome.jgi-psf.org/Poptr1/Poptr1.home.html), we were able to analyze the genomic constitution of *C4H* genes in a tree species.

In the *Populus* genome, we identified three *C4H* homologs as putative *PtriC4H* genes. Their gene models are estExt_Genewise1_v1.C_LG_XIX2612, eugene3.00131281 and fgenesh1_pg.C_scaffold_164000060, respectively (Table 1). Sequence comparison of the three *PtriC4H* gene models revealed that estExt_Genewise1_v1.C_LG_XIX2612 and eugene3.00131281 showed 92% homology at the nucleotide level and 97% identity at the amino acid level, while fgenesh1_pg.C_scaffold_164000060 had low homology (about 66%) at both the nucleotide and amino acid levels to estExt_Genewise1_v1.C_LG_XIX2612.
or eugene3.00131281. Comparing the PtriC4H genes with the isolated PtreC4H cDNAs, we found that estExt_Genewise1_v1.C_LG_XIX2612 showed the highest homology (about 98%) with PtreC4H1-1 and PtreC4H1-2, while eugene3.00131281 was more similar to PtreC4H2-1 and PtreC4H2-2 showing about 98% homology. The homology between fgenesh1_pg.C_scaffold_164000060 and four PtreC4H genes was only about 66%. Therefore, estExt_Genewise1_v1.C_LG_XIX2612, eugene3.00131281 and fgenesh1_pg.C_scaffold_164000060 were designated PtriC4H1, PtriC4H2 and PtriC4H3, respectively (Table 1).

To analyze the PtriC4H gene expression in P. trichocarpa, we extracted total RNAs from eight different tissues, including young leaves, mature leaves, young stems, or eugene3.00131281. Comparing the PtriC4H genes with the isolated PtreC4H cDNAs, we found that estExt_Genewise1_v1.C_LG_XIX2612 showed the highest homology (about 98%) with PtreC4H1-1 and PtreC4H1-2, while eugene3.00131281 was more similar to PtreC4H2-1 and PtreC4H2-2 showing about 98% homology. The homology between fgenesh1_pg.C_scaffold_164000060 and four PtreC4H genes was only about 66%. Therefore, estExt_Genewise1_v1.C_LG_XIX2612, eugene3.00131281 and fgenesh1_pg.C_scaffold_164000060 were designated PtriC4H1, PtriC4H2 and PtriC4H3, respectively (Table 1).

To analyze the PtriC4H gene expression in P. trichocarpa, we extracted total RNAs from eight different tissues, including young leaves, mature leaves, young stems, or eugene3.00131281. Comparing the PtriC4H genes with the isolated PtreC4H cDNAs, we found that estExt_Genewise1_v1.C_LG_XIX2612 showed the highest homology (about 98%) with PtreC4H1-1 and PtreC4H1-2, while eugene3.00131281 was more similar to PtreC4H2-1 and PtreC4H2-2 showing about 98% homology. The homology between fgenesh1_pg.C_scaffold_164000060 and four PtreC4H genes was only about 66%. Therefore, estExt_Genewise1_v1.C_LG_XIX2612, eugene3.00131281 and fgenesh1_pg.C_scaffold_164000060 were designated PtriC4H1, PtriC4H2 and PtriC4H3, respectively (Table 1).

To analyze the PtriC4H gene expression in P. trichocarpa, we extracted total RNAs from eight different tissues, including young leaves, mature leaves, young stems, or eugene3.00131281. Comparing the PtriC4H genes with the isolated PtreC4H cDNAs, we found that estExt_Genewise1_v1.C_LG_XIX2612 showed the highest homology (about 98%) with PtreC4H1-1 and PtreC4H1-2, while eugene3.00131281 was more similar to PtreC4H2-1 and PtreC4H2-2 showing about 98% homology. The homology between fgenesh1_pg.C_scaffold_164000060 and four PtreC4H genes was only about 66%. Therefore, estExt_Genewise1_v1.C_LG_XIX2612, eugene3.00131281 and fgenesh1_pg.C_scaffold_164000060 were designated PtriC4H1, PtriC4H2 and PtriC4H3, respectively (Table 1).
and mature leaves. The expression of internode, young stems, young roots, phloem, young leaves levels in the developing xylem from the eighth to the 15th to the sixth internode and the expression was at lower strongly expressed in the developing xylem from the third to the sixth internode, although both PtriC4H1 and PtriC4H2 were expressed strongly in the xylem tissues. However, the expression pattern of PtriC4H3 was very different; it was strongly expressed in the young stems and young roots.

**Analysis of Populus C4H gene promoters**

In order to understand the regulatory mechanisms underlying different expression patterns of *Populus C4H* genes, the promoter sequences from the *PtreC4H* genes were isolated through genome walking. The promoter sequences of *PtreC4H1-1* and *PtreC4H1-2* were highly similar to each other. However, the homology between group *PtreC4H1* and *PtreC4H2* promoters was in the range of 51–55%. The divergent upstream sequences among the two group *PtreC4H* genes suggested that the mechanisms of gene regulation might be different.

We also analyzed the promoter sequences of the *PtriC4H* genes in the *P. trichocarpa* genome database at http://genome.jgi-psf.org/Poptr1/Poptr1.home.html. A 1.1 kb sequence upstream of the coding sequence was extracted from each of the *PtriC4H* genes, named *PtriC4H1p*, *PtriC4H2p* and *PtriC4H3p*, respectively. Comparison of the promoter sequences indicated that *PtriC4H3p* has 44–45% homology with both *PtriC4H1p* and *PtriC4H2p*; and *PtriC4H1p* has 55% homology with *PtriC4H2p*. Meanwhile it was found that *PtriC4H1p* is highly homologous to *PtreC4H1p* (about 83% sequence homology) and *PtriC4H2p* is similar to *PtreC4H2p* (about 60% sequence homology); whereas, the homologies between *PtriC4H3p* and the promoters of *PtreC4Hs* are about 46%.

Further sequence analysis indicated that the potential cis-regulatory elements in the promoters were different (Table 2). The box L, one of the three sequence motifs (box P, YTYYMMCMAMCM; box A, CCGTCC; and box L, YCYYACCCWACC) that have previously been shown to be present in plant PAL and 4CL gene promoters (Logemann et al. 1995), is identified in both *PtreC4H1ps* and *PtriC4H1p* but not in either *PtreC4H2p*, *PtriC4H2p* or *PtriC4H3p*. Box P is present...
only inPtriC4H3p but not in the promoters of the other C4H genes. Moreover, PtriC4H3p contains an H box (CCTACCDNNNNNNT), which is essential for both light regulation and elicitor induction (Loake et al. 1992).

**Discussion**

Although only one C4H gene was described in Arabidopsis (Raes et al. 2003), four C4H homologs in P. tremuloides and three in P. trichocarpa were identified in this study. Using a yeast expression system (Osakabe et al. 1999), we engineered two quaking aspen cDNAs, PtreC4H1-1 and PtreC4H2-1, into S. cerevisiae and found that the recombinant proteins were able to convert trans-cinnamate into p-coumaric acid (Fig. 1), confirming that the isolated PtreC4Hs encode active C4Hs in P. tremuloides. Thus, Populus C4Hs are encoded by a small gene family with at least four genes in P. tremuloides and three in P. trichocarpa.

Based on sequence similarity, C4H proteins have been classified into two classes, class I and class II (Nedelkina et al. 1999, Betz et al. 2001, Raes et al. 2003). Except for the C4Hs from Mesembryanthemum crystallinum (AAD11427), Citrus sinensis (CsC4H1) (AF255013), Phaseolus vulgaris (CAAD7059), Nicotiana tabacum (AAG62344 and AAG62345) and Zea mays (CYP73A8), which belong to class II, most of the others including CsC4H2 (AF255014) are class I C4Hs (Nedelkina et al. 1999, Betz et al. 2001, Raes et al. 2003). Phylogenetic analysis suggested that the identified PtreC4H genes, PtriC4H1 and PtriC4H2, which show about 90% identity at the amino acid level to CsC4H2, belong to class I; while PtriC4H3, having 86% identity at the amino acid level to CsC4H1, is a member of class II. Although the class I C4H genes are widely identified in plant species, the class II C4H genes so far are found only in a few species, including P. trichocarpa, Zea mays (Potter et al. 1995) and Citrus (Betz et al. 2001).

PAL, 4CL and C4H are three enzymes catalyzing the core reactions of phenylpropanoid metabolism. Together with the results from PAL and 4CL gene studies (Hu et al. 1998, Kao et al. 2002), all the three genes have multiple isoforms in Populus. In P. tremuloides and P. trichocarpa, the PtreC4H (Figs. 3, 4) and PtriC4H genes (Fig. 5) are differentially expressed, indicating that the divergent Populus C4H isoforms play different physiological roles. PtreC4H2 genes were found to be more xylem specific than PtreC4H1 genes in P. tremuloides. Similar to PtreC4H genes, PtriC4H1 and PtriC4H2 display distinct expression patterns. PtriC4H1 is strongly expressed in the developing xylem from the third to the sixth internode and less strongly expressed in the developing xylem from the eighth to the 15th internode; whereas transcripts of PtriC4H2 in the xylem from the eighth to the 15th internode are significantly more abundant than that in the xylem from the third to the sixth internode. Thus, the physiological functions of PtreC4H1 and PtreC4H2 genes in P. tremuloides and PtriC4H1 and PtriC4H2 in P. trichocarpa could be distinct. In addition, PtriC4H3 shows an expression pattern distinct from that of the other two C4H genes; it may play a different role in Populus development. The PtriC4H3 promoter contained a light- and elicitor-responsive cis-acting regulatory element, box H (Table 2, Loake et al. 1992), indicating that PtriC4H3 could be involved in stress responses. However, the involvement of other poplar C4H genes in stresses remains to be clarified. Differential expression of C4H genes has been reported for CsC4H1 and CsC4H2 in C. sinensis (Betz et al. 2001). CsC4H2 that was constitutively expressed in orange flavedo was suggested to play a role as a ‘housekeeping’ gene in the phenylpropanoid pathway, whereas CsC4H1 is induced only after wounding treatment (Betz et al. 2001). Taken together, these results indicated that the biosynthesis of specific metabolites might require different C4H isoforms, as in the case of PAL and 4CL isoforms in phenylpropanoid metabolism (Hu et al. 1998, Ehltign et al. 1999, Kao et al. 2002, Raes et al. 2003, Hamberger and Hahlbrock 2004).

On the other hand, the abundant transcripts of PtreC4H genes, PtriC4H1 and PtriC4H2 in developing xylem suggested that all of them could be involved in the biosynthesis of lignin, a major phenolic polymer that deposits in cell walls. In angiosperm tree development, guaiacyl (G) lignin is preferentially deposited in the xylem of young stems in its predominant cell type, the tracheary vessels; whereas, during maturation of stem, the content of syringyl (S) lignin increases with the development of fiber

<table>
<thead>
<tr>
<th>cis-elements</th>
<th>PtreC4H1p1</th>
<th>PtreC4H1p2</th>
<th>PtreC4H2p</th>
<th>PtriC4H1p</th>
<th>PtriC4H2p</th>
<th>PtriC4H3p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Box P (YTYYMMCMACMCM)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>1</td>
</tr>
<tr>
<td>Box A (CCGCC)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Box L (YCYYACCCWACC)</td>
<td>2</td>
<td>2</td>
<td>–</td>
<td>1</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>AC element (CYCACWWACC)</td>
<td>2</td>
<td>2</td>
<td>–</td>
<td>1</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Box H (CCTACCDNNNNNNT)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>1</td>
</tr>
</tbody>
</table>
and ray cells (Fergus et al. 1970a, Fergus et al. 1970b, Musha and Goring 1975, Saka and Goring 1985). In this study, the accumulation of \textit{PtriC4H1} transcripts was more abundant in tissues (such as the developing xylem from the third to the sixth internode) where G lignin was preferred, and less in tissues (such as the developing xylem from the eighth to the 15th internode) where S lignin was mainly deposited (Fig. 5), showing a close relationship between the expression of \textit{PtriC4H1} and the deposition of G lignin. Furthermore, examining the cis-acting regulatory elements in the promoters of \textit{Populus C4H} genes, we found that both \textit{PtreC4H1ps} and \textit{PtriC4H1p} contained the box L element (Table 2) (Logemann et al. 1995). The box L, originally identified in the promoter of the parsley \textit{PAL1} gene by footprint (Lois et al. 1989), was also known to be an AC element (Hattton et al. 1995) and has been found to enhance the xylem-specific expression of monolignol biosynthesis genes (Leyva et al. 1992, Hauffe et al. 1993, Hatton et al. 1995) via interacting with a number of MYB proteins (Tamagnone et al. 1998, Borevitz et al. 2000, Jin et al. 2000, Patzlaff et al. 2003a, Patzlaff et al. 2003b). Examining the promoters of Arabidopsis lignin-related genes, Raes et al. (2003) found that all of the promoters of \textit{G} rather than \textit{S} lignin biosynthesis genes had the box L/AC element consensus sequences, CYCACCWACC. Thus, \textit{PtreC4H1} genes and \textit{PtriC4H1}, showing \textit{G} lignin-related expression patterns and being regulated by the box L/AC elements, seem to act in the biosynthesis of \textit{G} lignin. Conversely, the expression level of \textit{PtriC4H2} (Fig. 5) was in line with the biosynthesis of \textit{S} lignin, showing higher levels in tissues (such as the developing xylem from the eighth to the 15th internode) where \textit{S} lignin was accumulated and lower levels in tissues (such as the developing xylem from the third to the sixth internode) where \textit{G} lignin was mainly deposited. Furthermore, both \textit{PtreC4H2p} and \textit{PtriC4H2p} lack the box L/AC element (Table 2), like other genes associated with the biosynthesis of \textit{S} lignin (Raes et al. 2003). Therefore, we proposed that \textit{PtreC4H2} and \textit{PtriC4H2} could be involved in the biosynthesis of \textit{S} lignin.

\textbf{Materials and Methods}

cDNA isolation

To construct a cDNA library, total RNA was extracted from the developing xylem of \textit{P. tremuloides} by the cetyltrimethylammonium bromide (CTAB) method (Chang et al. 1993) and was used for isolation of mRNA by the Poly (A) Quik mRNA Isolation Kit (Stratagene, La Jolla, CA, USA). cDNA was synthesized and ligated into Uni-ZAP XR vector according to the instruction manual of the ZAP-cDNA Synthesis Kit (Stratagene). Ligated DNA was then packaged by Gigapack III Gold Packaging Extract (Stratagene). The constructed library was amplified and screened with \textsuperscript{32}P-labeled aspen \textit{C4H} cDNA (Ge and Chiang 1996) at 48°C. Thirty positive plaques were isolated after screening a total of 15,000 p.f.u. Ten of the 30 positive plaques were picked up and the pBluescript phagemids were excised from the Uni-ZAP XR vector in vivo by using ExAssist helper phage with \textit{Escherichia coli} host strain SOLR. The rescued clones were analyzed by end sequencing using M13 reverse or forward primer. Four clones with a full-length ORF and showing different end sequences from each other were designated \textit{PtreC4H1-1}, \textit{PtreC4H1-2}, \textit{PtreC4H2-1} and \textit{PtreC4H2-2}, respectively. Another six clones were identical to one of the above four clones. Complete sequencing of the four \textit{PtreC4H1} clones revealed that \textit{PtreC4H1-1} and \textit{PtreC4H1-2}, and \textit{PtreC4H2-1} and \textit{PtreC4H2-2} showed high homology, indicating that they were allelic genes.

\textit{Yeast expression vector constructs}

To construct the yeast expression vector, forward primer H1-ATG (5'-GTATAATGGATCTCCTCCTCTTGTG-3') and reverse primer H1-TAG (5'-GAAACTAAAAAGGACCTTGCTTTG-3') together with template \textit{PtreC4H1-1 cDNA} were used for PCR amplification of the \textit{PtreC4H1-1 ORF}. Forward primer H2-ATG (5'-GATCATGGATCTCCTCCTCTTG-3') and reverse primer H2-TAA (5'-GAAATTAAAAAGGACCTTGCTTTG-3') together with template \textit{PtreC4H2-1 cDNA} were used for PCR amplification of the \textit{PtreC4H2-1} full-length coding sequences. After 5 min heating at 95°C, 15 cycles of amplification were performed as follows: 1 min denaturation at 94°C, 1 min annealing at 58°C, and 3 min extension at 72°C. The reaction was completed by 25 min extension at 72°C to ensure that the PCR product contained sufficient 3' A-overhangs. After gel purification, the amplicons were ligated into an autonomously replicating TA expression vector pYES2.1/V5-His-TOPO (Invitrogen, Carlsbad, CA, USA). Two sequence-verified clones pYESC4H1-70 and pYESC4H2-9 contained the identical sequences of \textit{PtreC4H1-1} and \textit{PtreC4H2-1}, respectively, and were used for yeast transformation to express \textit{PtreC4H1-1} and \textit{PtreC4H2-1} driven by the GAL1 promoter.

\textit{In vivo analysis of \textit{PtreC4H1-1} and \textit{PtreC4H2-1 activity in yeast cells}

The host \textit{S. cerevisiae} strain INVSC2(CPR) used for expression of \textit{PtreC4H1-1} and \textit{PtreC4H2-1} was previously constructed in our laboratory (Osakabe et al. 1999). \textit{PtreC4H1-1} and \textit{PtreC4H2-1} expression vectors (pYESC4H1-70 and pYESC4H2-9) were introduced into INVSc2(CPR) to create the yeast strains pYESC4H1-70-1 and pYESC4H2-9-2 for co-expression of \textit{Arabidopsis thaliana CPR} and \textit{PtreC4H1-1} or \textit{PtreC4H2-1} cDNAs by using the lithium acetate method as described in the instruction manual of the pYES2.1 TOPO TA Cloning Kit (Invitrogen) and selected on YND/HTA agar plates (6.7 g l\textsuperscript{-1} yeast-nitrogen base without amino acids, 5 g l\textsuperscript{-1} ammonium sulfate, 20 g l\textsuperscript{-1} glucose, 40 mg l\textsuperscript{-1} L-tryptophan, 20 mg l\textsuperscript{-1} L-histidine, 40 mg l\textsuperscript{-1} adenine hemisulfate salt and 20 g l\textsuperscript{-1} agar). Control yeast strain pYES2.1 was also created by transferring the control vector (pYES2.1/V5-His-TOPO) into INVSc2(CPR) cells. To eliminate the uncertainty during microsome preparation, a \textit{PtreC4H1-1} and \textit{PtreC4H2-1} activity assay was conducted by modifying the in vivo method described previously (Schoendorf et al. 2001, Shimada et al. 2001). The pYESC4H1-70-1, pYESC4H2-9-2 and pYES2.1 clones were grown to stationary phase in 3 ml of YND/HTA medium at 30°C. Cells were harvested by centrifuging at 3,000 r.p.m. for 5 min. After washing twice with sterile distilled water, the pellet was then used as an inoculum for 10 ml of expression culture in galactose-supplemented induction medium (20 g l\textsuperscript{-1} bactopeptone, 20 g l\textsuperscript{-1} galactose, 10 g l\textsuperscript{-1} yeast extract) and incubated at 28°C for 7–8 h. After inducing the co-expression of \textit{CPR} and \textit{PtreC4H1} genes, the medium was replaced with 5 ml
of fresh induction medium supplemented with 100 μM of the C4H substrate cinnamate and incubated at 28°C for 16 h. The supernatant was collected after centrifugation at 3,000 r.p.m. for 5 min and extracted with 5 ml of sterile distilled water and then dried by a vacuum. The ethyl acetate-extracted and dried reaction mixtures were analyzed by HPLC-UV/MSD as described (Osakabe et al. 1999).

In situ mRNA localization

In situ hybridization was performed by following the procedure previously described (Harding et al. 2002). To generate the probe, *PtreC4H1*-1 cDNA in pBluescript SK- (Stratagene) obtained by screening the cDNA library was digested with *Hind*III and *Xho*I. Following blunting of the ends by T4 DNA polymerase and gel recovery, the 3.6 kb DNA fragment was self-ligated by T4 DNA ligase. The resulting plasmid contained 0.7 kb template amounts, *P. tremuloides* tissues. As an endogenous control that was used to normalize template amounts, *P. tremuloides* (AF206999) or *P. trichocarpa* (AY652861) 18S rRNA fragments were amplified using a forward primer (5'-CGAAGACGATCAGATACCGTCCTA-3') and a reverse primer (5'-TTTCTCTAAGTGCTGGCCGAGT-3'). PCRs were carried out in a 25 μl volume containing 5.5 μl of diluted cDNA, 200 nM of each primer and 1 x SYBR Green PCR Master Mix (ABI) using the following conditions: 50°C for 2 min, 95°C for 10 min, 60 cycles of 95°C for 15 s and 60°C for 1 min. A further amplification, a thermal denaturing cycle at 95°C for 15 s, 60°C for 15 s and 95°C for 15 s for determining the dissociation curves was applied. The dissociation curves were used for verifying the specificity of PCR amplification. The results from gene-specific amplification were analyzed using a comparative Ct method which uses an arithmetic formula, 2^-ΔΔCt, to achieve results for relative quantification as suggested by the manufacturer. Ct represents the threshold cycle.

Genome walking

The upstream *PtreC4H4* genes were amplified by using a Universal Genome Walker™ Kit (Clontech, Mountain View, CA, USA). Eight enzymes, namely *Pvu*II, *Hpa*I, *Pmel*, *Rsa*I, *Dral*, *EcorV*, *Ssp*I and *Stul*, were used to prepare aspen genomic DNA libraries according to the ClonTech protocol. Two *PtreC4H1* gene-specific primers, H1W1 (5'-GGAATATGATCTTCTTATCC A AAGG-3') and H1W2 (5'-GTACTGAAAGAGATTTTAGG AGAG-3'), and two *PtreC4H2* gene-specific primers, H2W1 (5'-CAGGGAGGAGGAGATCATGATC-3') and H2W2 (5'-GA GTGGGGAGCTTTATGGTAGAAG-3'), were designed according to the sequence of the 5' end of *PtreC4H1* and *PtreC4H2* cDNAs and used for genome walking of *PtreC4H* promoters. Three overlapping fragments were obtained from *Dral*, *Pvu*II and *Hpa*I libraries using *PtreC4H1* gene-specific primers. The longest one was designated *PtreC4H1p1* (accession No. DQ522296). Two PCR walking fragments were obtained from *Stul* and *Rsa*I libraries by using *PtreC4H2* gene-specific primers. The longer one is about 1.1 kb.

In order to eliminate the mutation due to the two rounds of amplification, forward primer H1-5' (5'-GAAGCTTATCACCTTTTG GATTTCTGGAAAGTG-3') and reverse primer H1-3' (5'-CGG A TCTTCTTACCAAGAAGAGGAGG-3') for *PtreC4H1p* and forward primer H2-5' (5'-ACAATAATGCTAACAAGAAA TAATC-3') and reverse primer H2-3' (5'-CGGATCCATCATGG GACTTGTTCTTTGT-3') for *PtreC4H2p* were designed according to the sequences obtained by genome walking and used to amplify the promoter fragment from the aspen genome directly by PCR. Four PCR clones of *PtreC4H1* gene promoters were picked up and completely sequenced. Sequencing analysis revealed that the four PCR clones (designated *PtreC4H1p2*, accession No. DQ522297) were identical and showed 90% homology to *PtreC4H1p1*, indicating that one of them is the promoter of *PtreC4H1-1* and another is the promoter of
Borevitz, J.O., Xia, Y.J., Blount, J., Dixon, R.A. and Lamb, C. (2002) ER15442 to V.L.C. accession No. DQ522298) were identical and showed a two nucleotide difference from the sequence cloned by genome walking.

Acknowledgments
The project was supported by the US Department of Energy, Office of Basic Energy Sciences, grant number DE-FG02-05-ER15442 to V.L.C.

References


(Received March 8, 2006; Accepted May 13, 2006)