MODIFICATION OF 4-COUMARATE:COA LIGASE PROTEINS VIA MUTAGENESIS AND DOMAIN-SWAPPING ALTERS HYDROXYCINNAMATE SUBSTRATE SELECTIVITY AND PHENYLPROPANOID METABOLISM IN *POPULUS*

by

HAN-YI CHEN

(Under the Direction of Chung-Jui Tsai)

ABSTRACT

4-coumarate:CoA ligase (4CL) activates hydroxycinnamates to their corresponding CoA esters for utilization in the biosynthesis of phenylpropanoid end products such as lignin and flavonoids. Multiple 4CL isoforms were found in all species examined, and these isoforms were grouped into three classes by phylogenetic analysis. Lignin biosynthesis-associated isoforms in angiosperms and gymnosperms segregated into Class I and Class III, respectively. Non-lignin phenylpropanoid biosynthesis-associated 4CLs of both angiosperms and gymnosperms clustered into Class II. This suggests that distinct 4CLs evolved before the divergence of spermatophytes. *In vitro* enzyme kinetic analysis revealed a correlation between 4CL phylogenetic classification and substrate selectivity. Homology modeling and site-directed mutagenesis identified residue 302 in the substrate binding pocket as a major determinant of 4CL substrate selectivity. Transgenic *Populus* over-expressing wild-type or mutant 4CLs with different substrate selectivity exhibited distinct phenylpropanoid phenotypes. Unexpectedly, little

metabolic change was observed in the transgenic *Populus* over-expressing a loblolly pine Pta4CL1 or a Pta4CL1 domain-swap mutant. The weak transgenic effects were in contrast to the significant metabolic changes in transgenic *Populus* expressing loblolly pine Pta4CL3. One explanation for the distinct transgenic effects is that the backbone of Pta4CL1, which shares low sequence identity with *Populus* 4CLs, may be incompatible with the phenylpropanoid metabolons in *Populus*. Class II Pta4CL3, on the other hand, is evolutionary conserved with angiosperm 4CLs, and can exert an effect on *Populus* metabolism.

INDEX WORDS: 4CL, phenylpropanoid, metabolism, *Populus*, loblolly pine

MODIFICATION OF 4-COUMARATE:COA LIGASE PROTEINS VIA MUTAGENESIS AND DOMAIN-SWAPPING ALTERS HYDROXYCINNAMATE SUBSTRATE SELECTIVITY AND PHENYLPROPANOID METABOLISM IN *POPULUS*

by

HAN-YI CHEN

BS, Tatung University, Taiwan, 2001

MS, Michigan Technological University, 2008

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2013

MODIFICATION OF 4-COUMARATE:COA LIGASE PROTEINS VIA MUTAGENESIS AND DOMAIN-SWAPPING ALTERS HYDROXYCINNAMATE SUBSTRATE SELECTIVITY AND PHENYLPROPANOID METABOLISM IN *POPULUS*

by

HAN-YI CHEN

Major Professor: Chung-Jui Tsai

Committee: Scott A. Harding
Sarah F. Covert
Zachary Wood

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August2013

ACKNOWLEDGEMENTS

Completing this Ph. D. dissertation is like running a marathon, which is a journey filled with sweat and tears. Fortunately, I was not alone. Many people I met in this journey have given me great help. Without their support, it would be impossible for me to reach the finish line.

I would like to express my deepest appreciation to my advisor, Dr. Chung-Jui Tsai, who gave me the opportunity to explore the world of science. She provided me the advice, encouragement, and vision that enabled me to proceed through this doctoral program. My appreciation also goes to Dr. Scott Harding, who was always willing to share his research experience with me. He patiently guided me through many obstacles. Both Dr. Tsai and Dr. Harding have been role models for me as a junior scientist.

I also want to extend thanks to my committee members, Dr. Sarah Covert and Dr. Zachary Wood. Dr. Covert was very supportive and gave me helpful suggestions. Dr. Wood provided me guidance that was crucial for my research and dissertation writing. I will always remember our meetings and discussions that were informative and filled with passion.

I am grateful to all the friends that I have made here in Athens, surely including my former and present lab mates, for their support and help. When I was down, they

cheered me up with kind words, smiles, and encouragement. I will cherish their friendship forever and remember all the bitter and sweet that we shared.

Mostly importantly, I would like to thank my family. My wife, Yu-Chi, has been my most important support through these years. To me, she is where my home is. My sisters and my stepmother always encouraged me and listened to me patiently. Finally, my heartfelt thanks go to my parents for their unconditional love. Special thanks to my father. He is the reason that I started this great journey. He planted the seed of "studying abroad" in my mind when I was young, and this experience has enriched my life. Therefore, I would like to dedicate this dissertation to him, my dearest father.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

4CL and the phenylpropanoid pathway

4-coumarate:CoA ligase (4CL) catalyzes the ATP-dependent esterification of phenylpropanoid hydroxycinnamates with co-enzyme A (CoA) (Hahlbrock and Grisebach, 1970; Dixon et al., 2002; Vogt, 2010). The product thioesters are utilized in downstream pathways for the biosynthesis of lignin monomers, flavonoids and a plethora of taxon-specific UV protectants, antioxidants, insect deterrents, and fragrances (La Camera et al., 2004; Vogt, 2010). Lignin is the second most abundant natural polymer on earth behind cellulose (Davin and Lewis, 2005). It reinforces the cell wall matrix, and its hydrophobicity is crucial to water transport in vascular plants. The various phenylpropanoid end products, though classified as secondary compounds, are essential for plant growth, reproduction, adaptation, and defense (Dixon et al., 2002; Humphreys and Chapple, 2002; Vogt, 2010).

p-Coumarate (PA), caffeate (CA), ferulate (FA), and sinapate all serve as 4CL substrates (Lindl et al., 1973; Gross et al., 1975; Stockigt and Zenk, 1975; Knobloch and Hahlbrock, 1977). Catalysis is initiated when ATP and a hydroxycinnamate interact to yield a hydroxycinnamoyl-adenylate intermediate (hydroxycinnamoyl-AMP) at the expense of pyrophosphate (Gross and Zenk, 1974; Knobloch and Hahlbrock, 1975). In the second step, a molecule of CoA reacts with the hydroxycinnamoyl-adenylate

intermediate, and the final products, hydroxycinnamoyl-CoA and AMP, are released. Similar catalytic mechanisms are observed in all acyl-AMP forming enzymes, including acyl-CoA synthetases, firefly luciferases, and the phenylalanine-activating domain (PheA) of gramicidin S synthetase from *Brevibacillus brevis* (de Wet et al., 1987; Babbitt et al., 1992; Conti et al., 1997). These enzymes belong to the superfamily of adenylate-forming enzymes (Fulda et al., 1994; Schneider et al., 2003). Many structurally characterized members of this superfamily exhibit similar overall folding patterns with a large N-terminal and a small C-terminal domain. Within the N-terminal domain, two conserved motifs (i.e. A3 and A6 motifs of PheA) flank the substrate binding domain. These conserved motifs correspond to Box I (SSGTTGLPKGV) and II (GEICIRG; Becker-Andre et al., 1991; Fulda et al., 1994; Dieckmann et al., 1997) of 4CLs. The Box I motif is highly conserved in 4CLs and some other adenylate-forming enzymes, and Box II is absolutely conserved in all 4CL proteins (Stuible et al., 2000). The contribution of these two motifs to 4CL catalysis is still unclear (Stuible et al., 2000).

4CL *In Vivo* Function and Classification

Multiple 4CL isoforms were first identified in soybean cell suspensions by Lindle et al. (1973). To date, all characterized angiosperm (eg. *Arabidopsis*, aspen, tobacco) and gymnosperm species (eg. Spruce) express multiple isoforms (Lee and Douglas, 1996; Hu et al., 1998; Ehlting et al., 1999; Lindermayr et al., 2002; Hamberger and Hahlbrock, 2004; Friedmann et al., 2007; Hamberger et al., 2007; Koutaniemi et al., 2007). Based on their predicted amino acid sequences, angiosperm 4CL isoforms are assigned into two major groups, Class I and II, and gymnosperm 4CLs into Class II (Chen, 2008) and a

third group, Class III (Ehlting et al., 1999; Lindermayr et al., 2002; Saballos et al., 2012). Within each Class, 4CL isoforms from different species often share similar tissue expression patterns and substrate selectivity (Ehlting et al., 1999; Harding et al., 2002; Lindermayr et al., 2002; Hamberger and Hahlbrock, 2004). Many Class I 4CLs exhibit their highest *in vitro* catalytic efficiency with CA, while Class II 4CLs exhibit maximum catalytic efficiency using PA (Ehlting et al., 1999; Harding et al., 2002; Lindermayr et al., 2002). Class I *4CLs*, including *ArabidopsisAt4CL1&2* and aspen *Prt4CL1*, are abundant in lignifying tissues (Hu et al., 1998; Ehlting et al., 1999). Down-regulation of Class I *4CLs* in transgenic *Arabidopsis*, tobacco, and aspen has led to reduced lignin content and altered monolignol composition (Kajita et al., 1996; Lee et al., 1997; Hu et al., 1999). Class I *4CLs* have therefore been proposed to be involved in lignin biosynthesis. *Gm4CL1* and *At4CL4* also encode Class I isoforms, and are most highly expressed in stem/hypocotyls and root tissues (Lindermayr et al., 2002; Soltani et al., 2006). Gm4CL1 and At4CL4 exhibit a rarely observed capacity to utilize sinapate as a substrate *in vitro* (Lindermayr et al., 2002; Hamberger and Hahlbrock, 2004). Based on their tissue localization and substrate utilization patterns, At4CL4 and Gm4CL1 are thought to be involved in the biosynthesis of both lignin and non-lignin phenylpropanoids (Lindermayr et al., 2002; Hamberger and Hahlbrock, 2004).

Class II *4CLs*, including *Ptr4CL2*, *At4CL3*, and *Gm4CL3 & 4*, are more highly expressed in green/soft tissues than in lignifying tissues (Hu et al., 1998; Ehlting et al., 1999; Lindermayr et al., 2002). The strong affinity of Class II 4CLs for PA *in vitro* distinguishes them from Class I isoforms (Ehlting et al., 1999; Harding et al., 2002; Lindermayr et al., 2002). Nonstructural phenolics such as flavonoids are typically

abundant in tissues where Class II isoforms are preferentially expressed (Hu et al., 1998; Ehlting et al., 1999; Winkel-Shirley, 2001; Lindermayr et al., 2002). The CoA thioester of PA, *p*-coumaroyl-CoA (PA-CoA), is utilized along with malonyl-CoA to form chalcone, the precursor of all flavonoids, and a metabolic link between Class II 4CLs and flavonoid biosynthesis (Ebel et al., 1974; Hahlbrock and Grisebach, 1979; Uhlmann and Ebel, 1993; Harding et al., 2002). The Class III gymnosperm-specific group is phylogenetically distinct form Class I and Class II, and loblolly pine Pta4CL1 and its ortholog in radiate pine are the only isoforms that have been characterized in this group (Voo et al., 1995; Zhang and Chiang, 1997; Ehlting et al., 1999; Wagner et al., 2009). Similar to angiosperm 4CL isoforms in Class I, Pta4CL1 is well expressed in lignifying tissues, and down-regulation of the ortholog in *P. radiata* reduced lignin content (Wagner et al., 2009). However, increased *Pta4CL1* expression in N-deficient needles and root tips suggests that Pta4CL1 may also have a role in non-lignin phenylpropanoid biosynthesis (Chen, 2008). Catalytically, Pta4CL1, unlike the Class I or II angiosperm 4CLs that have been characterized, exhibits little substrate selectivity between PA, CA, and FA (Zhang and Chiang, 1997; Harding et al., 2002).

Phenylpropanoid Metabolic Channeling

In the widely accepted phenylpropanoid pathway model, the CoA-activation by 4CL is thought to occur with PA (Humphreys and Chapple, 2002; Vogt, 2010). The 4CL product PA-CoA is then utilized to form CA-CoA via the combined action of hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyltransferase (HCT) and *p*-coumarate 3-hydroxylase (C3H; Humphreys and Chapple, 2002; Hoffmann et al., 2003;

Hoffmann et al., 2004). This scenario was supported by evidence that C3H exhibits weak activity toward PA, but can efficiently 3-hydroxylate *p*-coumaroyl shikimate/quinate *in vitro* to form the corresponding caffeoyl esters (Schoch et al., 2001; Franke et al., 2002; Schoch et al., 2006). However, several stable-isotope feeding studies using explants of *Carthamus tinctorius* and *Robinia psudoacacia* found that the free acid form of the major hydroxycinnamates, including caffeate, ferulate and sinapate, occurs *in vivo*, and can be CoA-activated and incorporated into syringyl lignin (Yamauchi et al., 2002; Tsuji et al., 2003; Yamauchi et al., 2003; Sakakibara et al., 2007). This suggests the existence of alternative pathways to the C3H-HCT model, involving 4CL activation of caffeate and other free hydroxycinnamtes for lignin biosynthesis (Yamauchi et al., 2003; Sakakibara et al., 2007; Umezawa, 2010). However, the *in vivo* evidence that lignin is actually formed via the alternative (free acid) pathway is not yet available.

Metabolic channeling has been proposed to sequester intermediates and to provide flux control in the central as well as various phenylpropanoid-dependent branch pathways (Stafford, 1974; Rasmussen and Dixon, 1999; Winkel-Shirley, 1999; Achnine et al., 2004; Jørgensen et al., 2005; Ralston and Yu, 2006). Among the phenylpropanoid enzymes, members of the cytochrome P450 family, including cinnamate 4-hydroxylase (C4H), C3H, and flavonoid 3'-hydroxylase (Maria Koch et al., 1995), are endoplasmic reticulum (ER)-localized. In the central phenylpropanoid pathway, a protein-protein interaction between phenylalanine ammonia-lyase (PAL) and C4H has been reported (Achnine et al., 2004). This suggests that C4H may serve as the ER anchor of other phenylpropanoid pathway enzymes to form a metabolon (Burbulis and Winkel-Shirley, 1999; Saslowsky and Winkel-Shirley, 2001; Winkel, 2004; Ralston and Yu, 2006). Recently, 4CL and

HCT were found to exhibit a variety of protein-protein interactions with C4H and/or C3H (Bassard et al., 2012). The detection of such interactions suggests that 4CL and other phenylpropanoid enzymes participate in metabolic channeling. In addition, *in vitro* interactions between *Populus* C4H and C3H were found to impact the affinity and catalytic efficiency of CA formation from PA (Chen et al., 2011). This result supports the idea that 4CL activation of both PA and CA are likely important *in vivo*, and that such functional diversity by 4CL is dependent on metabolic channeling complexes, or metabolons (Chen et al., 2011; Bassard et al., 2012; Chen et al., 2013).

4CL Protein Structure and Engineering

The catalytically important substrate-binding pocket residues of 4CL1 and 2 were first identified in *Arabidopsis* (Stuible et al., 2000; Ehlting et al., 2001; Stuible and Kombrink, 2001; Schneider et al., 2003). The major difference between At4CL1 and At4CL2 was that At4CL2 exhibits low or no *in vitro* activity toward FA (Ehlting et al., 1999). To identify the binding pocket residues that determine the substrate selectivity of 4CLs, domain swap experiment was conducted (Ehlting et al., 2001). The study narrowed the putative substrate binding residues to within the region flanked by the two conserved 4CL motifs, Boxes I and II. Replacing this region of At4CL2 with that of At4CL1 enabled the mutant At4CL2 to utilize FA (Ehlting et al., 2001). A homology modeling approach using the PheA crystal structure was then used to identify the putative 4CL substrate binding residues (Stuible and Kombrink, 2001). PheA, like 4CL, is a member of the adenylate-forming enzyme superfamily, and its crystal structure was obtained in a complex with AMP and phenylalanine, which is structurally similar to PA (Conti et al.,

1997; Stuible and Kombrink, 2001). From the 4CL and PheA sequence alignments, nine putative 4CL substrate binding resides were identified (Stuible and Kombrink, 2001). Replacing the methionine at position 293 with alanine or phenylalanine enabled At4CL2 to utilize FA, and similar effects were observed by replacing lysine at 320 with leucine or alanine (Stuible and Kombrink, 2001). Furthermore, deletion of valine 356 enabled At4CL2 to utilize both FA and sinapate as substrates (Schneider et al., 2003).

A *bona fide* 4CLcrystal structure, (*Populus tomentosa* Pto4CL1) in complex with adenosine 5'-(3-(4-hydroxyphenyl)propyl phosphate (APP), to mimic a transitional intermediate of the adenylate-forming step was recently published (Hu et al., 2010). Nine residues involved in forming the hydrophobic substrate binding pocket (SBP) were identified. The hydrophobic interaction between the SBP and the 4-hydroxyphenyl group of APP is the major contributor of substrate binding, because only two residues, Ser-240 and Lys-303, may form hydrogen bounds with the substrate (Hu et al., 2010). The distance (3.52 Å) between Lys-303 and the 4-hydroxy group of APP suggested that only weak H-bonding would be possible, and the role of Lys-303 in substrate binding was predicted to be minor, compared to Ser-240 (Hu et al., 2010; Saballos et al., 2012). Instead, residue 303 was suggested to be responsible for 4CL substrate specificity (Saballos et al., 2012), and the sequence alignment of 4CLs from different species showed that position 303 was the least conserved residue among the nine residues in the SBP (Hu et al., 2010).

Populus and 4CL

Phenylpropanoid metabolism supports a highly diverse array of phytochemicals among plant lineages (Boerjan et al., 2003; Tsai et al., 2006; Weng et al., 2010). In the foliage of *Populus*, large amounts of non-structural phenylpropanoid derivatives, including chlorogenic acids, condensed tannins and phenolic glycosides, are maintained constitutively (Tsai et al., 2006; Harding et al., 2009). These constitutive pools also vary in responses to disease and abiotic stresses, and are especially important for woody species like *Populus* that face environmental fluctuations for several years before reaching reproductive competence (Dixon et al., 2002; Jansson and Douglas, 2007). *Populus*, compared with *Arabidopsis* and other herbaceous species, sustains more vigorous secondary growth and lignin biosynthesis, and this makes *Populus* an important model system for the study of wood chemistry and other wood properties (Jansson and Douglas, 2007). The well-developed *Populus* transformation systems and availability of growing genomic resources further qualify *Populus* as a model for *in vivo* gene function analysis via reverse genetics (Fillatti et al., 1987).

The published genome sequence of *Populus trichocarpa* (Tuskan et al., 2006) and the annotation of phenylpropanoid pathway genes (Tsai et al., 2006) have unveiled new *Populus* 4CL gene sequences that have not been characterized. In addition, a new loblolly pine 4CL gene, *Pta4CL3*, phylogenetically related to angiosperm Class II 4CLs has recently been identified (Chen, 2008). Characterization of the gymnosperm Pta4CL3 and the newly identified *Populus* 4CLs may address questions about the evolution and functional divergence of 4CL isoforms. 4CL mutagenesis in earlier studies identified the

residues that are important for 4CL catalysis toward FA, a relatively minor substrate for angiosperm 4CLs (Stuible and Kombrink, 2001; Schneider et al., 2003). Substrate selectivity for the predominant substrates, CA and PA, of Class I and Class II 4CLs, respectively, has not been investigated. 4CL *in vivo* function has been previously studied through gene expression manipulation (Kajita et al., 1996; Lee et al., 1997; Hu et al., 1999), but the focus was only on the lignin-associated isoforms. Characterizing the metabolic consequences of transgenic *Populus* over-expressing 4CLs that have distinct substrate selectivity will inform on the *in vivo* functions of different 4CLs. Together with the published 4CL crystal structure, 4CL substrate specificity may be manipulated and engineered with precision for *in vivo* studies of 4CL catalysis and metabolic control.

Dissertation Organization

The rest of this dissertation is organized into two chapters. Chapter 2 describes the characterization of loblolly pine Pta4CL1 and Pta4CL3, using both biochemical and transgenic approaches. Recombinant proteins were used for *in vitro* analysis of Pta4CL1 and Pta4CL3 substrate preference. Both genes were over-expressed in transgenic *Populus* to investigate their *in vivo* functions. Chapter 3 presents the characterization of new *Populus* Ptr4CL isoforms, using again biochemical and transgenic approaches, in conjunction with domain swapping and/or site-directed mutagenesis to alter 4CL substrate selectivity.

References

Achnine L, Blancaflor EB, Rasmussen S, Dixon RA (2004) Colocalization of L-phenylalanine ammonia-lyase and cinnarnate 4-hydroxylase for metabolic channeling in phenylpropanoid biosynthesis. Plant Cell **16:** 3098-3109

Babbitt PC, Kenyon GL, Martin BM, Charest H, Slyvestre M, Scholten JD, Chang KH, Liang PH, Dunawaymariano D (1992) Ancestry of the 4-chlorobenzoate dehalogenase - analysis of amino acid sequence identities among families of acyl-adenyl ligase, enoyl-CoA hydratases isomerase, and acyl-CoA thioesteraes. Biochemistry **31:** 5594-5604

Bassard JE, Richert L, Geerinck J, Renault H, Duval F, Ullmann P, Schmitt M, Meyer E, Mutterer J, Boerjan W, De Jaeger G, Mely Y, Goossens A, Werck-Reichhart D (2012) Protein-protein and protein-membrane associations in the lignin pathway. Plant Cell **24:** 4465-4482

Becker-Andre M, Schulze-Lefert P, Hahlbrock K (1991) Structural comparison, modes of expression, and putative cis-acting elements of the two 4-coumarate: CoA ligase genes in potato. J. Biol. Chem. **266:** 8551-8559

Boerjan W, Ralph J, Baucher M (2003) Lignin biosynthesis. Annual Review of Plant Biology **54:** 519-546

Burbulis IE, Winkel-Shirley B (1999) Interactions among enzymes of the Arabidopsis flavonoid biosynthetic pathway. Proceedings of the National Academy of Sciences **96:** 12929-12934

Chen H-C, Li Q, Shuford CM, Liu J, Muddiman DC, Sederoff RR, Chiang VL (2011) Membrane protein complexes catalyze both 4- and 3-hydroxylation of cinnamic acid derivatives in monolignol biosynthesis. Proceedings of the National Academy of Sciences **108:** 21253-21258

Chen H-C, Song J, Williams CM, Shuford CM, Liu J, Wang JP, Li Q, Shi R, Gokce E, Ducoste J, Muddiman DC, Sederoff RR, Chiang VL (2013) Monolignol pathway 4-coumaric acid:coenzyme A ligases in Populus. trichocarpa: Novel specificity, metabolic regulation, and simulation of coenzyme A ligation fluxes. Plant Physiology **161:** 1501-1516

Chen H-Y (2008) Characterization of two distinct 4-coumarate:CoA ligase (4CL) genes in loblolly pine (*Pinus taeda*). Unpublished master thesis

Conti E, Stachelhaus T, Marahiel MA, Brick P (1997) Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. Embo Journal **16:** 4174-4183

Davin LB, Lewis NG (2005) Lignin primary structures and dirigent sites. Current Opinion in Biotechnology **16:** 407-415

de Wet JR, Wood KV, DeLuca M, Helinski DR, Subramani S (1987) Firefly luciferase gene: structure and expression in mammalian cells. Molecular and Cellular Biology **7:** 725-737

Dieckmann R, PavelaVrancic M, Pfeifer E, VonDohren H, Kleinkauf H (1997) The adenylation domain of tyrocidine synthetase 1 - Structural and functional role of the interdomain linker region and the (S/T)GT(T/S)GXPKG core sequence. European Journal of Biochemistry **247:** 1074-1082

Dixon RA, Achnine L, Kota P, Liu CJ, Reddy MSS, Wang LJ (2002) The phenylpropanoid pathway and plant defence - a genomics perspective. Molecular Plant Pathology **3:** 371-390

Ebel J, Schaller B, Knobloch KH, Wellman E, Grisebach H, Hahlbrock K (1974) Coordinated changes in enzyme-activities of phenylpropanoid metabolism during growth of soybean cell-suspension cultures. Biochimica Et Biophysica Acta **362:** 417-424

Ehlting J, Buttner D, Wang Q, Douglas CJ, Somssich IE, Kombrink E (1999) Three 4-coumarate : coenzyme A ligases in Arabidopsis thaliana represent two evolutionarily divergent classes in angiosperms. Plant Journal **19:** 9-20

Ehlting J, Shin JJK, Douglas CJ (2001) Identification of 4-coumarate : coenzyme A ligase (4CL) substrate recognition domains. Plant Journal **27:** 455-465

Fillatti JJ, Sellmer J, McCown B, Haissig B, Comai L (1987) Agrobacterium mediated transformation and regeneration of Populus. Molecular & General Genetics **206:** 192-199

Franke R, Humphreys JM, Hemm MR, Denault JW, Ruegger MO, Cusumano JC, Chapple C (2002) The Arabidopsis REF8 gene encodes the 3-hydroxylase of phenylpropanoid metabolism. Plant Journal **30:** 33-45

Friedmann M, Ralph SG, Aeschliman D, Zhuang J, Ritland K, Ellis BE, Bohlmann J, Douglas CJ (2007) Microarray gene expression profiling of developmental transitions in Sitka spruce (Picea sitchensis) apical shoots. Journal of Experimental Botany **58:** 593-614

Fulda M, Heinz E, Wolter FP (1994) The fadD gene of Escherichia coli K12 is located closeto rnd at 39.6 min of the cgromosomal map and is a new member of the AMP-binding protein family. Molecular & General Genetics **242:** 241-249

Gross GG, Mansell RL, Zenk MH (1975) Hydroxycinnamate-coenzyme A ligase from lignifying tissue of higher-plants - some properties and taxonomic distribution. Biochemie Und Physiologie Der Pflanzen **168:** 41-51

Gross GG, Zenk MH (1974) Isolation and properties of hydroxycinnamate-CoA ligase from lignifying tissue of *Forsthia*. European Journal of Biochemistry **42:** 453-459

Hahlbrock K, Grisebach H (1970) Formation of coenzyme-A esters of cinnamic acids with an enzyme preparation cell suspension ceultures of parsley. Febs Letters **11:** 62-&

Hahlbrock K, Grisebach H (1979) Enzymic controls in the biosynthesis of lignin and flavonoids. Annual Review of Plant Physiology and Plant Molecular Biology **30:** 105-130

Hamberger B, Hahlbrock K (2004) The 4-coumarate : CoA ligase gene family in Arabidopsis thaliana comprises one rare, sinapate-activating and three commonly occurring isoenzymes. Proceedings of the National Academy of Sciences of the United States of America **101:** 2209-2214

Hamberger Br, Ellis M, Friedmann M, de Azevedo Souza C, Barbazuk B, Douglas CJ (2007) Genome-wide analyses of phenylpropanoid-related genes in Populus trichocarpa, Arabidopsis thaliana, and Oryza sativa: the Populus lignin toolbox and conservation and diversification of angiosperm gene familiesThis article is one of a selection of papers published in the Special Issue on Poplar Research in Canada. Canadian Journal of Botany **85:** 1182-1201

Harding SA, Jarvie MM, Lindroth RL, Tsai CJ (2009) A comparative analysis of phenylpropanoid metabolism, N utilization, and carbon partitioning in fast- and slow-growing Populus hybrid clones. Journal of Experimental Botany **60:** 3443-3452

Harding SA, Leshkevich J, Chiang VL, Tsai C-J (2002) Differential substrate inhibition couples kinetically distinct 4-coumarate:coenzyme A ligases with spatially distinct metabolic roles in quaking aspen. Plant Physiol. **128:** 428-438

Hoffmann L, Besseau S, Geoffroy P, Ritzenthaler C, Meyer D, Lapierre C, Pollet B, Legrand M (2004) Silencing of hydroxycinnamoy-coenzyme A shikimate/quinate hydroxycinnamoyltransferase affects phenylpropanoid biosynthesis. Plant Cell **16:** 1446-1465

Hoffmann L, Maury S, Martz F, Geoffroy P, Legrand M (2003) Purification, cloning, and properties of an acyltransferase controlling shikimate and quinate ester intermediates in phenylpropanoid metabolism. Journal of Biological Chemistry **278:** 95-103

Hu WJ, Harding SA, Lung J, Popko JL, Ralph J, Stokke DD, Tsai CJ, Chiang VL (1999) Repression of lignin biosynthesis promotes cellulose accumulation and growth in transgenic trees. Nature Biotechnology **17:** 808-812

Hu WJ, Kawaoka A, Tsai CJ, Lung JH, Osakabe K, Ebinuma H, Chiang VL (1998) Compartmentalized expression of two structurally and functionally distinct 4-coumarate : CoA ligase genes in aspen (Populus tremuloides). Proceedings of the National Academy of Sciences of the United States of America **95:** 5407-5412

Hu Y, Gai Y, Yin L, Wang X, Feng C, Feng L, Li D, Jiang X-N, Wang D-C (2010) Crystal structures of a Populus tomentosa 4-coumarate:CoA ligase shed light on its enzymatic mechanisms. Plant Cell **22:** 3093-3104

Humphreys JM, Chapple C (2002) Rewriting the lignin roadmap. Current Opinion in Plant Biology **5:** 224-229

Jansson S, Douglas CJ (2007) Populus: A model system for plant biology. *In* Annual Review of Plant Biology, Vol 58. Annual Reviews, Palo Alto, pp 435-458

Jørgensen K, Rasmussen AV, Morant M, Nielsen AH, Bjarnholt N, Zagrobelny M, Bak S, Møller BL (2005) Metabolon formation and metabolic channeling in the biosynthesis of plant natural products. Current Opinion in Plant Biology **8:** 280-291

Kajita S, Katayama Y, Omori S (1996) Alterations in the biosynthesis of lignin in transgenic plants with chimeric genes for 4-coumarate:coenzyme a ligase. Plant and Cell Physiology **37:** 957-965

Knobloch K-H, Hahlbrock K (1975) Isoenzymes of p-coumarate: CoA ligase from cell suspension cultures of *Glycine max*. European Journal of Biochemistry **52:** 311-320

Knobloch KH, Hahlbrock K (1977) 4-coumarate - CoA ligase from cell- suspension cultures of Petroselinum hortense Hoffm - partial -purification, substrate-specificity, and further properties. Archives of Biochemistry and Biophysics **184:** 237-248

Koutaniemi S, Warinowski T, KÃ ¤rkÃ ¶nen A, Alatalo E, Fossdal C, SaranpÃ ¤Ã ¤ P, Laakso T, Fagerstedt K, Simola L, Paulin L, Rudd S, Teeri T (2007)

Expression profiling of the lignin biosynthetic pathway in Norway spruce using EST sequencing and real-time RT-PCR. Plant Molecular Biology **65:** 311-328

La Camera S, Gouzerh G, Dhondt S, Hoffmann L, Fritig B, Legrand M, Heitz T (2004) Metabolic reprogramming in plant innate immunity: the contributions of phenylpropanoid and oxylipin pathways. Immunological Reviews **198:** 267-284

Lee D, Douglas CJ (1996) Two divergent members of a tobacco 4-coumarate:coenzyme A ligase (4CL) gene family - cDNA structure, gene inheritance and expression, and properties of recombinant proteins. Plant Physiology **112:** 193-205

Lee D, Meyer K, Chapple C, Douglas CJ (1997) Antisense suppression of 4-coumarate:coenzyme A ligase activity in Arabidopsis leads to altered lignin subunit composition. The Plant Cell **9:** 1985-1998

Lindermayr C, Mollers B, Fliegmann J, Uhlmann A, Lottspeich F, Meimberg H, Ebel J (2002) Divergent members of a soybean (Glycine max L.) 4-coumarate : coenzyme A ligase gene family - Primary structures, catalytic properties, and differential expression. European Journal of Biochemistry **269:** 1304-1315

Lindl T, Kreuzale.F, Hahlbroc.K (1973) Synthesis of para-coumaroyl coenzyme A with a partially purified para-coumarate-CoA ligase from cell suspension cultures of soybean (*Glycine max*). Biochimica Et Biophysica Acta **302:** 457-464

Maria Koch B, Sibbesen O, Halkier BA, Svendsen I, Lindberg MÃ¸ller B (1995) The primary sequence of cytochrome P450tyr, the multifunctional N-hydroxylase catalyzing the conversion of L-tyrosine to p-hydroxyphenylacetaldehyde oxime in the biosynthesis of the cyanogenic glucoside dhurrin in Sorghum bicolor(L.) Moench. Archives of Biochemistry and Biophysics **323:** 177-186

Ralston L, Yu O (2006) Metabolons involving plant cytochrome P450s. Phytochemistry Reviews **5:** 459-472

Rasmussen S, Dixon RA (1999) Transgene-mediated and elicitor-Induced perturbation of metabolic channeling at the entry point into the phenylpropanoid pathway. The Plant Cell Online **11:** 1537-1551

Saballos A, Sattler SE, Sanchez E, Foster TP, Xin Z, Kang C, Pedersen JF, Vermerris W (2012) Brown midrib2 (Bmr2) encodes the major 4-coumarate:coenzyme A ligase involved in lignin biosynthesis in sorghum (Sorghum bicolor (L.) Moench). The Plant Journal **70:** 818-830

Sakakibara N, Nakatsubo T, Suzuki S, Shibata D, Shimada M, Umezawa T (2007) Metabolic analysis of the cinnamate/monolignol pathway in Carthamus tinctorius seeds by a stable-isotope-dilution method. Organic & Biomolecular Chemistry **5:** 802-815

Saslowsky D, Winkel-Shirley B (2001) Localization of flavonoid enzymes in Arabidopsis roots. The Plant Journal **27:** 37-48

Schneider K, Hovel K, Witzel K, Hamberger B, Schomburg D, Kombrink E, Stuible HP (2003) The substrate specificity-determining amino acid code of 4-coumarate : CoA ligase. Proceedings of the National Academy of Sciences of the United States of America **100:** 8601-8606

Schoch G, Goepfert S, Morant M, Hehn A, Meyer D, Ullmann P, Werck-Reichhart D (2001) CYP98A3 from Arabidopsis thaliana is a 3 '-hydroxylase of phenolic esters, a missing link in the phenylpropanoid pathway. Journal of Biological Chemistry **276:** 36566-36574

Schoch GA, Morant M, Abdulrazzak N, Asnaghi C, Goepfert S, Petersen M, Ullmann P, Werck-Reichhart D (2006) The meta-hydroxylation step in the phenylpropanoid pathway: a new level of complexity in the pathway and its regulation. Environmental Chemistry Letters **4:** 127-136

Soltani B, Ehlting Jr, Hamberger Br, Douglas C (2006) Multiple cis-regulatory elements regulate distinct and complex patterns of developmental and wound-induced expression of Arabidopsis thaliana 4CL gene family members. Planta **224:** 1226-1238

Stafford HA (1974) The metabolism of aromatic compounds,

Stockigt J, Zenk MH (1975) Chemical synthesis and properties of hdroxycinnamoyl coenzyme A derivatives. Zeitschrift Fur Naturforschung C-a Journal of Biosciences **30:** 352-358

Stuible HP, Buttner D, Ehlting J, Hahlbrock K, Kombrink E (2000) Mutational analysis of 4-coumarate : CoA ligase identifies functionally important amino acids and verifies its close relationship to other adenylate-forming enzymes. Febs Letters **467:** 117-122

Stuible HP, Kombrink E (2001) Identification of the substrate specificity-conferring amino acid residues of 4-coumarate : coenzyme A ligase allows the rational design of mutant enzymes with new catalytic properties. Journal of Biological Chemistry **276:** 26893-26897

Tsai CJ, Harding SA, Tschaplinski TJ, Lindroth RL, Yuan YN (2006) Genome-wide analysis of the structural genes regulating defense phenylpropanoid metabolism in Populus. New Phytologist **172:** 47-62

Tsuji Y, Chen F, Yasuda S, Fukushima K (2003) The behavior of deuterium-labeled monolignol and monolignol glucosides in lignin biosynthesis in angiosperms. Journal of Agricultural and Food Chemistry **52:** 131-134

Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao RR, Bhalerao RP, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen GL, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, Cunningham R, Davis J, Degroeve S, Dejardin A, Depamphilis C, Detter J, Dirks B, Dubchak I, Duplessis S, Ehlting J, Ellis B, Gendler K, Goodstein D, Gribskov M, Grimwood J, Groover A, Gunter L, Hamberger B, Heinze B, Helariutta Y, Henrissat B, Holligan D, Holt R, Huang W, Islam-Faridi N, Jones S, Jones-Rhoades M, Jorgensen R, Joshi C, Kangasjarvi J, Karlsson J, Kelleher C, Kirkpatrick R, Kirst M, Kohler A, Kalluri U, Larimer F, Leebens-Mack J, Leple JC, Locascio P, Lou Y, Lucas S, Martin F, Montanini B, Napoli C, Nelson DR, Nelson C, Nieminen K, Nilsson O, Pereda V, Peter G, Philippe R, Pilate G, Poliakov A, Razumovskaya J, Richardson P, Rinaldi C, Ritland K, Rouze P, Ryaboy D, Schmutz J, Schrader J, Segerman B, Shin H, Siddiqui A, Sterky F, Terry A, Tsai CJ, Uberbacher E, Unneberg P, Vahala J, Wall K, Wessler S, Yang G, Yin T, Douglas C, Marra M, Sandberg G, Peer Y, Rokhsar D (2006) The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). Science **313:** 1596 - 1604

Uhlmann A, Ebel J (1993) Molecular-cloning and expression of 4-coumarate-coenzyme A ligase, an enzyme involved in resistance reponse of soybean (Glycine Max L) against pathogen attack. Plant Physiology **102:** 1147-1156

Umezawa T The cinnamate/monolignol pathway. Phytochemistry Reviews **9:** 1-17

Vogt T (2010) Phenylpropanoid Biosynthesis. Molecular Plant **3:** 2-20

Voo KS, Whetten RW, O'Malley DM, Sederoff RR (1995) 4-coumarate:coenzyme A ligase from loblolly pine xylem (isolation, characterization, and complementary DNA cloning). Plant Physiol. **108:** 85-97

Wagner A, Donaldson L, Kim H, Phillips L, Flint H, Steward D, Torr K, Koch G, Schmitt U, Ralph J (2009) Suppression of 4-Coumarate-CoA Ligase in the Coniferous Gymnosperm Pinus radiata. Plant Physiol. **149:** 370-383

Weng J-K, Akiyama T, Bonawitz ND, Li X, Ralph J, Chapple C (2010) Convergent evolution of syringyl lignin biosynthesis via distinct pathways in the Lycophyte selaginella and fowering plants. Plant Cell **22:** 1033-1045

Winkel-Shirley B (1999) Evidence for enzyme complexes in the phenylpropanoid and flavonoid pathways. Physiologia Plantarum **107:** 142-149

Winkel-Shirley B (2001) Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. Plant Physiology **126:** 485-493

Winkel BSJ (2004) Metabolic channeling in plants. Annual Review of Plant Biology **55:** 85-107

Yamauchi K, Yasuda S, Fukushima K (2002) Evidence for the biosynthetic pathway from sinapic acid to syringyl lignin using labeled sinapic acid with stable isotope at both methoxy groups in Robinia pseudoacacia and Nerium indicum. Journal of Agricultural and Food Chemistry **50:** 3222-3227

Yamauchi K, Yasuda S, Hamada K, Tsutsumi Y, Fukushima K (2003) Multiform biosynthetic pathway of syringyl lignin in angiosperms. Planta 216: 496-501

Zhang XH, Chiang VL (1997) Molecular cloning of 4-coumarate:coenzyme A ligase in loblolly pine and the roles of this enzyme in the biosynthesis of lignin in compression wood. Plant Physiol. 113: 65-74

CHAPTER 2

LOBLOLLY PINE CONIAINS PHYLOGENETICALLY DIVERGENT 4-COUMARATE:COA LIGASES THAT CONFER DISTINCT METABOLIC PHENOTYPES WHEN OVER-EXPRESSED IN *POPULUS*[1]

[1]Chen HY, Babst BA, Nyamdari B, Skyes R, Davis MF, Harding SA, and Tsai CJ. To be submitted to the Plant Journal.

Abstract

4-Coumarate:CoA ligase (4CL) catalyzes the formation of hydroxycinnamoyl-CoA esters for biosynthesis of various phenylpropanoid metabolites in plants. Phylogenetically distinct Class I and Class II 4CL isoforms occur in angiosperms, and are associated with lignin and non-lignin phenylpropanoid biosynthesis, respectively. In contrast, the few experimentally characterized gymnosperm 4CLs that are associated with lignin biosynthesis belong to the conifer-specific Class III. Loblolly pine (*Pinus taeda*) was recently shown to harbor a new isoform Pta4CL3 that is more closely related to Class II 4CLs than to other conifer (Class III) orthologs. Like angiosperm Class II 4CLs, *Pta4CL3* transcript levels were higher in needles and roots than in developing xylem, and its recombinant protein exhibited a substrate preference for *p*-coumaric acid. Ectopic expression of *Pta4CL3* in *Populus* led to significant changes in hydroxycinnamates-conjugates, including an increase of cinnamoyl-quinate in the leaves. This suggested a novel *in vivo* Pta4CL3 activity toward cinnamic acid. *Pta4CL1*-expressing transgenic *Populus*, on the other hand, exhibited little change in soluble phenylpropanoids relative to the wild-type control. This 'silent' metabolic phenotype may be ascribed to the evolutionarily divergence of Class III Pta4CL1, which renders it incompatible with the *Populus* metabolons for biosynthesis of phenylpropanoids. The evolutionarily conserved Class II Pta4CL3 can readily participate in the metabolic channeling, but exhibited additional *in vivo* enzyme properties not predicted by *in vitro* assays.

Introduction

Phenylpropanoids are plant-specific natural products that serve important functions during plant growth, development and interactions with biotic and abiotic environments (Vogt, 2010). They are synthesized from phenylalanine via the central phenylpropanoid pathway steps catalyzed by phenylalanine ammonia-lyase (PAL), cinnamate 4-hydroxylase (C4H) and 4-coumarate:CoA ligase (4CL; Dixon et al., 2002; Boerjan et al., 2003). 4CL enzymes exhibit substrate affinities toward 4-coumaric acid and its hydroxylated and/or methoxylated derivatives. The 4CL product hydroxycinnamoyl-CoAs are then channeled into different phenylpropanoid branch pathways to give rise to a diverse array of natural products, including lignin and flavonoids (Vogt, 2010). 4CL therefore occupies an important branch point in the regulation of phenylpropanoid metabolism.

4CL has been studied most extensively in angiosperms. It is encoded by a small gene family in *Arabidopsis* (Ehlting et al., 1999; Hamberger and Hahlbrock, 2004), *Glycine* (soybean; Lindermayr et al., 2002), *Populus* (Hu et al., 1998; Harding et al., 2002; Hamberger et al., 2007), *Nicotiana* (tobacco; Lee and Douglas, 1996), and *Oryza* (rice; Gui et al., 2011). These 4CLs are generally classified into two phylogenetic groups (Fig. 2-1) that share moderate sequence homology (<65% amino acid identity; Cukovic et al., 2001; Lindermayr et al., 2002; Hamberger and Hahlbrock, 2004). Several Class I members have been associated with lignin biosynthesis based on transgenic or mutant characterization (Kajita et al., 1996; Lee et al., 1997; Hu et al., 1999; Li et al., 2003). Class I recombinant proteins (*e.g.*, Arabidopsis At4CL1, At4CL2, aspen Ptr4CL1,

soybean Gm4CL2) typically exhibited a substrate preference for caffeic acid (CA) over *p*-coumaric acid (PA) or ferulic acid (FA) *in vitro* (Hu et al., 1998; Ehlting et al., 1999; Harding et al., 2002; Lindermayr et al., 2002). Class II 4CLs are thought to be involved in biosynthesis of non-structural phenolics, such as flavonoids and phytoalexins, with preferential expression in green/soft tissues (Hu et al., 1998; Ehlting et al., 1999; Lindermayr et al., 2002). Unlike Class I isoforms, all biochemically characterized Class II 4CLs in *Populus, Arabidopsis* and soybean favor PA over CA or FA *in vitro* (Ehlting et al., 1999; Harding et al., 2002; Lindermayr et al., 2002).

The gymnosperm loblolly pine (*Pinus taeda*) also contains two phylogenetically distinct 4CL isoforms, Pta4CL1 and Pta4CL3 (Fig. 2-1; Voo et al., 1995; Zhong et al., 1998; Chen, 2008). The xylem-abundant Class III 4CL is encoded by two genes, *Pta4CL1* and *Pta4CL2*, with identical coding sequences but different introns (Zhang and Chiang, 1997). *Pta4CL1/2* is highly expressed in developing xylem, especially during the formation of lignin-rich compression wood (Zhang and Chiang, 1997). It is also well expressed in needles and roots (Chen, 2008). The predicted amino acid sequence is phylogenetically distinct from that of the angiosperm 4CLs (Ehlting et al., 1999; Lindermayr et al., 2002; Gui et al., 2011). Recombinant Pta4CL1 exhibits no apparent substrate discrimination among PA, CA and FA (Harding et al., 2002). These findings led to the proposition that Pta4CL1 may have a comparatively broad suite of functional associations in pine stem (Voo et al., 1995; Zhang and Chiang, 1997; Harding et al., 2002). RNAi silencing of a radiata pine (*Pinus radiata*) ortholog *Pra4CL1* (sharing a 99.4% amino acid sequence identity with Pta4CL1) reduced lignin content in wood (Wagner et al., 2009), supporting a role in lignin biosynthesis. Pta4CL3 is

phylogenetically related to angiosperm Class II 4CLs. The *Pta4CL3* transcripts are detected in needles and root tips, and essentially absent in xylem (Chen, 2008). The function of Pta4CL3 was thought to be involved in flavonoid biosynthesis like other angiosperm Class II 4CLs (Ehlting et al., 1999; Harding et al., 2002; Lindermayr et al., 2002). However, *in vivo* support for the function is still lacking.

The research presented in this chapter aimed to address the *in vitro* enzyme properties and *in vivo* function of Pta4CL3. Consistent with the phylogenetic relatedness of Pta4CL3 to angiosperm Class II 4CLs, Pta4CL3 exhibited a substrate utilization pattern that is characteristic of Class II isoforms. This suggests an ancient origin of Class II 4CLs prior to the angiosperm-gymnosperm split during land plant evolution. While Class II 4CLs are relatively conserved between gymnosperm and angiosperm, the lignin-associated Class I and III isoforms appear to have evolved separately. Ectopic expression of *Pta4CL1* and *Pta4CL3* in *Populus* provided further support for their functional and evolutionary divergence. Over-expression of the evolutionarily conserved *Pta4CL3* in *Populus* resulted in altered phenylpropanoid profiles, whereas similar manipulation with *Pta4CL1* had little impact on *Populus* metabolism.

Materials and Methods

Phylogenetic analysis

The amino acid sequences of representative 4CLs were subjected to phylogenetic analysis using tools implemented in the Phylogeny.fr server (Dereeper et al., 2008). The

work flow included multiple sequence alignment by MUSCLE (Edgar, 2004), alignment refinement by Gblocks (Castresana, 2000), and phylogeny inference using the maximum likelihood algorithm PhyML (Guindon et al., 2010), with Jones-Taylor-Thornton matrix (Jones et al., 1992) as the substitution model and approximated likelihood ratio test (aLRT) for branch support (Anisimova and Gascuel, 2006). The inferred tree in Newick format was then imported into MEGA5 (Tamura et al., 2011) for visualization.

Real-time Quantitative RT-PCR (qRT-PCR) analysis

Total RNA was extracted using the Plant RNA Reagent (Invitrogen). First-strand cDNA synthesis was carried out using 2μg of total RNA, Oligo-dT primer, and SuperScript II reverse transcriptase (Applied Biosystem). Gene-specific primers (Table 1) were designed to span the coding sequence and/or 3'-UTR of *Pta4CL1*, *Pta4CL3*, *Ptr4CL1* and *Ptr4CL2*. Real-time PCR was performed with cDNA from 1.25 ng of total RNA per reaction, using the Absolute QPCR SYBR Green Mix kit (ABgene) and the Mx3000P real-time PCR system (Stratagene). MxPro software (Stratagene) was used to assess specificity of amplifications. Duplicate analysis of two biological replicates was performed for each sample. Relative transcript abundance was determined by the ΔCT method (Tsai et al., 2006) using aspen ubiquitin-conjugating enzyme E2 and elongation factor 1-β as housekeeping genes.

Recombinant Pta4CL1 and Pta4CL3 protein expression in *E. coli*

A recombinant Pta4CL1 expression construct was already available (Harding et al., 2002). To create a recombinant Pta4CL3 construct, the coding sequence of *Pta4CL3* was

PCR-amplified with a 5' *Bam*H I site introduced by primers (Table 1), TA-cloned into pCRII and sequence-confirmed. A *Bam*H I and *Xho* I restricted fragment was then subcloned into pET-30a (Novagen) and transformed into *E. coli* strain Rosetta 2 (DE3) pLysS for recombinant protein expression according to manufacturer's instructions. The His-tagged recombinant Pta4CL1 and Pta4CL3 was purified by Nickel resin (His-bind kit, Novagen), desalted with Sephadex G25 in buffer containing 100 mM Tris-HCl pH 7.5, 100 mM NaCl, 5 mM $MgCl_2$, 2.5 μg/ml leupeptin, 0.25 mM DTT, and 15% ethylene glycol. The protein concentration was determined using the Bradford method (1976) with bovine serum albumin (BSA) as standards. Protein samples were aliquoted and snap frozen in liquid nitrogen prior to storage at -80℃.

Enzyme assays

Each enzyme assay reaction (100 μl) contained 100 ng of recombinant Pta4CL1 or 300 ng of recombinant Pta4CL3 in 100 mM Tris-HCl (pH 7.5), 2.5 mM $MgCl_2$, 2.5 mM ATP, and 5 μg BSA. Cinnamic acid, 4-coumaric acid (PA), caffeic acid (CA), or ferulic acid (FA) were used as substrates, with concentrations ranging from 5 to 4000 μM. Mixed-substrate assays contained 200 ng (Pta4CL1) or 400 ng (Pta4CL3) recombinant proteins with equal-molar combinations of substrates from 5 to 120μM. Reactions were initiated by addition of 0.2 mM coenzyme A, incubated at 30℃ for 3 min and terminated by heating at 95℃ for 3 min. The 3 min incubation time was chosen from the linear range of the reaction product vs. time catalytic curves in preliminary enzyme assays for all substrates tested. After termination, the reaction mixtures were centrifuged at 15,000

rpm for 15 min, and the supernatants analyzed immediately or snap-frozen in liquid nitrogen and stored at -80℃ until use.

The reaction products (10 µl) were analyzed using an Agilent 1200 HPLC coupled to a 6220 time-of-flight mass spectrometer (HPLC-TOF/MS), equipped with a ZORBAX Eclipse Plus C18 column (2.1 mm x 15 cm x 3.5 µm, Agilent). Two mobile phase solvents were used: mobile phase A was 2% acetonitrile and 98% water with 10 mM formic acid (pH 7.0), while mobile phase B was 98% acetonitrile and 2% water. The elution gradient was 11.8% (v/v) mobile phase B for 2 minutes, and increased from 11.8% to 15% over 5 minutes with a hold at 15% for 5 min, at a flow rate of 0.4 ml/min. 4CL products were detected by MS at *m/z* 100 to 1200. Reaction product quantification was based on extracted ion chromatograms at *m/z* 898 for cinnamoyl-CoA, *m/z* 914 for 4-coumaroyl-CoA, *m/z* 930 for caffeoyl-CoA, and *m/z* 944 for feruloyl-CoA. Calibration curves were constructed using authentic CoA thioesters synthesized *in vitro* with recombinant Pta4CL1, and purified by flash chromatography according to Meng and Campbell (1997). Quantification of cinnamoyl-CoA was based on the calibration curve of 4-coumaroyl-CoA.

Populus transformation and transgenic plant characterization

The coding sequences of *Pta4CL1* and *Pta4CL3* were PCR amplified with an introduced 5' *Xba* I site, TA-cloned into pCRII, sequence-confirmed and subcloned into pBI121 binary vector at *Xba* I and *Bam*H I sites. Both constructs were mobilized into *Agrobacterium tumefaciens* strain C58/pMP90 using the freeze and thaw method (Holsters et al., 1978). Transformation of hybrid poplar (*Populus tremula* × *alba* clone

717-1B4) was performed as described by Han et al. (2000). PCR-positive transgenic lines were further screened by qRT-PCR for transgene expression, using gene-specific *Pta4CL1* and *Pta4CL3* primers (Table 2-1). Two transgenic events each, B16 and B22 for *Pta4CL1*-plants and H6 and H21 for *Pta4CL3*-plants, with the highest levels of transgene expression were propagated by rooted cuttings.

Hydroponic nitrogen stress treatments

The hydroponic nitrogen stress experiments were conducted as described in Harding et al. (2009). The rooted cuttings were transplanted into perlite pots for hydroponic growth in aerated nutrient solution. Nitrate- (2 mM) and ammonium-nitrogen (0.5 mM) were added daily, and the entire nutrient solution was replaced weekly. The pH of nutrient solution was monitored and maintained at 4.5 to 5.7 by adding 10 M potassium hydroxide or hydrochloric acid. After one month, the nitrogen level for the treatment group was reduced from 2.5 mM (N+) to 0.125 mM (N-). After 15 days of treatment, leaf plastochron index (LPI)-2, LPI-5, bark and xylem between internodes 16 and 21, and actively growing root tissues were harvested in liquid nitrogen for molecular and metabolite analyses.

Metabolite Profiling

Freeze-dried tissue powders (10 mg) were extracted with 315 µl of methanol and 200 µl of water containing 255 µM ortho-anisic acid, 127.5 µM adonitol, 1.3 mM resorcinol, 400µM $^{13}C_6$-cinnamic acid and 700 µM of D_5-benzoic acid as internal standards. After adding 165 µl of chloroform, the mixture was vortexed and sonicated in

a 70^{o}C water bath for 5 min, followed by centrifugation for 1 min. The supernatant was transferred to a new tube and the residues re-extracted (without internal standards) once more. A 700 µl aliquot of the combined extracts was evaporated to dryness in a speedvac for solid phase extraction, and the remaining was saved for HPLC-MS/TOF analysis. Vacuum-dried samples were resuspended in 10% methanol solution and incubated with Advanta resin (Applied Separations) for 20 min with agitation. After the resin settled, the extract was partitioned into relatively polar (bound) and less polar (unbound supernatant) fractions. The polar metabolites that were bound to Advanta resin were eluted by 600 µl acetonitrile. An aliquot of both Advanta-bound and unbound fractions was vacuum-dried and resuspended in 5 µl pyridine that contained alkanes as retention index standards. Methoxyamine hydrochloride and N-methyl-N-(trimethylsilyl)-trifluoroacetamide were then used for derivatization as described in Jeong et al. (2004). GC-MS (Agilent 7890A and 5975C) analysis conditions were as described (Jeong et al. 2004, Frost et al. 2012), using a DB5-MS column (30 m x 0.25 mm x 0.25 mm, with a 10 m DuraGuard pre-column, Agilent) and splitless injection (1 µl) with helium carrier gas at 1.2 ml/min. Mass spectra were recorded at 3.21 scans/sec from 6.2 to 15 min with a scan range of *m/z* 50 to 500, and at 2.66 scans/sec after 15 min with a scan range of *m/z* 50 to 600. The spectral data were searched against NIST, Agilent Fiehn GC/MS Metabolomics RTL Library, and our in-house authentic standard library by AnalyzerPro (SpectralWorks). The alignment of peaks from each sample was conducted by a custom web-based portal MetaLab (http://128.192.158.63/x/MetaLab/). The metabolites with confidence level over 70% and presented in 80% of all samples were selected for further analysis.

The unfractionated total extracts were analyzed by reverse phase HPLC-MS/TOF using a ZORBAX Rapid Resolution High Throughput RRHT) column (4.6 x 50 mm, 1.8 µm, Agilent). The separation gradient consisted of mobile phase A (97% water, 3% acetonitrile, 0.1% formic acid) and mobile phase B (97% acetonitrile, 3% water, 0.1% formic acid) as follows: 0-1 min 3% B, linear gradient to 17% B at 3 min, isocratic at 17% B to 5 min, linear gradient to 60% B at 9 min, and then linear gradient to 98% B at 11 min, at a flow rate of 1 ml/min. Metabolites were detected by DAD at 254, 280, and 310 nm, and by MS at *m/z* 100-1200. The parameters of MS/TOF were: negative mode, gas temperature 350℃, drying gas 13 l/min, nebulizer pressure 60 psig, Vcap 3500 V, and fragmentor 125 V. Data was processed by molecular feature extraction using MassHunter Qualitative Analysis (Agilent). The resultant compound lists from different samples were aligned by Mass Profiler, and peak areas obtained by MassHunter Quantitative Analysis (Agilent), followed by manual curation.

The peak area of each metabolite was normalized to that of the internal standard to obtain relative abundance values. The statistical significance between wild-type (WT) and transgenic plants was assessed by *R* using the SLIM package (Wang et al., 2010). The response ratio of each metabolite (transgenics over WT) was log2 transformed and visualized as heatmaps using the HeatMapper Plus tool (http://bar.utoronto.ca/ntools/cgi-bin/ntools_heatmapper_plus.cgi).

Metabolite identification

The unknown peak (*m/z* 321.098, RT 4.43 min) was predicted as cinnamoyl-quinate by KNApSAck database. To further confirm its identity, the methanol extract of

leaf LPI 5 from metabolic profile experiment was dried by centrivap and resuspended in 1/10 the original volume of methanol. The β-glucosidase from almond (Sigma) was first dissolved in 20 mM NaOAc, pH5.0, and the β-glucosidase was added to the resuspended sample. The glucosidase reaction was incubated at 37ºC for 1 hr. The β-glucosidase digestion yielded two highly abundant peaks (*m/z* 423.13 and 469.13) that overlapped with the target peak, while the target peak was not affected (Fig. 2-2A). Additional sample aliquot was used for fraction collection. The fraction collected sample that contained enriched target peak was then alkaline hydrolyzed by 0.1 N NaOH for 1 hr at room temperature and neutralized by 3M sodium acetate, pH 5.2. The target peak was completely hydrolyzed after the treatment (Fig. 2-2B), indicating a possible ester bounding (Lim et al., 2000). The hydrolyzed sample was analyzed by GC-MS, and the presence of cinnamic acid (Fig. 2-2C; 64% confidence) and quinic acid (Fig. 2-2D; 82 % confidence) was identified.

Crude protein extraction

The leaf (LPI 5) and xylem tissue harvested from the hydroponic experiment were ground to a fine powder in liquid nitrogen. Approximately 250 µl of tissue powder from individual plants or pooled from two biological replicates were used for crude protein extraction, with 2 or 3 replicates from each genotype × treatment group. Two volumes of ice-cold extraction buffer containing 100 mM Tris-HCl (pH7.5), 100 mM NaCl, 25 mM ascorbate, 1 mM DTT, 2 µg/ml leupeptin, and 5% (w/v) polyvinylpolypyrrolidone were added to the tissue aliquots. The samples were vortexed for 30 sec twice with a 30 sec interval on ice, followed by 15 min centrifugation at 15000 g. The supernatant was

desalted with Sephadex G-25 equilibrated in buffer containing 100 mM Tris-HCl (pH7.5), 5mM $MgCL_2$, 0.1 mM DTT, 2.5 μg/ml leupeptin, and 15% ethylene glycol. The crude protein concentration was similarly determined as for recombinant proteins. Samples were snap-frozen in liquid nitrogen and stored at -80℃. The crude protein extract was used for mixed substrate assays (100 μM equal molar of PA, CA, and FA) with the same conditions as recombinant protein assays, except 30 μg of crude protein was used in each reaction, and the reactions were incubated for 8 min.

Lignin analysis

The bark and developing xylem tissues were removed from the stem (internodes 16-21) of plants from the hydroponic experiment, and the remaining wood material was air dried and ball-milled to a fine powder. The powder was used for lignin analysis by Pyrolysis Molecular Beam Mass Spectrometry (PyMBMS) at the National Renewable Energy Laboratory (NREL). A subset of vegetatively propagated WT, transgenic control and *Pta4CL1*- (B16 and B22) and *Pta4CL3*- (H5 and H21) transgenic plants were grown in soil pots. After the plants reached 5-6 feet in height, the woody stem between internodes 26 and 40 was harvested, debarked and air-dried. The wood material was processed by Wiley mill to pass through a 20-mesh screen. Lignin content was determined by the Klason method according to Sluiter et al. (2010), using 200 mg of extractive-free wood meal. Samples were hydrolyzed in 2 ml of 72% H_2SO_4 for 1 hr with gentle shaking, diluted to 4% H_2SO_4 and autoclaved for 1 hr at 121^oC. The acid soluble lignin content of clarified hydrolysate was estimated by UV spectroscopy at 205 nm, using an extinction coefficient of ε =110 L/g-cm. The acid-insoluble lignin residue was

vacuum-filtered through a pre-weighed fiberglass filter and rinsed with water until the filtrate was acid-free. The residue and filter set was oven-dried and weighed to obtain the acid-insoluble lignin content. The sum of the acid soluble and acid-insoluble lignin is the total lignin amount.

Condensed tannin analysis

The actively growing root tissue was processed to a fine powder in liquid nitrogen and freeze-dried for condensed tannin analysis. About 10 mg of the powder was used for MeOH extraction as described in Harding et al, (2005). The MeOH extract supernatant was further extracted by 200 μl water, 200 μl chloroform, and 50 μl hexanes, followed by drying in a CentryVap. The dried supernatant and MeOH extract residue were resuspended in 980 μl of n-butanol-HCl (95% n-butanol and 5% 2M HCl) and 20 μl ferric-ammonium sulfate (2% ferric-ammonium-sulfate in 2 M HCl), followed by heating at 95℃ for 20 min. The soluble and residue-bound condensed tannins were estimated by spectrophotometer based on A_{554} as previously described (Porter et al., 1985; Payyavula et al., 2009).

Results

Biochemical characterization of recombinant Pta4CL1 and Pta4CL3

Recombinant Pta4CL1 exhibited similar substrate preference toward PA, CA, and FA based on their K_m values (Table 2-2). Its turnover rate (k_{cat}) was higher with FA than

with PA and CA, resulting in a higher catalytic efficiency (k_{cat}/K_m) toward FA. Pta4CL3, on the other hand, had a lower K_m and higher catalytic efficiency with PA than with CA or FA, a feature that has been reported for several Class II 4CL isoforms (Ehlting et al., 1999; Harding et al., 2002; Lindermayr et al., 2002). Cinnamic acid was a relatively poor *in vitro* substrate for either protein. The results revealed distinct kinetic properties of the two recombinant proteins: while Pta4CL1 exhibited a broad substrate utilization pattern, Pta4CL3 showed a substrate specificity toward PA (Table 2-2).

Our laboratory previously developed a mixed-substrate assay system for 4CL proteins that reliably predicted patterns of potential substrate competition (Harding et al., 2002). The broad substrate specificity of Pta4CL1 as revealed by kinetics analysis (Table 2-2) is consistent with its non-discriminating substrate utilization profile observed in mixed-substrate assays (Fig. 2-3A), as reported by Harding et al. (2002). The average CA-: PA-: FA-CoA product ratio was 1:1.3: 1.1. For recombinant Pta4CL3, PA-CoA was the predominant product in three-substrate assays, with an average product ratio of 1:4.5:0.6 (Fig. 2-3B). Likewise, PA was preferred over CA or FA in two-substrate assays (data not shown). In the absence of PA, CA was preferred to FA by recombinant Pta4CL3 (Fig. 2-3C). These results are consistent with the competitive inhibition patterns reported for the Class II aspen Ptr4CL2 (Harding et al., 2002), and support the kinetics analysis (Table 2-2) that PA is the preferred substrate of Pta4CL3, followed by CA, with FA being a relatively poor substrate.

Ectopic expression of *Pta4CL1* and *Pta4CL3* in transgenic *Populus*

To investigate their functions *in planta*, the two pine genes were over-expressed in *Populus* under control of the constitutive CaMV 35S promoter. Eight and nine independent transgenic events were obtained for the *Pta4CL1* and *Pta4CL3* constructs, respectively. The transgene expression was first examined in xylem by qRT-PCR (Figs. 2-4A & B). Three events from each transgenic group that showed high levels of transgene expression were further analyzed for transgene expression in leaf (Figs. 2-4C and D) and bark (Figs. 2-4E & F) tissues. Based on the data, two high expressers of each transgenic group along with the WT were vegetatively propagated and subjected to hydroponic N stress experiments.

4CL transcript levels were evaluated in leaf (LPI5), bark, xylem and root tissues of WT and transgenic plants maintained under N-replete (N+) or low-N (N-). The transcript levels of endogenous *4CL1* and *4CL2* were overall similar among WT and different transgenic lines, suggesting the absence of non-target effects due to ectopic expression of pine *4CL*. In both leaf and bark tissues, the transcript abundance of *Pta4CL1* or *Pta4CL3* transgene was 6- to 20-fold higher than that of the endogens (Figs. 2-5 & 2-6). In contrast to the situation in green tissues, the transgene transcript levels in xylem and root tissues were more comparable to those of endogenous *4CL1* and/or *4CL2* (Figs. 2-7 & 2-8). The transcript levels of endogenous *4CLs* and transgenes in N- treated plants were trending higher than in N+ treated plants, but the relative abundance between endogenous and transgenic 4CLs remained similar.

To gauge the effects at the protein level, crude proteins were extracted from leaves (LPI-5) and xylem of WT and transgenic plants for 4CL enzyme assays using mixed substrates of PA, CA and FA. Total 4CL activity was calculated by the sum of PA-CoA, CA-CoA and FA-CoA products, while the ratio of individual CoA thioester products over the total was used to represent the substrate utilization patterns of 4CL from different samples. The 4CL activity of WT leaves increased under N- growth (Fig. 2-9A), consistent with the transcriptional activation of the endogenous *4CL1* and *4CL2* by N stress (Fig. 2-5). This response was not observed in any of the transgenic lines, nor were there consistent differences between WT and transgenics, except for Pta4CL3-H6 which showed a weakly significant (p= 0.052) increase under N+ conditions (Fig. 2-9A). The substrate utilization patterns of leaf 4CL were not significantly changed among genotypes grown under N+, but the Pta4CL3-H6 line showed a trend of increased preference toward PA and a decrease toward CA, compares with wild-type (Fig. 2-9B). This pattern of altered substrate utilization became statistically significant under N- treatment (Fig. 2-9C). No difference in 4CL substrate utilization was observed for leaf extracts derived the PtaC4L1-plants. The xylem extracts from all plants showed the expected substrate preference for CA as reported (Harding et al. 2002), but there were no N treatment or transgenic effects on 4CL activity or substrate utilization patterns (Fig. 2-10).

Metabolite profiles of transgenic *Populus* in response to N treatments

Metabolite profiling revealed both genotypic and N stress-induced differences in several classes of soluble phenylpropanoids (Fig. 2-11). The overall phenolic profiles of

the *Pta4CL1* transgenics differed little from those of the WT in all above-ground tissues examined (sink and source leaves, bark and xylem), regardless of the N regime (Fig. 2-11). In contrast, the *Pta4CL3* transgenic lines showed considerable metabolic adjustments relative to the WT, and the response was more pronounced under N-limitation. At the tissue level, the metabolite differences were strongest in leaves and phloem, and insignificant in xylem. Levels of free hydroxycinnamates and their glucose esters were uniformly reduced in leaves and phloem of the *Pta4CL3* transgenics, while several hydroxycinnamoyl-quinate/shikimate esters increased (Fig. 2-11). Specifically, caffeoyl-quinate/shikimate esters increased in bark, whereas cinnamoyl-quinate increased in sink and source leaves. The occurrence of cinnamoyl-quinate, whose identity was confirmed in the present study (Fig. 2-11), has never been reported in *Populus* before. Flavonoids were the only soluble phenolics we detected that were similarly affected by both transgenes. While kaempferol levels increased in stressed leaves and roots, catechin levels decreased in leaves. Coniferyl alcohol and sinapyl alcohol changed little in all tissues (Fig. 2-11). Most primary metabolites detected by GC-MS, including sugars and amino acids, exhibited no significant difference between WT and either transgenic group (data not shown).

In contrast to the above-ground tissues, roots from the *Pta4CL*1 and *Pta4CL3* transgenic lines showed similar metabolite changes relative to WT (Fig. 2-11). The levels of chlorogenic acids in transgenic roots were lower than those of WT, especially under N-limitation, unlike the patterns observed in leaves and phloem. Levels of kaempferol trended higher in both transgenic groups during N-limited growth, suggesting diversion of phenylpropanoid intermediates into the flavonoid pathway under stress. To further test

this idea, we also measured levels of condensed tannins (CTs), a major end product of the flavonoid pathway in *Populus* roots (Tsai et al., 2006). Root CT levels did not differ between genotypes under N-replete conditions (Fig. 2-12). However, CT levels increased significantly under N stress in most transgenic lines but not WT. This resulted in significant increases of CTs in the transgenic roots relative to the WT under N-limitation, similar to the response of kaempferol.

Lignin analysis of transgenic *Populus*

The Klason lignin content, both soluble and insoluble, was not different between WT and either transgenic group (Fig. 2-13).The lignin content and syringyl-to-guaiacyl (S/G) lignin ratio were also analyzed by Pyrolysis Molecular Beam Mass Spectrometry (PyMBMS). No significant lignin or S/G ratio difference was observed between WT and transgenic groups.

Discussion

Our results showed that Pta4CL1 and Pta4CL3 in loblolly pine, like those in angiosperm species, are transcriptionally and catalytically distinct. Pta4CL1 and Pta4CL3 exhibited complementary expression profiles and contrasting substrate preferences that are consistent with their distinct phylogenetic groupings. Pta4CL3 shared multiple attributes with previously characterized Class II 4CLs from angiosperms. These included higher transcript abundance in foliar and root tissues than in lignifying xylem (Hu et al., 1998; Ehlting et al., 1999) and an *in vitro* substrate preference toward PA (Hu et al., 1998;

Hutzler et al., 1998; Ehlting et al., 1999; Lindermayr et al., 2002). Our findings suggest that Class II 4CLs are evolutionarily conserved among gymnosperms and angiosperms, and likely share similar functions in non-structural phenylpropanoid metabolism.

The xylem-abundant Pta4CL1, on the other hand, exhibited a uniquely broad substrate utilization pattern not found in other angiosperm 4CL isoforms involved in lignin biosynthesis. This is consistent with the weak phylogenetic association (*i.e.*, low branch support) of Class I and Class III 4CLs (Fig. 2-1), suggesting their divergent evolution. Lignin composition is known to differ between gymnosperms and angiosperms, with the former consisting primarily of G monolignols with low level of *p*-hydroxyphenyl (H) units, and the latter containing predominantly G and S monolignols (Boerjan et al., 2003). Whether the distinct characteristics of lignin-associated 4CLs in gymnosperms (Class III) and angiosperms (Class I) are linked to differential evolution of lignin biosynthesis in these taxa is an interesting possibility to consider.

Previously, suppression of a Class III *Pra4CL1* in radiata pine was shown to decrease lignin accumulation in wood, with a concomitant increase of flavonoids or flavonoid derivatives in bark (Wagner et al., 2009). The increase in flavonoids was thought to be a redirection of phenylpropanoid metabolites in response to reduced lignin pathway flux, and depend on a distinct 4CL member (Wagner et al., 2009) – presumably the *Pta4CL3*ortholog. Given that *Pta4CL1* and *Pta4CL3* share 66% sequence identity at the nucleotide level, the *Pra4CL1*RNAi construct was unlikely to affect expression of the *Pta4CL3*ortholog. The two studies thus complement each other in support of two functionally distinct 4CLs in gymnosperms.

Ectopic expression of *Pta4CL1* and *Pta4CL3* in transgenic *Populus* resulted in different metabolic phenotypes in green (leaf and bark) tissues, consistent with their distinct functions. An overall stronger effect was observed for the *Pta4CL3*-lines, with the *Pta4CL1*-transgenics largely indistinguishable from the WT. The predominant response in *Pta4CL3*-transgenics was the reduction of various hydroxycinnamates and their glucose esters. However, levels of the abundant chlorogenic acids and related hydroxycinnamoyl-quinate/shikimate esters were sustained or increased. This suggested that Pta4CL3 activity preferentially directed the phenylpropanoid pathway toward maintaining the chlorogenic acid-related metabolite pool at the expense of the hydroxycinnamate glucose ester pool. Of particular interest was the significant increase of an unusual hydroxycinnamate conjugate, cinnamoyl-quinate, in leaves of *Pta4CL3* plants. Two possible routes have been proposed for the synthesis of cinnamoyl-quinate, via cinnamoyl-CoA or *p*-coumaroyl-quinate (Liu et al., 1995; Niggeweg et al., 2004; Sonnante et al., 2010). Elevated 4CL activity could potentially increase the level of *p*-coumaroyl-CoA and its quinate ester. However, enzymes that catalyze the reductive dehydroxylation of hydroxycinnamoyl-quinate to cinnamoyl-quinate have yet to be identified, making this route less likely. By contrast, the occurrence of cinnamoyl-CoA and its derivatives has been reported in many species, including spinach, beet, soybean, cactus and scots pine (Walton and Butt, 1971; Lindl et al., 1973; Fliegmann et al., 1992; Liu et al., 1995). In fact, the occurrence of a trace amount of cinnamoyl-quinate was first reported in potato tuber as a potential intermediate of chlorogenic acid biosynthesis (Levy and Zucker, 1960). This supports the possibility that cinnamoyl-quinate may be derived from cinnamoyl-CoA in the *Pta4CL3*transgenics.

Cinnamic acid is a relatively poor substrate of 4CL *in vitro*, however, enzyme activity *in vivo* may be modulated by interactions among substrates, enzymes or other factors. In tobacco (*Nicotiana tabacum*), for instance, unknown macromolecules in the crude extract affected the ability of recombinant Nt4CL1 and Nt4CL2 to utilize cinnamic acid (Lee and Douglas, 1996). In another example, recombinant *p*-coumaroyl ester 3-hydroxylase (C3H) is known to catalyze the 3-hydroxylation of *p*-coumaroyl-shikimate, but not PA, in single-enzyme assays (Schoch et al., 2001; Chen et al., 2011). However, 3-hydroxylation of PA becomes effective when recombinant *Populus* PtrC3H3 complexes with cinnamate 4-hydroxylases PtrC4H1 and/or PtrC4H2 (Chen et al., 2011). 4CL and hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyltransferase (HCT) are enzymes that catalyze the lignin biosynthetic steps between C4H and C3H. Although 4CL and HCT are cytosolic, they can relocalize toward the endoplasmic reticulum (ER) upon co-expression of C4H and/or C3H (Bassard et al., 2012). These findings suggest that *in vivo* conversion of cinnamic acid by Pta4CL3 remains a possibility to account for the increased cinnamoyl-quinate level in the *Pta4CL3*-transgenics.

The weak metabolic phenotype of the *Pta4CL1*-transgenics was unexpected. As Pta4CL1 is structurally and kinetically divergent from both Class I and Class II 4CLs, it may not participate in the multi-enzyme metabolon formation and metabolic channeling that normally involve *Populus* 4CLs, while the evolutionarily conserved Pta4CL3 may. In addition, the recent reports of Achnine et al. (2004), Chen et al. (2011) and Bassard et al. (2012) suggested the involvement of other phenylpropanoid enzymes in metabolic channeling. Tobacco PAL1 and PAL2 exhibit differential affinities for association with ER-localized C4H (Achnine et al., 2004). This suggests that different isoforms of the

same gene family may have distinct involvement in the metabolons, and hence, *in vivo* functions. Our results support the idea that Pta4CL3 functionally resembles the Class II 4CL in *Populus* and participates in phenylpropanoid channeling. The more divergent Pta4CL1 may be less compatible for the multi-enzyme complexes, leading to the observed 'silent' metabolic phenotypes in transgenic *Populus*. Therefore, the *in vivo* roles of 4CL isoforms may be different from expectations based on *in vitro* assays.

References

Achnine L, Blancaflor EB, Rasmussen S, Dixon RA (2004) Colocalization of L-phenylalanine ammonia-lyase and cinnarnate 4-hydroxylase for metabolic channeling in phenylpropanoid biosynthesis. Plant Cell **16:** 3098-3109

Anisimova M, Gascuel O (2006) Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. Systematic Biology **55:** 539-552

Babbitt PC, Kenyon GL, Martin BM, Charest H, Slyvestre M, Scholten JD, Chang KH, Liang PH, Dunawaymariano D (1992) Ancestry of the 4-chlorobenzoate dehalogenase - analysis of amino acid sequence identities among families of acyl-adenyl ligase, enoyl-CoA hydratases isomerase, and acyl-CoA thioesteraes. Biochemistry **31:** 5594-5604

Bassard JE, Richert L, Geerinck J, Renault H, Duval F, Ullmann P, Schmitt M, Meyer E, Mutterer J, Boerjan W, De Jaeger G, Mely Y, Goossens A, Werck-Reichhart D (2012) Protein-protein and protein-membrane associations in the lignin pathway. Plant Cell **24:** 4465-4482

Becker-Andre M, Schulze-Lefert P, Hahlbrock K (1991) Structural comparison, modes of expression, and putative cis-acting elements of the two 4-coumarate: CoA ligase genes in potato. J. Biol. Chem. **266:** 8551-8559

Boerjan W, Ralph J, Baucher M (2003) Lignin biosynthesis. Annual Review of Plant Biology **54:** 519-546

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Analytical Biochemistry **72:** 248-254

Burbulis IE, Winkel-Shirley B (1999) Interactions among enzymes of the Arabidopsis flavonoid biosynthetic pathway. Proceedings of the National Academy of Sciences **96:** 12929-12934

Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Molecular Biology and Evolution **17:** 540-552

Chen H-C, Li Q, Shuford CM, Liu J, Muddiman DC, Sederoff RR, Chiang VL (2011) Membrane protein complexes catalyze both 4- and 3-hydroxylation of cinnamic acid derivatives in monolignol biosynthesis. Proceedings of the National Academy of Sciences **108:** 21253-21258

Chen H-C, Song J, Williams CM, Shuford CM, Liu J, Wang JP, Li Q, Shi R, Gokce E, Ducoste J, Muddiman DC, Sederoff RR, Chiang VL (2013) Monolignol pathway 4-coumaric acid:coenzyme A ligases in Populus. trichocarpa: Novel specificity, metabolic regulation, and simulation of coenzyme A ligation fluxes. Plant Physiology **161:** 1501-1516

Chen H-Y (2008) Characterization of two distinct 4-coumarate:CoA ligase (4CL) genes in loblolly pine (*Pinus taeda*). Unpublished master thesis

Conti E, Stachelhaus T, Marahiel MA, Brick P (1997) Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. Embo Journal **16:** 4174-4183

Cukovic D, Ehlting J, VanZiffle JA, Douglas CJ (2001) Structure and evolution of 4-coumarate : coenzyme A ligase (4CL) gene families. Biological Chemistry **382:** 645-654

Davin LB, Lewis NG (2005) Lignin primary structures and dirigent sites. Current Opinion in Biotechnology **16:** 407-415

de Wet JR, Wood KV, DeLuca M, Helinski DR, Subramani S (1987) Firefly luciferase gene: structure and expression in mammalian cells. Molecular and Cellular Biology **7:** 725-737

Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard JF, Guindon S, Lefort V, Lescot M, Claverie JM, Gascuel O (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Research **36:** W465-W469

Dieckmann R, PavelaVrancic M, Pfeifer E, VonDohren H, Kleinkauf H (1997) The adenylation domain of tyrocidine synthetase 1 - Structural and functional role of the interdomain linker region and the (S/T)GT(T/S)GXPKG core sequence. European Journal of Biochemistry **247:** 1074-1082

Dixon RA, Achnine L, Kota P, Liu CJ, Reddy MSS, Wang LJ (2002) The phenylpropanoid pathway and plant defence - a genomics perspective. Molecular Plant Pathology **3:** 371-390

Ebel J, Schaller B, Knobloch KH, Wellman E, Grisebach H, Hahlbrock K (1974) Coordinated changes in enzyme-activities of phenylpropanoid metabolism during growth of soybean cell-suspension cultures. Biochimica Et Biophysica Acta **362:** 417-424

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research **32:** 1792-1797

Ehlting J, Buttner D, Wang Q, Douglas CJ, Somssich IE, Kombrink E (1999) Three 4-coumarate : coenzyme A ligases in Arabidopsis thaliana represent two evolutionarily divergent classes in angiosperms. Plant Journal **19:** 9-20

Ehlting J, Shin JJK, Douglas CJ (2001) Identification of 4-coumarate : coenzyme A ligase (4CL) substrate recognition domains. Plant Journal **27:** 455-465

Fillatti JJ, Sellmer J, McCown B, Haissig B, Comai L (1987) Agrobacterium mediated transformation and regeneration of Populus. Molecular & General Genetics **206:** 192-199

Fliegmann J, Schroder G, Schanz S, Britsch L, Schroder J (1992) Molecular analysis of chalcone and dihydropinosylvin synthase from scots pine (Pinus sylvestris), and differential regulation of these and related enzyme activities in stressed plants. Plant Molecular Biology **18:** 489-503

Franke R, Humphreys JM, Hemm MR, Denault JW, Ruegger MO, Cusumano JC, Chapple C (2002) The Arabidopsis REF8 gene encodes the 3-hydroxylase of phenylpropanoid metabolism. Plant Journal **30:** 33-45

Friedmann M, Ralph SG, Aeschliman D, Zhuang J, Ritland K, Ellis BE, Bohlmann J, Douglas CJ (2007) Microarray gene expression profiling of developmental transitions in Sitka spruce (Picea sitchensis) apical shoots. Journal of Experimental Botany **58:** 593-614

Fulda M, Heinz E, Wolter FP (1994) The fadD gene of Escherichia coli K12 is located closeto rnd at 39.6 min of the cgromosomal map and is a new member of the AMP-binding protein family. Molecular & General Genetics **242:** 241-249

Gross GG, Mansell RL, Zenk MH (1975) Hydroxycinnamate-coenzyme A ligase from lignifying tissue of higher-plants - some properties and taxonomic distribution. Biochemie Und Physiologie Der Pflanzen **168:** 41-51

Gross GG, Zenk MH (1974) Isolation and properties of hydroxycinnamate-CoA ligase from lignifying tissue of *Forsthia*. European Journal of Biochemistry **42:** 453-459

Gui JS, Shen JH, Li LG (2011) Functional characterization of evolutionarily divergent 4-coumarate: coenzyme A ligases in rice. Plant Physiology **157:** 574-586

Guindon Sp, Dufayard J-Fo, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. Systematic Biology **59:** 307-321

Hahlbrock K, Grisebach H (1970) Formation of coenzyme-A esters of cinnamic acids with an enzyme preparation cell suspension ceultures of parsley. Febs Letters **11:** 62-&

Hahlbrock K, Grisebach H (1979) Enzymic controls in the biosynthesis of lignin and flavonoids. Annual Review of Plant Physiology and Plant Molecular Biology **30:** 105-130

Hamberger B, Hahlbrock K (2004) The 4-coumarate : CoA ligase gene family in Arabidopsis thaliana comprises one rare, sinapate-activating and three commonly occurring isoenzymes. Proceedings of the National Academy of Sciences of the United States of America **101:** 2209-2214

Hamberger Br, Ellis M, Friedmann M, de Azevedo Souza C, Barbazuk B, Douglas CJ (2007) Genome-wide analyses of phenylpropanoid-related genes in Populus trichocarpa, Arabidopsis thaliana, and Oryza sativa: the Populus lignin toolbox and conservation and diversification of angiosperm gene familiesThis article is one of a selection of papers published in the Special Issue on Poplar Research in Canada. Canadian Journal of Botany **85:** 1182-1201

Han KH, Meilan R, Ma C, Strauss SH (2000) An Agrobacterium tumefaciens transformation protocol effective on a variety of cottonwood hybrids (genus Populus). Plant Cell Reports **19:** 315-320

Harding SA, Jarvie MM, Lindroth RL, Tsai CJ (2009) A comparative analysis of phenylpropanoid metabolism, N utilization, and carbon partitioning in fast- and slow-growing Populus hybrid clones. Journal of Experimental Botany **60:** 3443-3452

Harding SA, Jiang HY, Jeong ML, Casado FL, Lin HW, Tsai CJ (2005) Functional genomics analysis of foliar condensed tannin and phenolic glycoside regulation in natural cottonwood hybrids. Tree Physiology **25:** 1475-1486

Harding SA, Leshkevich J, Chiang VL, Tsai C-J (2002) Differential substrate inhibition couples kinetically distinct 4-coumarate:coenzyme A ligases with spatially distinct metabolic roles in quaking aspen. Plant Physiol. **128:** 428-438

Hoffmann L, Besseau S, Geoffroy P, Ritzenthaler C, Meyer D, Lapierre C, Pollet B, Legrand M (2004) Silencing of hydroxycinnamoy-coenzyme A shikimate/quinate hydroxycinnamoyltransferase affects phenylpropanoid biosynthesis. Plant Cell **16:** 1446-1465

Hoffmann L, Maury S, Martz F, Geoffroy P, Legrand M (2003) Purification, cloning, and properties of an acyltransferase controlling shikimate and quinate ester

intermediates in phenylpropanoid metabolism. Journal of Biological Chemistry **278:** 95-103

Holsters M, Waele D, Depicker A, Messens E, Montagu M, Schell J (1978) Transfection and transformation of Agrobacterium tumefaciens. Molecular and General Genetics MGG **163:** 181-187

Hu WJ, Harding SA, Lung J, Popko JL, Ralph J, Stokke DD, Tsai CJ, Chiang VL (1999) Repression of lignin biosynthesis promotes cellulose accumulation and growth in transgenic trees. Nature Biotechnology **17:** 808-812

Hu WJ, Kawaoka A, Tsai CJ, Lung JH, Osakabe K, Ebinuma H, Chiang VL (1998) Compartmentalized expression of two structurally and functionally distinct 4-coumarate : CoA ligase genes in aspen (Populus tremuloides). Proceedings of the National Academy of Sciences of the United States of America **95:** 5407-5412

Hu Y, Gai Y, Yin L, Wang X, Feng C, Feng L, Li D, Jiang X-N, Wang D-C (2010) Crystal structures of a Populus tomentosa 4-coumarate:CoA ligase shed light on its enzymatic mechanisms. Plant Cell **22:** 3093-3104

Humphreys JM, Chapple C (2002) Rewriting the lignin roadmap. Current Opinion in Plant Biology **5:** 224-229

Hutzler P, Fischbach R, Heller W, Jungblut TP, Reuber S, Schmitz R, Veit M, WeissenbÃ¶ck G, Schnitzler Jr-P (1998) Tissue localization of phenolic compounds in plants by confocal laser scanning microscopy. Journal of Experimental Botany **49:** 953-965

Jørgensen K, Rasmussen AV, Morant M, Nielsen AH, Bjarnholt N, Zagrobelny M, Bak S, Møller BL (2005) Metabolon formation and metabolic channeling in the biosynthesis of plant natural products. Current Opinion in Plant Biology **8:** 280-291

Jansson S, Douglas CJ (2007) Populus: A model system for plant biology. *In* Annual Review of Plant Biology, Vol 58. Annual Reviews, Palo Alto, pp 435-458

Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. Computer Applications in the Biosciences **8:** 275-282

Kajita S, Katayama Y, Omori S (1996) Alterations in the biosynthesis of lignin in transgenic plants with chimeric genes for 4-coumarate:coenzyme a ligase. Plant and Cell Physiology **37:** 957-965

Knobloch K-H, Hahlbrock K (1975) Isoenzymes of p-coumarate: CoA ligase from cell suspension cultures of *Glycine max*. European Journal of Biochemistry **52:** 311-320

Knobloch KH, Hahlbrock K (1977) 4-coumarate - CoA ligase from cell- suspension cultures of Petroselinum hortense Hoffm - partial -purification, substrate-specificity, and further properties. Archives of Biochemistry and Biophysics **184:** 237-248

Koutaniemi S, Warinowski T, KÃ¤rkÃ¶nen A, Alatalo E, Fossdal C, SaranpÃ¤Ã¤ P, Laakso T, Fagerstedt K, Simola L, Paulin L, Rudd S, Teeri T (2007) Expression profiling of the lignin biosynthetic pathway in Norway spruce using EST sequencing and real-time RT-PCR. Plant Molecular Biology **65:** 311-328

La Camera S, Gouzerh G, Dhondt S, Hoffmann L, Fritig B, Legrand M, Heitz T (2004) Metabolic reprogramming in plant innate immunity: the contributions of phenylpropanoid and oxylipin pathways. Immunological Reviews **198:** 267-284

Lee D, Douglas CJ (1996) Two divergent members of a tobacco 4-coumarate:coenzyme A ligase (4CL) gene family - cDNA structure, gene inheritance and expression, and properties of recombinant proteins. Plant Physiology **112:** 193-205

Lee D, Meyer K, Chapple C, Douglas CJ (1997) Antisense suppression of 4-soumarate:coenzyme A ligase activity in Arabidopsis leads to altered lignin subunit composition. The Plant Cell **9:** 1985-1998

Levy CC, Zucker M (1960) Cinnamyl and coumaryl ester intermediates in the synthesis of chlorogenic aicd. Federation Proceedings **19:** 41-41

Li L, Zhou YH, Cheng XF, Sun JY, Marita JM, Ralph J, Chiang VL (2003) Combinatorial modification of multiple lignin traits in trees through multigene cotransformation. Proceedings of the National Academy of Sciences of the United States of America **100:** 4939-4944

Lim KB, Walker CRB, Guo L, Pellett S, Shabanowitz J, Hunt DF, Hewlett EL, Ludwig A, Goebel W, Welch RA, Hackett M (2000) Escherichia coli Î±-Hemolysin (HlyA) Is Heterogeneously Acylated in Vivo with 14-, 15-, and 17-Carbon Fatty Acids. Journal of Biological Chemistry **275:** 36698-36702

Lindermayr C, Mollers B, Fliegmann J, Uhlmann A, Lottspeich F, Meimberg H, Ebel J (2002) Divergent members of a soybean (Glycine max L.) 4-coumarate : coenzyme A ligase gene family - Primary structures, catalytic properties, and differential expression. European Journal of Biochemistry **269:** 1304-1315

Lindl T, Kreuzale.F, Hahlbroc.K (1973) Synthesis of para-coumaroyl coenzyme A with a partially purified para-coumarate-CoA ligase from cell suspension cultures of soybean (*Glycine max*). Biochimica Et Biophysica Acta **302:** 457-464

Liu Q, Bonness MS, Liu M, Seradge E, Dixon RA, Mabry TJ (1995) Enzymes of B-ring-deoxy flavonoid biosynthesis in elicited cell cultures of "old man" cactus (Cephalocereus senilis). Archives of Biochemistry and Biophysics **321:** 397-404

Maria Koch B, Sibbesen O, Halkier BA, Svendsen I, Lindberg MÅ¸ller B (1995) The primary sequence of cytochrome P450tyr, the multifunctional N-hydroxylase catalyzing the conversion of L-tyrosine to p-hydroxyphenylacetaldehyde oxime in the biosynthesis of the cyanogenic glucoside dhurrin in Sorghum bicolor(L.) Moench. Archives of Biochemistry and Biophysics **323:** 177-186

Meng HB, Campbell WH (1997) Facile enzymic synthesis of caffeoyl CoA. Phytochemistry **44:** 605-608

Niggeweg R, Michael AJ, Martin C (2004) Engineering plants with increased levels of the antioxidant chlorogenic acid. Nature Biotechnology **22:** 746-754

Payyavula R, Babst B, Nelsen M, Harding S, Tsai C-J (2009) Glycosylation-mediated phenylpropanoid partitioning in Populus tremuloides cell cultures. Bmc Plant Biology **9:** 151

Porter LJ, Hrstich LN, Chan BG (1985) The conversion of procyanidins and prodelphinidins to cyanidin and delphinidin. Phytochemistry **25:** 223-230

Ralston L, Yu O (2006) Metabolons involving plant cytochrome P450s. Phytochemistry Reviews **5:** 459-472

Rasmussen S, Dixon RA (1999) Transgene-mediated and elicitor-Induced perturbation of metabolic channeling at the entry point into the phenylpropanoid pathway. The Plant Cell Online **11:** 1537-1551

Saballos A, Sattler SE, Sanchez E, Foster TP, Xin Z, Kang C, Pedersen JF, Vermerris W (2012) Brown midrib2 (Bmr2) encodes the major 4-coumarate:coenzyme A ligase involved in lignin biosynthesis in sorghum (Sorghum bicolor (L.) Moench). The Plant Journal **70:** 818-830

Sakakibara N, Nakatsubo T, Suzuki S, Shibata D, Shimada M, Umezawa T (2007) Metabolic analysis of the cinnamate/monolignol pathway in Carthamus tinctorius seeds by a stable-isotope-dilution method. Organic & Biomolecular Chemistry **5:** 802-815

Saslowsky D, Winkel-Shirley B (2001) Localization of flavonoid enzymes in Arabidopsis roots. The Plant Journal **27:** 37-48

Schneider K, Hovel K, Witzel K, Hamberger B, Schomburg D, Kombrink E, Stuible HP (2003) The substrate specificity-determining amino acid code of 4-coumarate : CoA ligase. Proceedings of the National Academy of Sciences of the United States of America **100:** 8601-8606

Schoch G, Goepfert S, Morant M, Hehn A, Meyer D, Ullmann P, Werck-Reichhart D (2001) CYP98A3 from Arabidopsis thaliana is a 3 '-hydroxylase of phenolic esters, a missing link in the phenylpropanoid pathway. Journal of Biological Chemistry **276:** 36566-36574

Schoch GA, Morant M, Abdulrazzak N, Asnaghi C, Goepfert S, Petersen M, Ullmann P, Werck-Reichhart D (2006) The meta-hydroxylation step in the phenylpropanoid pathway: a new level of complexity in the pathway and its regulation. Environmental Chemistry Letters **4:** 127-136

Sluiter JB, Ruiz RO, Scarlata CJ, Sluiter AD, Templeton DW (2010) Compositional analysis of lignocellulosic feedstocks. 1. Review and description of methods. Journal of Agricultural and Food Chemistry **58:** 9043-9053

Soltani B, Ehlting Jr, Hamberger Br, Douglas C (2006) Multiple cis-regulatory elements regulate distinct and complex patterns of developmental and wound-induced expression of Arabidopsis thaliana 4CL gene family members. Planta **224:** 1226-1238

Sonnante G, D'Amore R, Blanco E, Pierri CL, De Palma M, Luo J, Tucci M, Martin C (2010) Novel hydroxycinnamoyl-coenzyme A quinate transferase genes from artichoke are involved in the synthesis of chlorogenic acid. Plant Physiology **153:** 1224-1238

Stafford HA (1974) The metabolism of aromatic compounds,

Stockigt J, Zenk MH (1975) Chemical synthesis and properties of hdroxycinnamoyl coenzyme A derivatives. Zeitschrift Fur Naturforschung C-a Journal of Biosciences **30:** 352-358

Stuible HP, Buttner D, Ehlting J, Hahlbrock K, Kombrink E (2000) Mutational analysis of 4-coumarate : CoA ligase identifies functionally important amino acids and verifies its close relationship to other adenylate-forming enzymes. Febs Letters **467:** 117-122

Stuible HP, Kombrink E (2001) Identification of the substrate specificity-conferring amino acid residues of 4-coumarate : coenzyme A ligase allows the rational

design of mutant enzymes with new catalytic properties. Journal of Biological Chemistry **276:** 26893-26897

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution **28:** 2731-2739

Tsai CJ, Harding SA, Tschaplinski TJ, Lindroth RL, Yuan YN (2006) Genome-wide analysis of the structural genes regulating defense phenylpropanoid metabolism in Populus. New Phytologist **172:** 47-62

Tsuji Y, Chen F, Yasuda S, Fukushima K (2003) The behavior of deuterium-labeled monolignol and monolignol glucosides in lignin biosynthesis in angiosperms. Journal of Agricultural and Food Chemistry **52:** 131-134

Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao RR, Bhalerao RP, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen GL, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, Cunningham R, Davis J, Degroeve S, Dejardin A, Depamphilis C, Detter J, Dirks B, Dubchak I, Duplessis S, Ehlting J, Ellis B, Gendler K, Goodstein D, Gribskov M, Grimwood J, Groover A, Gunter L, Hamberger B, Heinze B, Helariutta Y, Henrissat B, Holligan D, Holt R, Huang W, Islam-Faridi N, Jones S, Jones-Rhoades M, Jorgensen R, Joshi C, Kangasjarvi J, Karlsson J, Kelleher C, Kirkpatrick R, Kirst M, Kohler A, Kalluri U, Larimer F, Leebens-Mack J, Leple JC, Locascio P, Lou Y, Lucas S, Martin F, Montanini B, Napoli C, Nelson DR, Nelson C, Nieminen K, Nilsson O, Pereda V, Peter G, Philippe R, Pilate G, Poliakov A, Razumovskaya J, Richardson P, Rinaldi C, Ritland K, Rouze P, Ryaboy D, Schmutz J, Schrader J, Segerman B, Shin H, Siddiqui A, Sterky F, Terry A, Tsai CJ, Uberbacher E, Unneberg P, Vahala J, Wall K, Wessler S, Yang G, Yin T, Douglas C, Marra M, Sandberg G, Peer Y, Rokhsar D (2006) The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). Science **313:** 1596 - 1604

Uhlmann A, Ebel J (1993) Molecular-cloning and expression of 4-coumarate-coenzyme A ligase, an enzyme involved in resistance reponse of soybean (Glycine Max L) against pathogen attack. Plant Physiology **102:** 1147-1156

Umezawa T The cinnamate/monolignol pathway. Phytochemistry Reviews **9:** 1-17

Vogt T (2010) Phenylpropanoid Biosynthesis. Molecular Plant **3:** 2-20

Voo KS, Whetten RW, O'Malley DM, Sederoff RR (1995) 4-coumarate:coenzyme A ligase from loblolly pine xylem (isolation, characterization, and complementary DNA cloning). Plant Physiol. **108:** 85-97

Wagner A, Donaldson L, Kim H, Phillips L, Flint H, Steward D, Torr K, Koch G, Schmitt U, Ralph J (2009) Suppression of 4-Coumarate-CoA Ligase in the Coniferous Gymnosperm Pinus radiata. Plant Physiol. **149:** 370-383

Walton E, Butt VS (1971) The demonstration of cinnamyl-CoA synthetase activity in leaf extracts. Phytochemistry **10:** 295-304

Wang H-Q, Tuominen LK, Tsai C-J (2010) SLIM: a sliding linear model for estimating the proportion of true null hypotheses in datasets with dependence structures. Bioinformatics **27:** 225-231

Weng J-K, Akiyama T, Bonawitz ND, Li X, Ralph J, Chapple C (2010) Convergent evolution of syringyl lignin biosynthesis via distinct pathways in the Lycophyte selaginella and fowering plants. Plant Cell **22:** 1033-1045

Winkel-Shirley B (1999) Evidence for enzyme complexes in the phenylpropanoid and flavonoid pathways. Physiologia Plantarum **107:** 142-149

Winkel-Shirley B (2001) Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. Plant Physiology **126:** 485-493

Winkel BSJ (2004) Metabolic channeling in plants. Annual Review of Plant Biology **55:** 85-107

Yamauchi K, Yasuda S, Fukushima K (2002) Evidence for the biosynthetic pathway from sinapic acid to syringyl lignin using labeled sinapic acid with stable isotope at both methoxy groups in Robinia pseudoacacia and Nerium indicum. Journal of Agricultural and Food Chemistry **50:** 3222-3227

Yamauchi K, Yasuda S, Hamada K, Tsutsumi Y, Fukushima K (2003) Multiform biosynthetic pathway of syringyl lignin in angiosperms. Planta **216:** 496-501

Zhang XH, Chiang VL (1997) Molecular cloning of 4-coumarate:coenzyme A ligase in loblolly pine and the roles of this enzyme in the biosynthesis of lignin in compression wood. Plant Physiol. **113:** 65-74

Zhong R, Morrison WH, Negrel J, Ye Z-H (1998) Dual methylation pathways in lignin biosynthesis. The Plant Cell Online **10:** 2033-2045

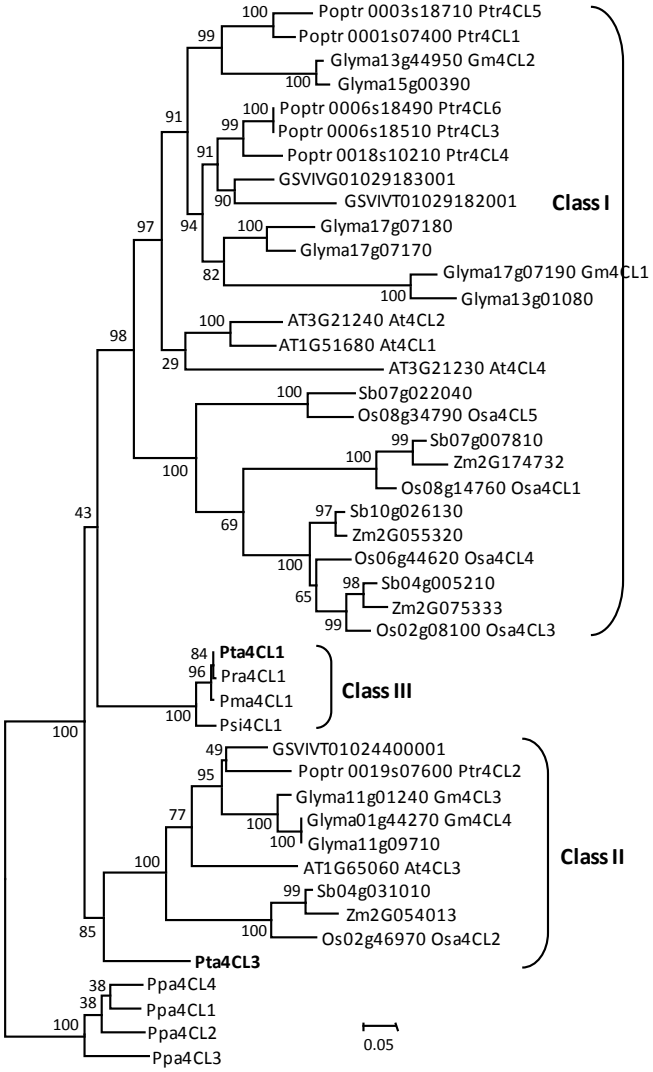


Figure 2-1. Phylogenetic analysis of representative 4CLs using the full length amino acid sequences. The classification is based on their enzyme kinetic properties and phylogenetic relations. The tree was constructed using the Approximate Likelihood-Ratio Test (aLRT) method implemented in the PhyML program, and moss Ppa4CL group was used as the root. The isoforms presented with locus names were obtained from Phytozome, and the others were derived from GenBank. The GenBank IDs are as follow: Pta4CL1, AAA92669; Pra4CL1, ACF35279; Psi4CL1, ABR17998; Pma4CL1, ACO40513; Ppa4CL1, ABY21308; Ppa4CL2, ABY21309; Ppa4CL3, ABY21310; Ppa4CL4, ABY21311.

Table 2-1. Primer Sequences. Sequences of primers that were used for cloning and real-time RT-PCR.

Primer sequence	
Cloning	
Pta4CL3 (pET)	F: AAGGATCCATTGAGGTACAGTCAGCTCC
	R: TCATGCTTTGATTTCAACGG
Pta4CL1 (pBI121)	F: ATTCTAGATGCCAACGGAATCAGAA
	R: TCATTTTGCTGCCAGTCTGCTTC
Pta4CL3 (PBI121)	F: ATGATTGAGGTACAGTCAGC
	R: TCATGCTTTGATTTCAACGG
qRT-PCR	
Pta4CL1	F: GATTTGAGAAGCAGAYTGGCAGCA
	R: AGAGTGTAGGGCGTTGACAATCCA
Pta4CL3	F: AAGGATTTGAGGGCGAAGCTCTCT
	R: CCACCCTTACAGAGAAGCAGGAAT
Ptr4CL1	F: CCAGGCATATAACTGAAGA
	R: TTGAATTACTCCATATGGCGACAGGAG
Ptr4CL2	F: AATTCATTCCGTTAAATCTGCATT
	R: CAGAATGATGGGTTTGTAGTAATT
Aspen ubiquitin-conjugating enzyme E2	F: CTGAAGAAGGAGATGACARCMCCA
	R: GCATCCCTTCAACACAGTTTCAMG
Aspen elongation factor 1-β	F: AAGAGGACAAGAAGGCAGCA
	R: CTAACCGCCTTCTCCAACAC

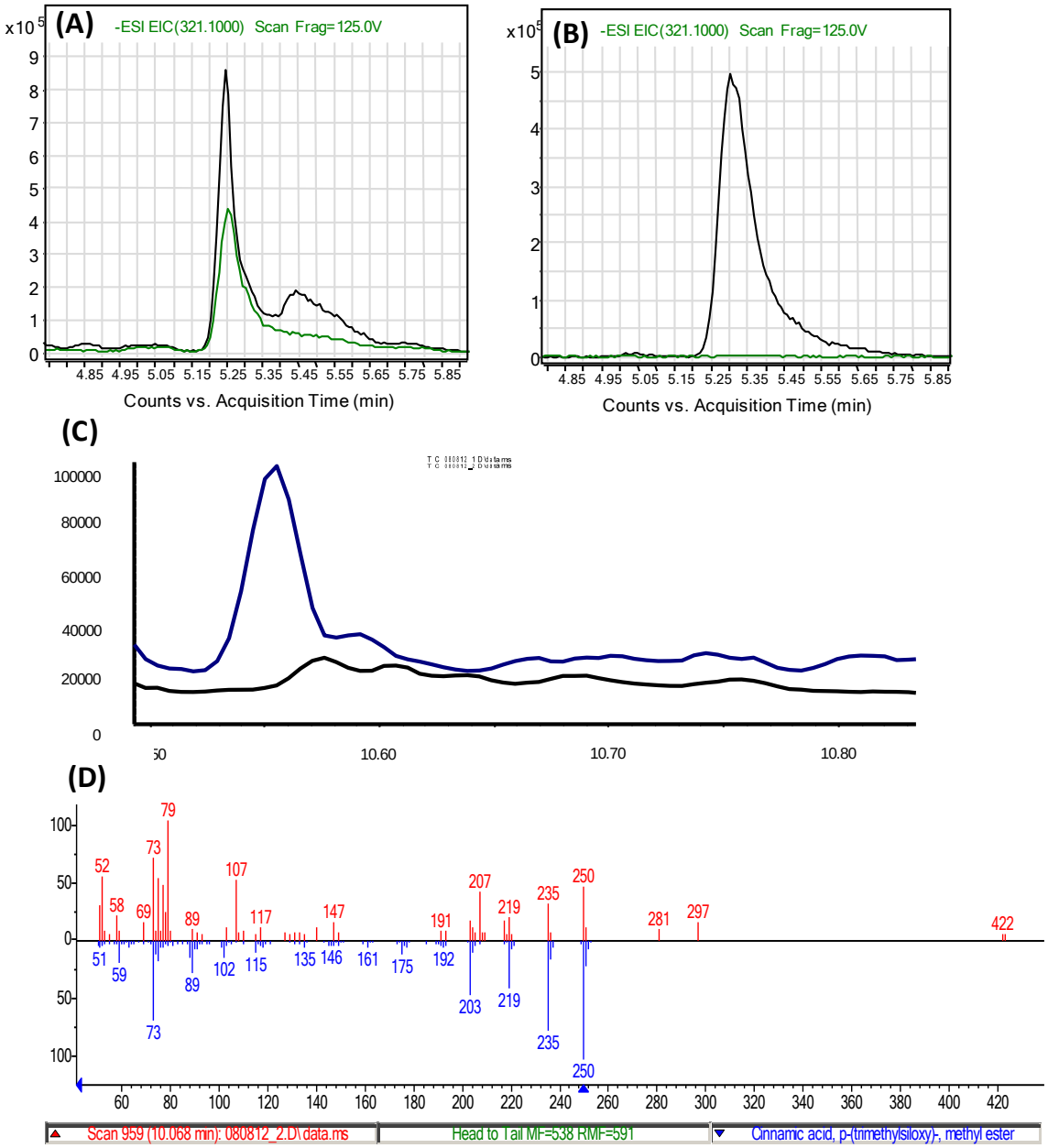


Figure 2-2. Cinnamoyl-quinate confirmation. The target peak (*m/z* 321.098, RT 5.25 min) cannot be digested by β-glucosidase but can be hydrolyzed by 0.1 N NaOH. (A) The extraction ion chromatogram (EIC) showed the target peak (*m/z* 321.098) before (black) and after (green) β-glucosidase treatment. (B) EIC indicated the target peak before (black) and after (green) alkaline hydrolysis. (C) The head to tail fragmentation alignment of cinnamate standard (red) and the library-matched cinnamate peak from the hydrolyzed sample. (D) The quinate peak (blue) was identified in the hydrolyzed sample but not found in control (black).

Table 2-2. Enzyme kinetic properties of recombinant Pta4CL1 and Pta4CL3. The indicated values are the mean ± error range of two protein purification batches, each with three technical replicates.

Enzyme	Substrate	K_m μM	k_{cat} S^{-1}	k_{cat}/K_m
Pta4CL1	**cinnamate**	1080.25±53.60	5.22±0.486	0.0048
	4-coumarate	30.23±8.56	4.15±1.368	0.1373
	caffeate	29.59±6.84	1.60±0.122	0.0541
	ferulate	52.15±2.00	14.99±0.084	0.2874
Pta4CL3	**cinnamate**	887.50±201.38	0.27±0.023	0.0003
	4-coumarate	10.97±3.18	0.19±0.007	0.0173
	caffeate	42.27±3.53	0.13±0.013	0.0031
	ferulate	84.92±4.74	0.15±0.002	0.0018

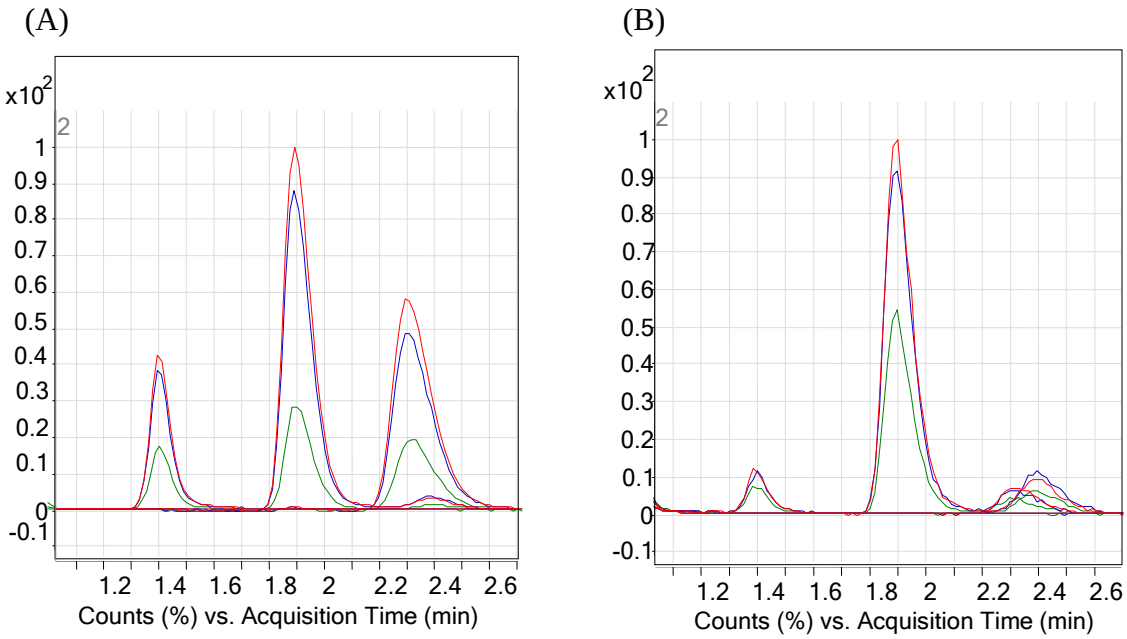


Figure 2-3. Representative chromatograms of CoA thioester products from mixed-substrate enzyme assays with recombinant Pta4CL1 and Pta4CL3. Shown are UV chromatograms (Abs 350 nm) resulting from three-substrate assay (PA+CA+FA) of Pta4CL1 (A) and Pta4CL3 (B) with 100 µM (red), 40 µM (blue), or 10 µM (green) of equal molar substrates.

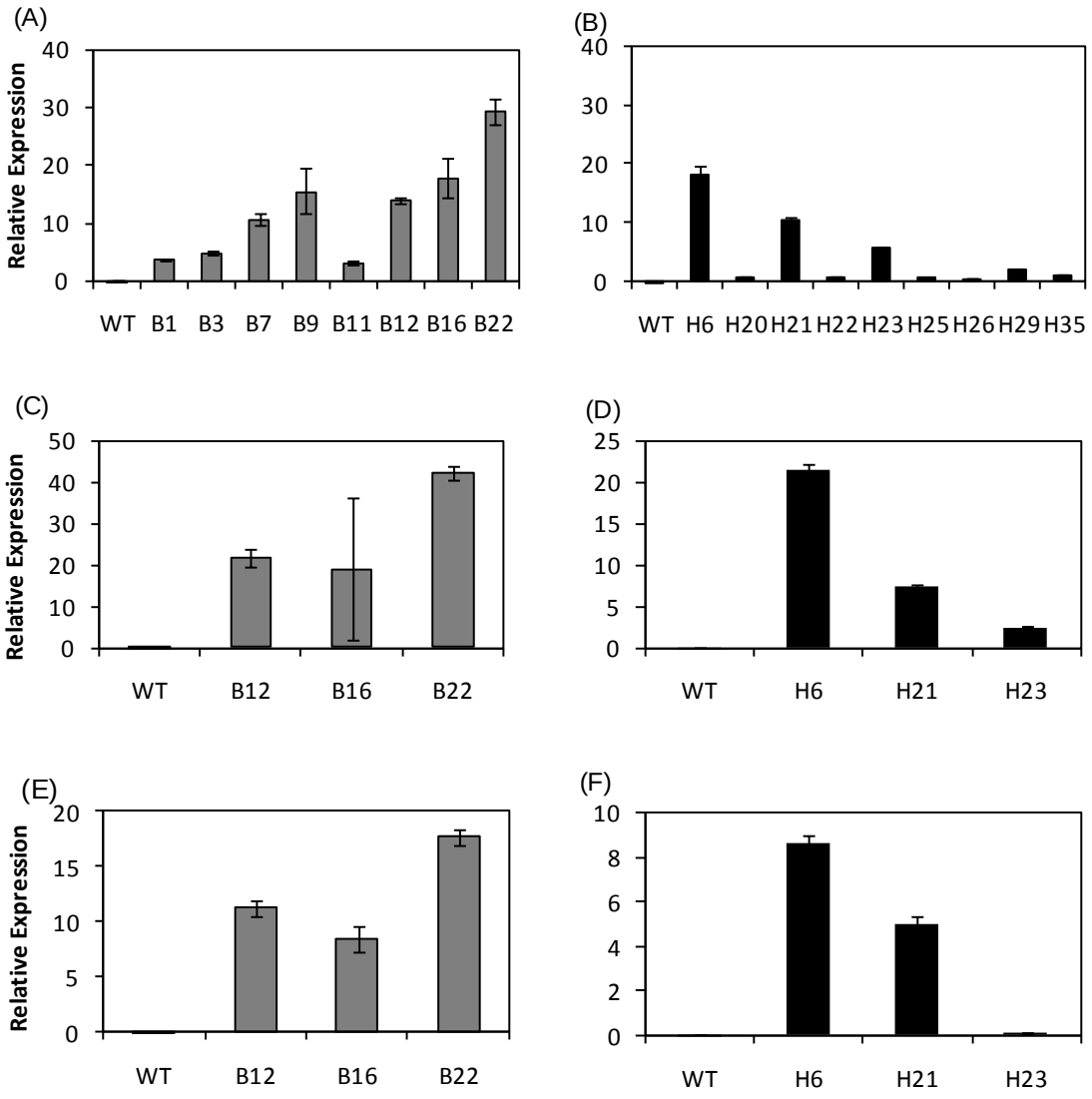


Figure 2-4. Screening of transgenic lines by qRT-PCR. The transcript levels of Pta4CL1 (grey) and Pta4CL3 (black) were analyzed in xylem (A and B), leaf (C and D), and bark (E and F). The error bars indicates the standard deviation of two technical replicates.

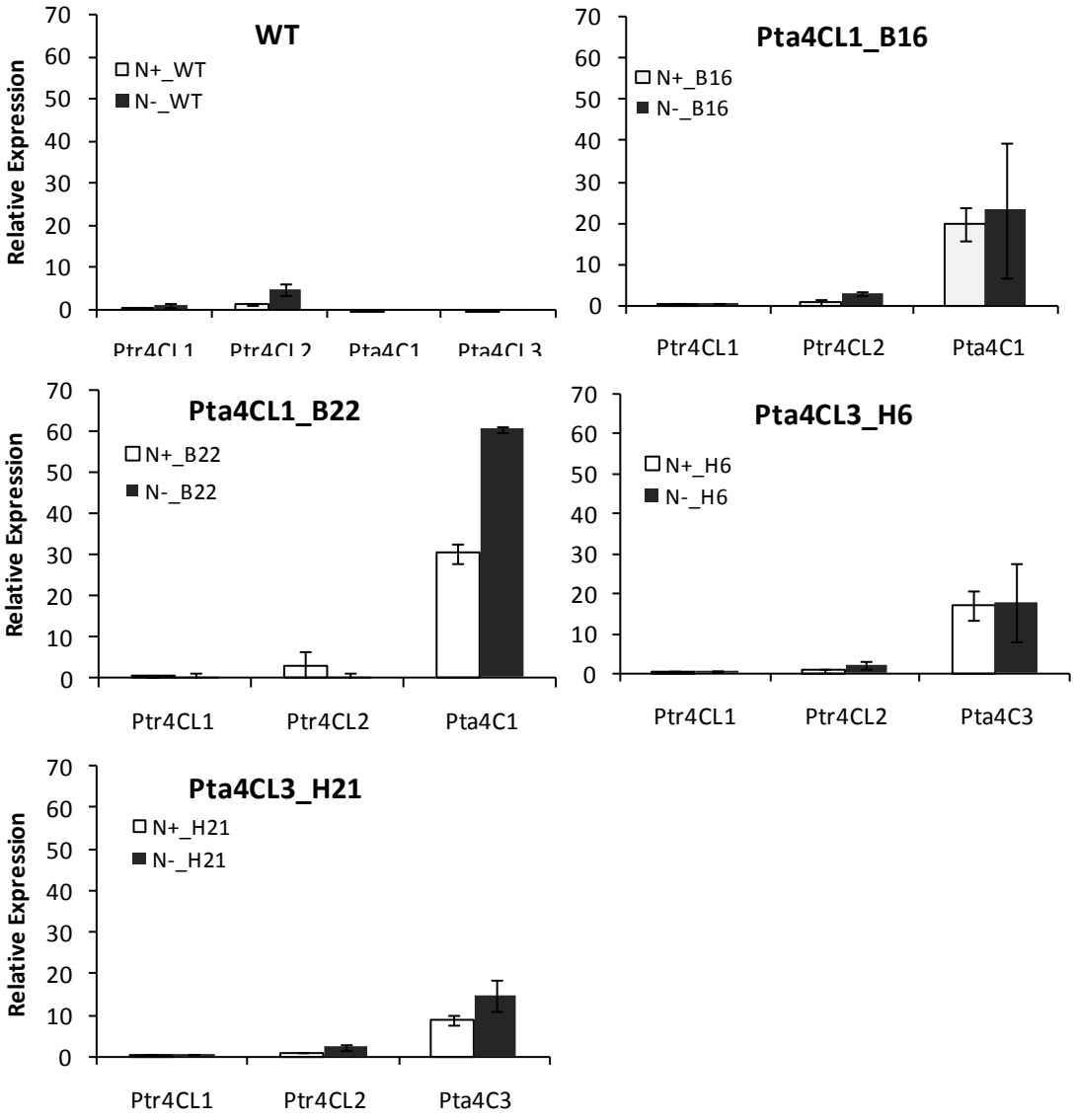


Figure 2-5. The transcription levels of endogenes and transgenes in leaf LPI 5 of wild type and transgenic *Populus*. Two replicates of the N+_B22 group and three replicates of all the other groups were used for the real-time RT-PCR analysis. The expression was normalized to the geometric mean of *elongation factor1-β* (EF1B) and *ubiquitin-conjugating enzyme E2* (UBCc) using the delta-CT method. Error bars indicate the standard deviation of the biological replicates.

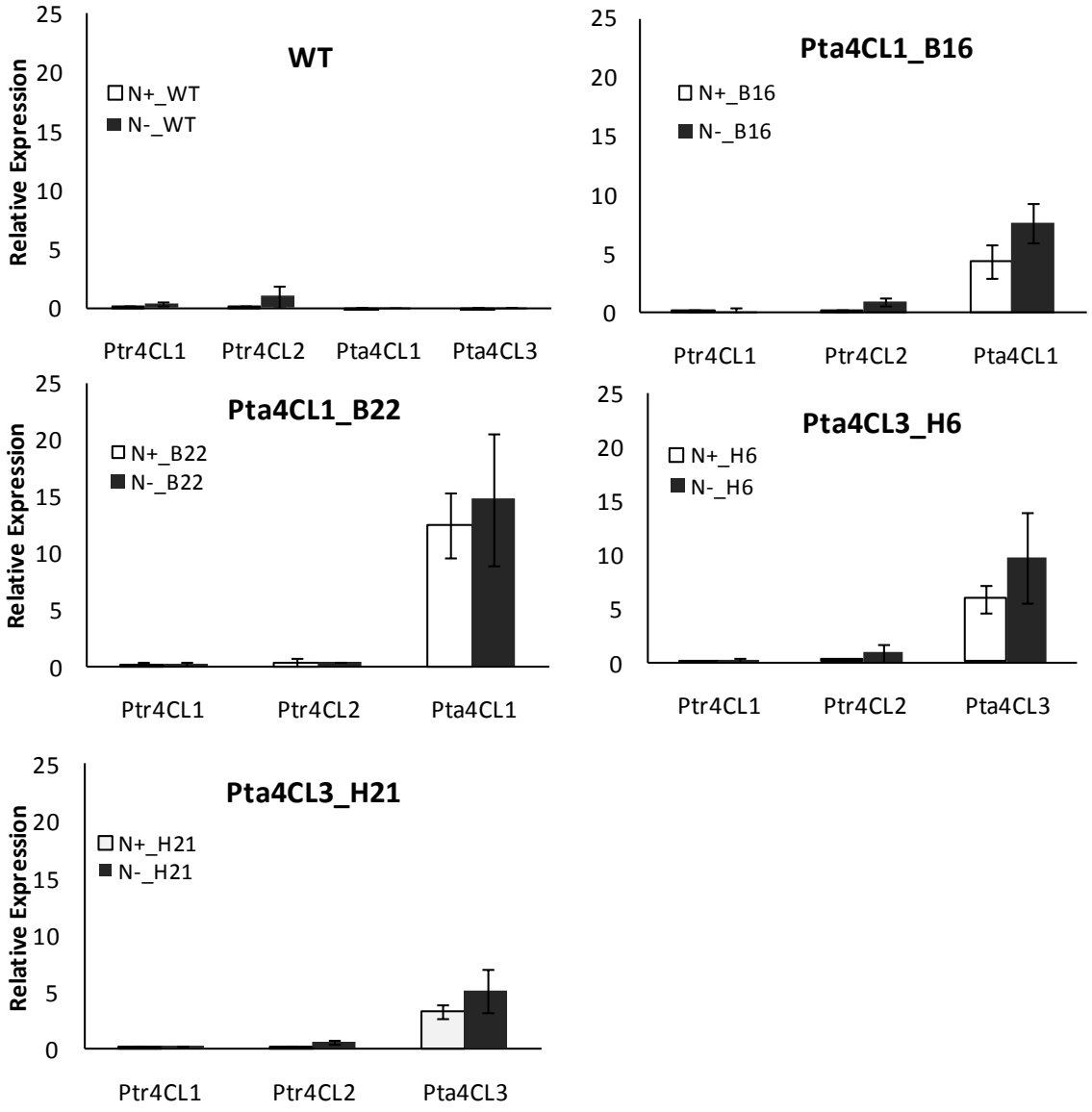


Figure 2-6. The transcription levels of endogenes and transgenes in bark of wild type and transgenic *Populus*. Two replicates of the N+_B22 group and three replicates of all the other groups were used for the real-time RT-PCR analysis. The expression was normalized to the geometric mean of *elongation factor1-β* (EF1B) and *ubiquitin-conjugating enzyme E2* (UBCc) using the delta-CT method. Error bars indicate the standard deviation of the biological replicates.

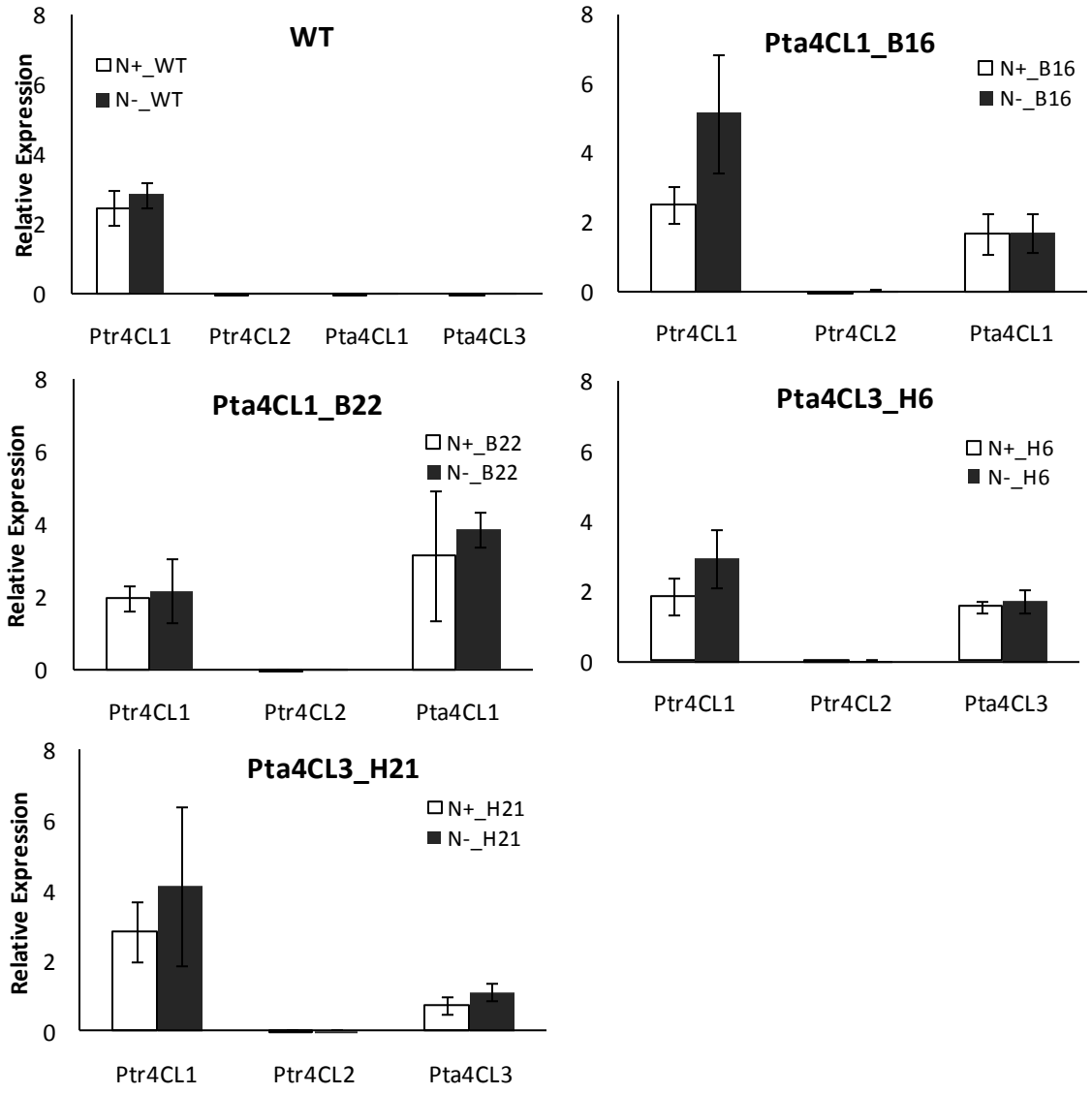


Figure 2-7. The transcription levels of endogenes and transgenes in xylem of wild type and transgenic *Populus*. Two replicates of the N+_B22 group and three replicates of all the other groups were used for the real-time RT-PCR analysis. The expression was normalized to the geometric mean of *elongation factor1-β* (EF1B) and *ubiquitin-conjugating enzyme E2* (UBCc) using the delta-CT method. Error bars indicate the standard deviation of the biological replicates.

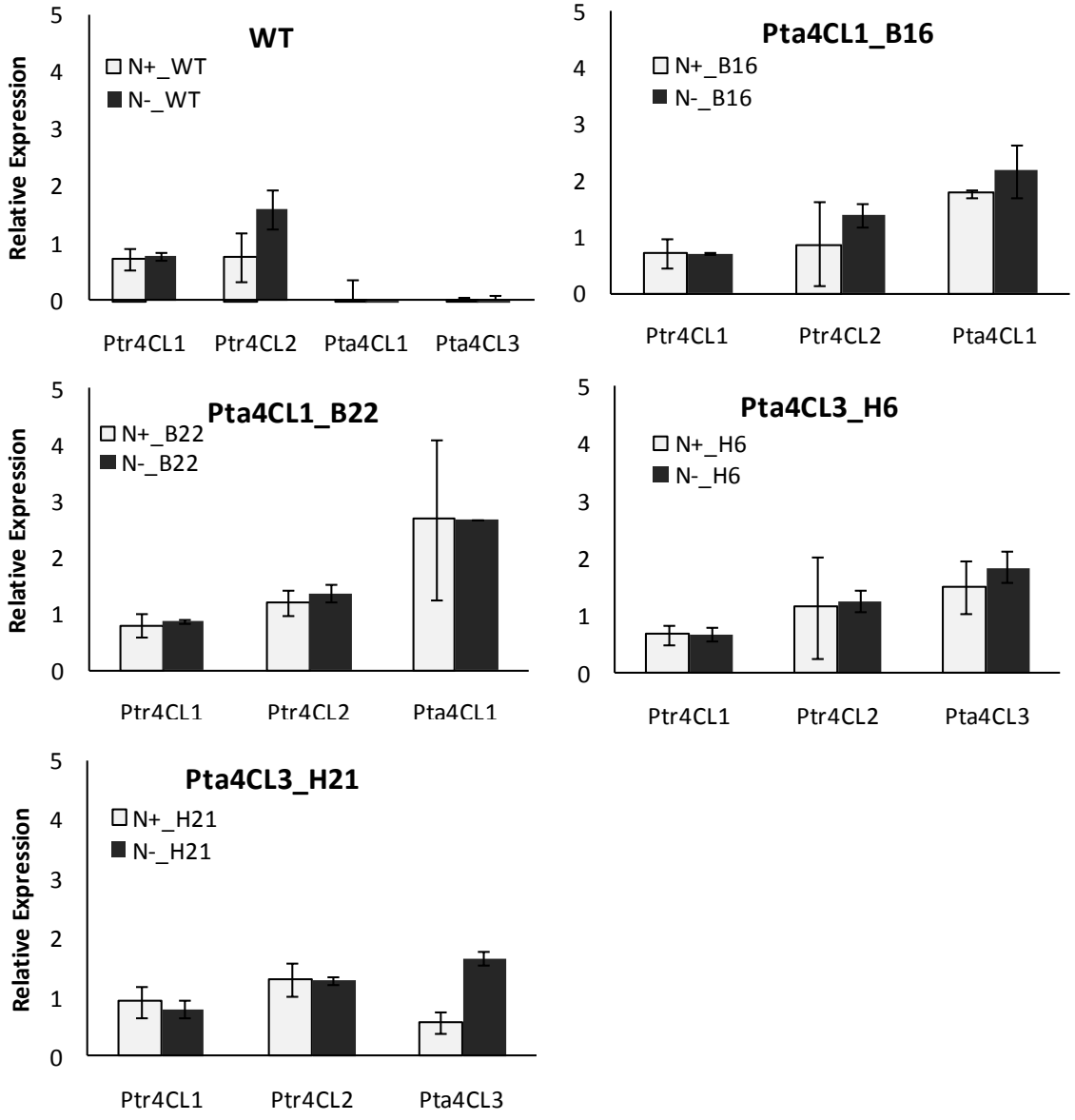


Figure 2-8. The transcription levels of endogenes and transgenes in actively growing roots of wild type and transgenic *Populus*. Two replicates of the N+_B22 group and three replicates of all the other groups were used for the real-time RT-PCR analysis. The expression was normalized to the geometric mean of *elongation factor1-β* (EF1B) and *ubiquitin-conjugating enzyme E2* (UBCc) using the delta-CT method. Error bars indicate the standard deviation of the biological replicates.

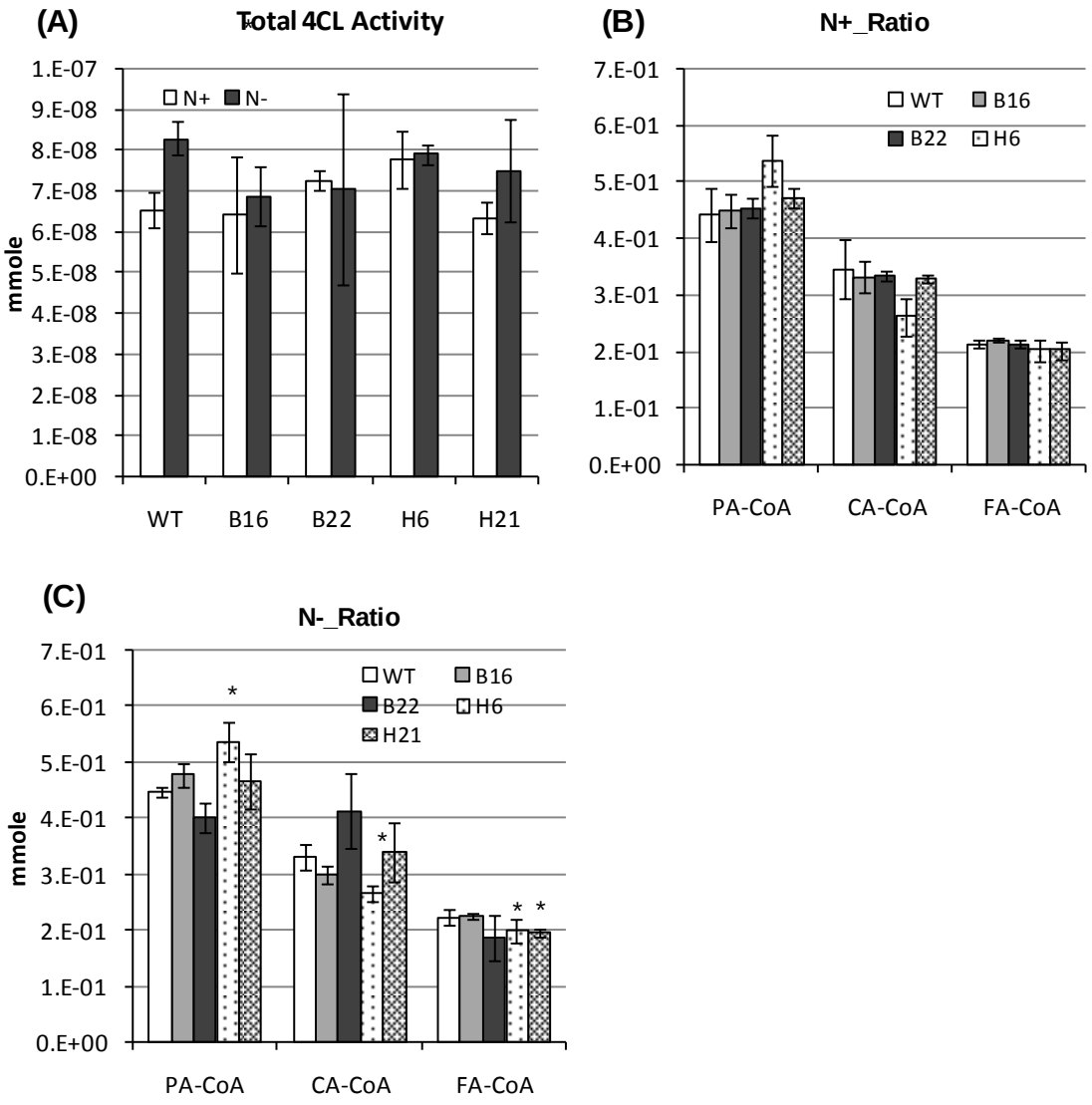


Figure 2-9. 4CL mixed substrate enzyme assays using crude protein extracted from leaf LPI 5. Total 4CL activity (A) was the sum of all hydroxycinnamoyl-CoA in each reaction. The substrate utilization patterns of each transgenic line under N+ (B) and N- (C) were presented by the ratio of the amount of individual product over total product amount. The asterisk marks indicated statistical significance ($p < 0.05$) between the transgenic line and the wild type control.

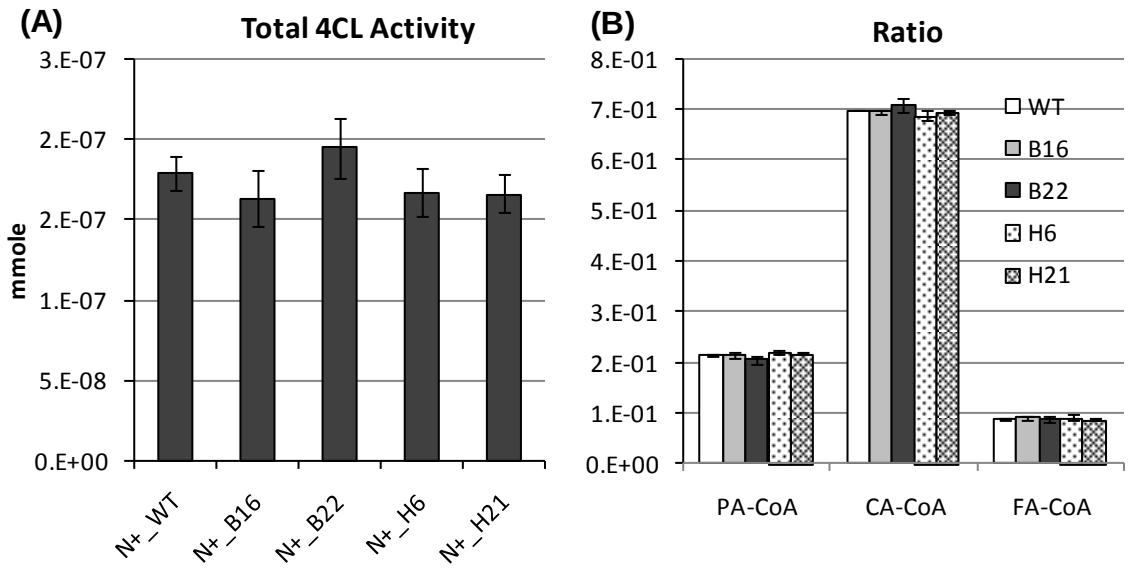
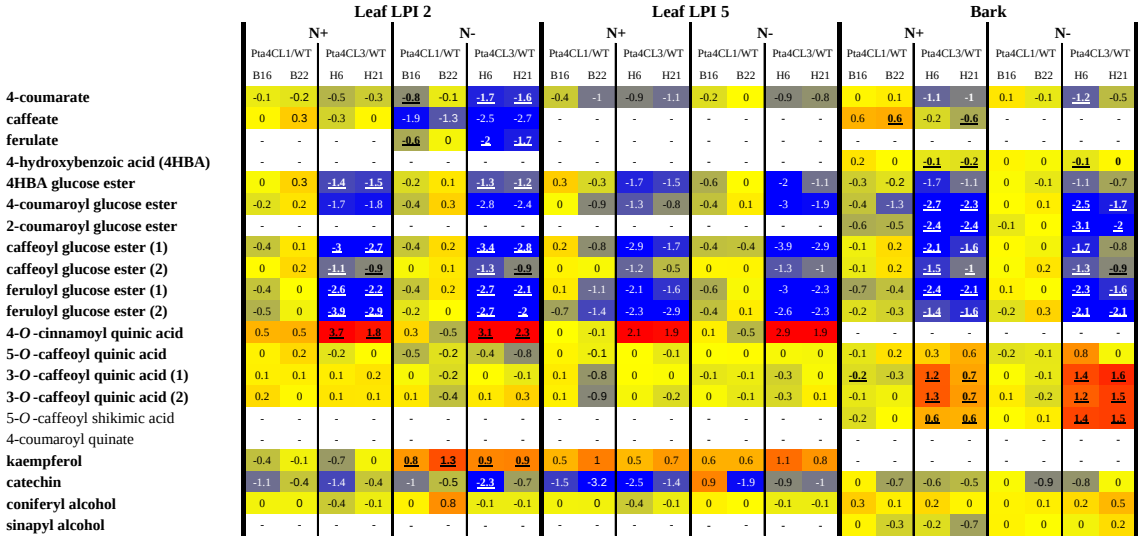


Figure 2-10. 4CL mixed substrate enzyme assays using crude protein extracted from N+ treated xylem. Total 4CL activity (A) was the sum of all hydroxycinnamoyl-CoA in each reaction. The substrate utilization patterns (B) of each transgenic event were presented by the ratio of the amount of individual product over total product.



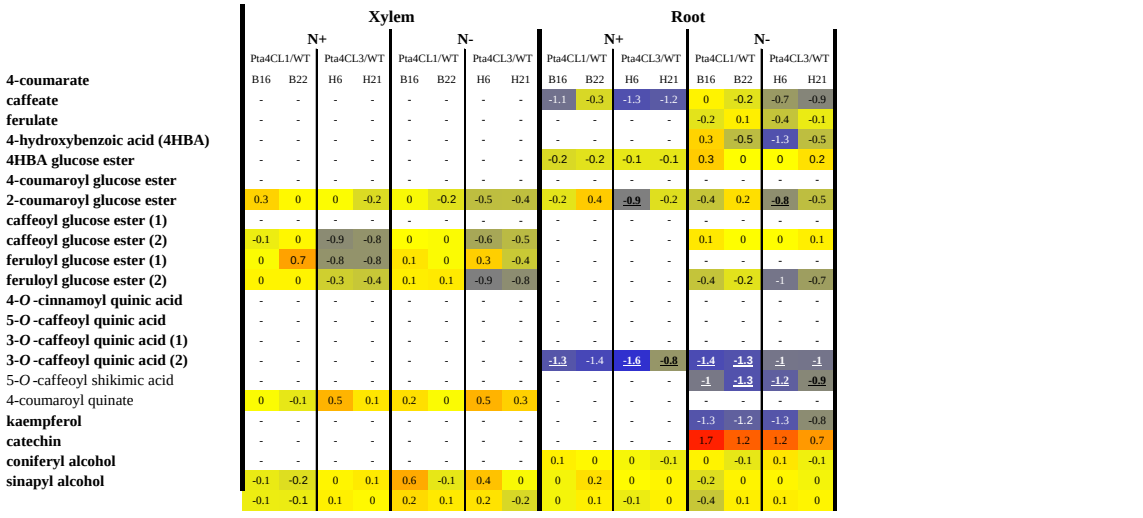


Figure 2-11. The results of *Pta4CL1* and *Pta4CL3* transgenic lines presented in heat maps. Five tissues, including young leaf, mature leaf, bark, root, and xylem were analyzed by LC-MS/TOF and GC-MS. The fold change between transgenic events and wild-type control of each compound was Log 2 transformed. Statistical significance of metabolite level between the transgenic events and control was determined by pair-wise comparison. Bold fonts and underlines indicate significant difference at $p < 0.05$. Metabolite in grey color indicates the predicted results by searching KNApSAcK database using the *m/z* or calculated molecular formula.

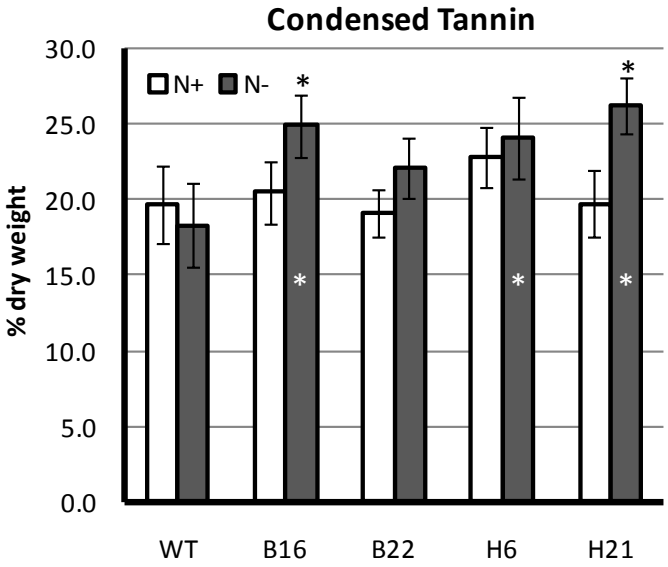


Figure 2-12. Root condensed tannins (CTs) concentrations of each group in percentage of dry weight were shown. Statistical significance ($p < 0.05$) between regular nitrogen (N+) and low nitrogen stress (N-) treatments of each genotypes as determined by t-test and indicated by black asterisks. The white asterisks indicated the significant difference between wild-type and transgenic lines that were under N- treatment.

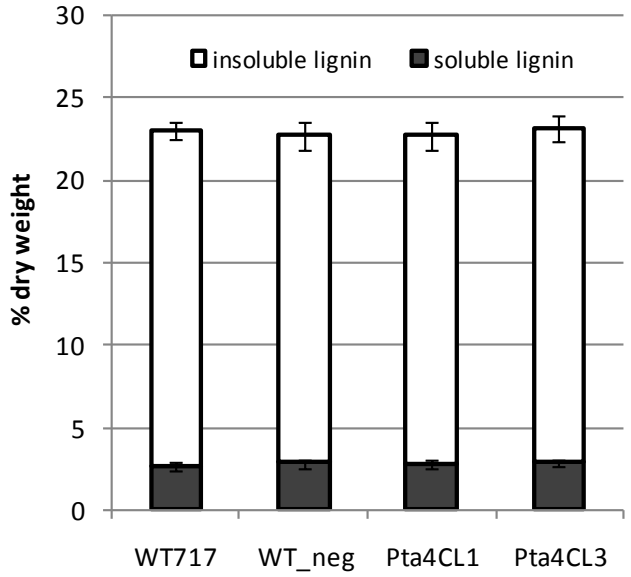


Figure 2-13. The soluble lignin, insoluble lignin, and total lignin content of wild-type (WT717), negative transformant (WT_neg), Pta4CL1, and Pta4CL3 transgenic lines in percent dry weight.

CHAPTER 3

A KEY RESIDUE IN THE SUBSTRATE BINDING POCKET AFFECTS THE SUBSTRATE SELECTIVITY OF 4-COUMARATE:COA LIGASES AND THE PHENYLPROPANOID METABOLISM IN TRANSGENIC *POPULUS*[1]

[1]Chen HY, Mark PY, Zhang D, Wood Z, Harding SA, and Tsai CJ. To be submitted to New Phytologist.

Abstract

4-coumarate:CoA ligase (4CL) mediates the CoA-activation of hydroxycinnamates for their entry into lignin, flavonoid and other phenylpropanoid biosynthetic pathways. Recombinant 4CL isoforms utilize various hydroxycinnamates *in vitro* with distinct substrate specificity. 4CL isoforms from different species are phylogenetically separated into three major clusters, and 4CLs within each cluster generally exhibit similar substrate utilization patterns. Homology modeling was used to aid identification of key residues that determine the substrate selectivity of 4CLs. Among the nine residues predicted to form the hydroxycinnamate binding pocket, position 302 is least conserved. Class I 4CLs typically feature lysine-302, while Classes II and III feature leucine-302 and methionine-302, respectively. Mutagenesis studies confirm that position 302 is a major determining factor of 4CL substrate selectivity *in vitro*. Over-expressing wild-type or mutant 4CLs with substrate preference toward *p*-coumarate (PA) or caffeate (CA) in transgenic *Populus* caused distinct phenylpropanoid metabolite changes, especially in bark. The different transgenic effects suggested that 4CL utilizes CA *in vivo*, which contradicts the current phenylpropanoid pathway model with its emphasis on PA as the sole 4CL substrate. Contrasting effects on metabolism were observed in two transgenic lines that over-express domain swapped 4CL mutants. The mutants share the same substrate binding domain, but in different backbone sequences. The importance of the 4CL backbone to catalytic function was discussed and the findings offer preliminary evidence that the backbone sequence influences 4CL function *in vivo*.

Introduction

The central phenylpropanoid pathway of plants converts the amino acid phenylalanine into a series of C6-C3 hydroxycinnamates (Bravo, 1998). CoA-activated forms of these hydroxycinnamates are utilized for the biosynthesis of higher order protective and structural end-products such as flavonoids and lignin monomers (La Camera et al., 2004; Vogt, 2010). Conversion of hydroxycinnamates into their corresponding CoA thioesters is mediated by multiple isoforms of 4-coumarate:CoA ligase (4CL) with differing hydroxycinnamate substrate specificities (Hu et al., 1998; Ehlting et al., 1999; Lindermayr et al., 2002; Hamberger and Hahlbrock, 2004; Hamberger et al., 2007; Silber et al., 2008; Vogt, 2010; Gui et al., 2011). Based on their amino acid sequences, angiosperm 4CLs are classified into two major groups, Class I and Class II (Ehlting et al., 1999; Lindermayr et al., 2002). Within each class, 4CL isoforms usually have similar gene expression patterns and enzyme kinetic properties (Hu et al., 1998; Ehlting et al., 1999; Lindermayr et al., 2002).

Class II has fewer representatives than Class I, as only a single Class II isoforms generally occurs in the species characterized so far (Hamberger et al., 2007). Class II isoforms are usually expressed most abundantly in green/soft tissues. These tissues are replete with flavonoids which are biosynthesized via conjugation of CoA-activated *p*-coumaric acid (PA) with several units of malonyl-CoA (Ebel et al., 1974; Hahlbrock and Grisebach, 1979). The Class II enzymes are well-suited for a role in flavonoid biosynthesis since they exhibit a stronger substrate preference for PA than do the Class I enzymes (Ehlting et al., 1999; Harding et al., 2002; Lindermayr et al., 2002). Multiple

Class I 4CL isoforms usually occur in angiosperms. Some of these isoforms are highly expressed in lignifying tissues, and they show high *in vitro* affinity toward caffeate. Transgenic studies in various species confirm that Class I 4CLs function in lignin biosynthesis (Kajita et al., 1996; Lee et al., 1997; Hu et al., 1999). Whereas most angiosperm 4CLs fall into Class I and Class II, loblolly pine Pta4CL1 and other conifer 4CL isoforms form a conifer-specific Class III. *Pta4CL1* is well-expressed in lignifying tissues and green/soft tissues (Chen, 2008). It is able to utilize 4-coumarate, caffeate, and ferulate with equal catalytic efficiency (Zhang and Chiang, 1997; Harding et al., 2002). Pta4CL1 may in fact be involved both in the biosynthesis of lignin monomers and of non-lignin phenolics (Voo et al., 1995; Zhang and Chiang, 1997; Chen, 2008). Roles of Class I and III 4CLs in lignin biosynthesis have been investigated *in vivo* using transgenic tobacco, *Arabidopsis*, aspen and radiata pine (Kajita et al., 1996; Lee et al., 1997; Hu et al., 1999; Anterola and Lewis, 2002; Wagner et al., 2009). There are no published reports of Class II 4CLs or their functional characterization in transgenic plants.

4CL catalyzes the ATP-dependent formation of a hydroxycinnamoyl-adenylate intermediate from which the hydroxycinnamoyl moiety is then transferred to the sulfhydryl group of coenzyme A (Fulda et al., 1994; Pietrowska-Borek et al., 2003). This catalytic mechanism is characteristic of adenylate-forming enzymes (Babbitt et al., 1992). The crystal structure of another adenylate-forming enzyme, the phenylalanine-activating domain (PheA) of gramicidin S synthetase, has been resolved (Fulda et al., 1994; Conti et al., 1997). With the available structure information, PheA was used as the early structural model to identify putative 4CL substrate binding residues (Ehlting et al., 2001; Schneider et al., 2003). By a domain swap strategy, the adenylate binding sites of *Arabidopsis*

At4CL were localized to a region flanked by conserved 4CL motifs, Box I and Box II (Ehlting et al., 2001). The corresponding motifs, A3 and A6 in PheA and other peptide synthetases, flank the substrate binding sites (Conti et al., 1997; Ehlting et al., 2001; Schneider et al., 2003). Alignment of the At4CLs and PheA identified 9 residues that line the substrate binding pocket (SBP), including two positions that exhibited variability between At4CL isoforms (Stuible and Kombrink, 2001). Site-directed mutagenesis showed that replacing the bulky residues lysine 320 (K320) and methionine 293 (M293) of At4CL2 with less bulky residues resulted in a gain of activity toward ferulate and a decrease toward caffeate (Stuible and Kombrink, 2001). The crystal structure of *Populus tomentosa* Pto4CL1 in complex with a 4-hydroxycoumaroyl adenylate intermediate mimic has recently been reported (Hu et al., 2010). The structure supports the idea that nine amino acids are involved in hydroxycinnamate binding, though residues at only six positions were predicted correctly in the *Arabidopsis* At4CL2 study (Stuible and Kombrink, 2001). All nine residues can interact with the 4-hydroxyphenyl group of the intermediate mimic by either their main or side chains (Hu et al., 2010). During catalysis, the 4-hydroxyphenyl group of the adenylate-forming intermediate is stabilized by the hydrophobic environment of the substrate binding pocket. The main chain of tyrosine 330 (Y330) and glycine 331 (G331) and the side chain of Y236 are in close proximity within the SBP. While Y236 interacts with the substrate through H-bonding, Y330 and G331 form tight van der Waals interactions with the 4-hydroxyphenyl group of the substrate mimic. The 4-hydroxyl group can H-bond with residue serine 240 (S240) and K303. Among those nine residues of *Populus tomentosa*, K303 corresponds to Ptr4CL1 K302 and At4CL2 K320 (Stuible and Kombrink, 2001; Hu et al., 2010).

For the present work, four new aspen 4CL isoforms, Ptr4CL3 through Ptr4CL6, were identified based on the *Populus* genome sequence. Published crystal structure data from PheA and Pto4CL1 were used to identify putative substrate binding residues of the aspen Ptr4CLs. From sequence alignment and recombinant protein assays, position 302 exhibited the strongest influence on isoform-specific substrate utilization and was targeted for mutagenesis studies. A series of mutant 4CLs was generated for both *in vitro* enzyme assays and *Populus* transformation experiments. Mutations at residue 302 (Ptr4CL1 numbering) altered 4CL substrate preference. Transgenic *Populus* ectopically expressing mutant 4CLs that represent the substrate preferences of Class I and Class II were produced, and transgenic effects were analyzed by metabolic profiling.

Materials and Methods

Recombinant Ptr4CL3, Ptr4CL4 and Ptr4CL5 protein expression in *E. coli*

The coding sequences of *Ptr4CL3*, *Ptr4CL4*, and *Ptr4CL5* were amplified by PCR. Primers (Table 3-1) were designed to introduce a *Kpn I* site at the 5' end without the start codon. The amplified coding sequences were TA-cloned into pCRII and sequenced. The cloned fragment was then excised using *Kpn I* and *Xho I*, subcloned into pET-30a (Novagen), and transformed into *E. coli* strain Rosetta 2 (DE3) pLysS for recombinant protein expression. Recombinant protein induction and purification steps were described in Chapter two.

Sequence alignment and 4CL phylogeny

The amino acid sequences of previously characterized aspen Ptr4CL1-2 (Hu et al., 1998), newly identified Ptr4CL3-6, loblolly pine Pta4CL1 & 3 (Voo et al., 1995; Chen, 2008), *Arabidopsis* At4CL1-4 (Ehlting et al., 1999; Hamberger and Hahlbrock, 2004), soybean Gm4CL1-4 (Lindermayr et al., 2002) and *Populus tomentosa* Pto4CL1 (Hu et al., 2010) were aligned using the ClustalW program (Larkin et al., 2007) to identify the putative substrate binding sites. The sequences of additional 4CL isoforms from various species, including *Medicago truncatula*, *Vitis vinifera*, *Eucalyptus grandis*, *Zea mays*, *Sorghum bicolor*, *Brachypodium distachyon*, *Oryza sativa*, *Pinus radiate*, *Pinus massoniana*, and *Picea sitchensis*, were included for phylogenetic tree construction using MUSCLE (Edgar, 2004), Gblocks (Castresana, 2000), PhyML (Guindon et al., 2010), Jones-Taylor-Thornton matrix (Jones et al., 1992), and approximated likelihood ratio test (aLRT) (Anisimova and Gascuel, 2006), implemented in the Phylogeny.fr server (Dereeper et al., 2008). The workflow was described in Chapter Two.

4CL mutagenesis and recombinant protein expression

4CL expression constructs *pQE-Ptr4CL1* and *pQE-Ptr4CL2* (Hu et al., 1998; Harding et al., 2002) were used for the introduction of point mutations at position 302 (Dongyan Zhang, unpublished work) using primers listed in Table 1. The domain-swap mutants Hyb4CL1-a and Hyb4CL1-b were generated by replacing the putative substrate binding regions of Pta4CL1 and Ptr4CL2, respectively, with the corresponding region of Ptr4CL1 (Mark, 2002). The mutant 4CLs were cloned into pQE protein expression vectors and transformed into *E. coli* host strain M15 (Qiagen).

Enzyme assays

Recombinant 4CL enzyme activities were measured following protocols described in the Materials and Methods of Chapter Two. Single substrate enzyme assays contained cinnamate (CiA), *p*-coumarate (PA), caffeate (CA), or ferulate (FA), and substrate concentrations ranging from 5 to 5000 μM. Mixed-substrate assays included equal-molar PA, CA, and FA, ranging from 5 to 500 μM. For single substrate assays, 100 ng (Ptr4CL4) or 300 ng (other 4CL) of recombinant proteins were used. For mixed-substrate assays, 200 ng (Ptr4CL4) or 400 ng (other 4CL) were used. Except noted below, assay products were analyzed by HPLC-MS/TOF and the kinetic properties were determined as described in Chapter two. Products of the mutant 4CL single-substrate assays were analyzed spectrophotometrically by UV-absorbance at wavelengths 333 nm (PA-CoA), and 346 nm (CA-CoA and FA-CoA), and quantified using extinction coefficients (PA-CoA, 21; CA-CoA, 18; FA-CoA, 19 cm-1mM-1; Gross and Zenk, 1966). K_m and V_{max} values were calculated by fitting the product formation data to Lineweaver-Burk plots. The k_{cat} values were calculated using ($k_{cat} = V_{max}/[E]_T$) where E was the amount of protein (mole) assayed.

Populus transformation and transgenic plant characterization

Hyb4CL1-a, Hyb4CL1-b, and Ptr4CL2 sequences were cloned into the *Xba* I and *Kpn* I restriction sites of binary vector pBI121, downstream of the CaMV 35S promoter (S. Hubscher). The transformation constructs were mobilized into *Agrobacterium tumefaciens* strain C58/pMP90, followed by co-cultivation with *Populus* explants as described in Chapter two. *Populus* transformants were transferred to 1-gal pots

containing potting mix (2:1:1 (v/v) topsoil: peat moss: perlite). Genomic DNA was obtained from leaf tissue (Cheung et al., 1993) for PCR confirmation of positive transformants. Transgene transcript levels were estimated by qRT-PCR of RNA from leaves, xylem and bark using transgene-specific primers (Table 3-1). The two transformation events with the strongest transgene expression for each construct were selected for subsequent experiments. The selected events (transgenic lines) were A3 and A7 for Hyb4CL1-a; E26 and E27 for Hyb4CL1-b; and G8 and G9 for Ptr4CL2. The selected lines were vegetatively propagated by rooted cuttings and maintained in a mist chamber. After the cuttings regenerated, the mist was gradually decreased for the cuttings to acclimate to the normal greenhouse conditions. Ten replicates of A3, A7, and G9; nine replicates of E26 and E27; and four replicates of G8 were obtained for the hydroponic nitrogen stress experiment. Plants were grown in ebb-and-flow hydroponic tubs under greenhouse conditions as described in Harding et al. (2009). Supplemental lighting (400 W metal-halide lamps) was used to ensure a 16 h photoperiod. Plants were grown in aerated nutrient solution containing nitrate- (2 mM) and ammonium-nitrogen (0. 5 mM) which was replaced weekly. After a month long acclimation, the nitrogen level for the N stress treatment group was reduced from 2.5 mM (N+) to 0.125 mM (N-). The treatment was maintained for 14 days. The third youngest fully unfurled leaf, leaf plastochron index 2 (LPI 2), along with bark and xylem between internodes 7 and 12 were harvested, snap-frozen in liquid nitrogen, and stored at -80℃. The tissues were ground to a fine powder in liquid nitrogen. One aliquot of each sample was freeze-dried for metabolic profiling. For lignin analysis, rooted cuttings of each selected transgenic event, wild type controls, and the transgenic negative control (a PCR-negative transformant) were grown to a height of

two meters in 5-gal pots containing soil mixture as described earlier. The stem between internodes 26 and 40 was harvested, and the bark and the developing xylem tissue were removed. The remaining wood was air-dried until use.

Metabolic profiling and lignin analysis

Freeze-dried tissue powders were used for metabolic profiling by HPLC-MS/TOF and GC-MS as described in Material and Methods of Chapter two. For lignin analysis, the previously air-dried wood tissue of pot-grown plants was ground to pass through a 20-mesh screen using a Wiley Mill (Thomas Scientific for the Wiley Mill). Soluble extractives were removed by soxhlet extraction (Extraction Unit E-816, BUCHI) with ethanol for 99 cycles. Klason lignin was analyzed as described in Chapter Two. The syringyl-to-guaiacyl (S/G) lignin ratio was determined using Pyrolysis Molecular Beam Mass Spectrometry (PyMBMS) at the Complex Carbohydrate Research Center (CCRC) of the University of Georgia.

4CL structural homology modeling

The 3D structural coordinates of Pto4CL1 were downloaded from the Protein Data Bank (accession number 3NI2). The automated modeling mode of SWISS-MODEL program (Arnold et al., 2006) was used to create the 3D structural alignment model of Ptr4CL1, Ptr4CL2, and various 4CL mutants. The proposed substrate binding model was created with the help of Dr. Zachary Wood based on the crystal structure information of Pto4CL1 (Hu et al., 2010) and the 4CL mutagenesis results from this study.

Results

Characterization of four 4CL isoforms from *Populus*

Blast search against the *Populus* genome v3 (Phytozome; Goodstein et al.) , using previously characterized *Ptr4CL1* (Potri.001G036900.1) or Ptr4CL2 (Ptori. 019G049500.1; Hu et al., 1998) as query, yielded the following new *4CL*genes: *Ptr4CL3* (Potri.006G169700.1), *Ptr4CL4* (Potri.018G094200.1), *Ptr4CL5* (Potri.003G188500.1), and *Ptr4CL6* (Potri.006G169600.1). *Ptr4CL4* and *Ptr4CL5* are located on chromosomes 18 and 3, respectively. *Ptr4CL3* and *Ptr4CL6* are tandem duplicate on chromosome 6, with a 98.7% identity (99.1% similarity) at the amino acid sequence level. All four new 4CLs clustered phylogenetically with Class I isoforms (Fig. 3-1). *Ptr4CL5* is closely related to previously characterized Ptr4CL1 in Class I-A, while the other three isoforms clustered into a separate subclade, Class I-B. Recombinant proteins of each new 4CL were produced, except Ptr4CL6, which is essentially identical to Ptr4CL3. Among the three recombinant proteins, Ptr4CL5 exhibited the strongest *in vitro* substrate selectivity (Table 3-2). It did not utilize CiA, and it exhibited a K_m differential of more than one order of magnitude with respect to CA, PA and FA. Ptr4C 3 and 4 both utilized CiA and exhibited only a 2-3 fold K_m differential among the other three substrates. Ptr4CL4 was less catalytically discriminating than Ptr4CL5, but both isoforms clearly exhibited substrate preferences; Ptr4CL4 for PA, and Pt4CL5 for CA (Table 3-2). Ptr4CL3 exhibited comparatively weak substrate discrimination, both in terms of K_m and catalytic efficiency (k_{cat}/K_m). These substrate preference patterns were also evident in mixed substrate assays under conditions of equal-molar PA, CA, and FA (Fig. 3-2). Ptr4CL4

exhibited much higher activity rates than the other aspen 4CL isoforms regardless of substrate or assay condition (Table 3-2 and Fig 3-2).

The putative substrate binding sites of Ptr4CL1 and Ptr4CL2

The crystal structure of Pto4CL1 was used as a homology model for identifying the substrate binding residues responsible for 4CL catalysis. Nine putative substrate binding residues were identified and sequence alignment was used to map the corresponding residues in the six aspen Ptr4CLs as well as in 4CL isoforms from *Arabidopsis*, soybean, and loblolly pine (Fig. 3-3). Position 302 (Ptr4CL1 numbering) was found to exhibit the most variation, while the other positions were relatively conserved. Leucine (L) occupies position 302 in all Class II 4CL isoforms, while all Class III members feature methionine (M) at that position. Various residues occupy position 302 in the different Class I isoforms. Ptr4CL1, Ptr4CL5, and other isoforms in Class I-A feature lysine (K) at that position. Class I-B members Ptr4CL3 and Ptr4CL6 have M302, while L occupies position 302 in Ptr4CL4. Position 302 in aspen Ptr4CL1 and Ptr4CL2 isoforms was targeted for 4CL mutagenesis studies for several reasons. The different residues at position 302 appear to segregate with enzyme substrate preference. The L302-containing 4CL isoforms exhibit a preference for PA regardless of class, while pine Pta4CL1 and aspen Ptr4CL3 (M302) exhibit poor substrate selectivity. K302-containing isoforms (Ptr4CL1 and Ptr4CL5) have low K_m toward CA. Gm4CL1 is the only characterized isoform with valine (V) at position 302, and Gm4CL1 exhibits a rarely observed activity toward sinapate (Lindermayr et al., 2002). The homology model and kinetic evidence are consistent with an important role for residue 302 in substrate

utilization. Therefore, a strategy based on overlapping-PCR was utilized to replace K302 of Ptr4CL1 with L, M, or V; while L302 of Ptr4CL2 was replaced with K, M, or V. The mutant 4CLs were cloned into pQE or pET expression vectors for recombinant protein purification. The effects of the mutations on enzyme catalysis were analyzed *in vitro*.

Biochemical characterization of mutant Ptr4CL1 and Ptr4CL2

Single-substrate enzyme assays of the mutant 4CL1 enzymes revealed that K_m values were affected, but k_{cat} values were not (Table 3-3). Replacement of K302 with M (K302M) had little effect on PA or CA catalysis, but replacement with L (K302L) or V (K302V) increased PA and decreased CA catalysis, primarily through differential effects on K_m values for those substrates (Table 3-3). With increased K_m toward CA, the catalytic efficiency of K302L for CA decreased by half relative to the wild-type Ptr4CL1. Unlike Ptr4CL1, the K302L, K302M or K302V mutants were all able to utilize FA as substrate. These single substrate assay results were consistent with the mixed substrate enzyme assays analyzed by LC-MS/TOF (Fig. 3-4). For the mixed substrate enzyme assays, the average CA-: PA-: FA-CoA product ratio of each Ptr4CL1 mutant was 0.2: 1: 0.4 (K302L), 0.5: 1: 0.7 (K302M) and 0.2: 1: 1 (K302V); and the average product ratio of wild type Ptr4CL1 was 4.8: 1: 0.4. Activities of the Ptr4CL2 mutants were low and could not be measured using UV-spectrometry. The activities of Ptr4CL2 L302M and L302V were weak but detectable using HPLC-MS/TOF. In mixed substrate assays, PA was the preferred substrate for both Ptr4CL2 L302M and L302V, similar to wild-type Ptr4CL2 (Fig. 3-5). Four batches of recombinant Ptr4CL2 L302K were purified, and no activity was detected.

Characterization of mutant 4CLs generated by domain swapping

Two domain-swapped proteins, Hyb4CL1-a and Hyb4CL1-b, were designed by replacing the sequence between Box I and Box II of loblolly pine Pta4CL1 and aspen Ptr4CL2, respectively, with that of Ptr4CL1. Both recombinant proteins exhibited a preference for CA over PA, and FA (Table 3-4). However, Hyb4CL1-b exhibited very poor overall activity, and its catalysis of PA and FA was too weak to measure. The mixed substrate assays indicated that both Hyb4CL1-a and Hyb4CL1-b exhibited stronger CoA-esterification of CA than of PA (Fig. 3-6). The product CA-: PA-: FA-CoA ratio of Hyb4CL1-a and Hyb4CL1-b was 2.2: 1: 0.3 and 2: 1: 0.5, respectively. The CA-preference property of these two mutant 4CLs was similar to those of Ptr4CL1 and Ptr4CL5 in Class I, and was not found in other Class of 4CL isoforms.

Ectopic expression of wild-type and mutant 4CLs in transgenic *Populus*

Hyb4CL1(-a and b), and Ptr4CL2 were selected as catalytic representatives of Class I and Class II 4CLs, respectively, for transformation into *Populus* under control of the CaMV 35S constitutive promoter. Transgene expression in xylem (Fig. 3-7), leaf, and bark (Fig. 3-8) of the putative transformants was assessed by qRT-PCR, and two high expressing events of each transgenic group were selected for nitrogen stress experiments. Transcript levels of endogenous *Ptr4CL1* and *Ptr4CL2* were similar in wild type and all transgenic lines (Fig. 3-9 and 3-10). Endogenous *Ptr4CL2* transcripts were more abundant than *Ptr4CL1* transcripts in leaf (Fig. 3-9), while *Ptr4CL1* transcripts predominated in xylem tissue (Fig. 3-10). In both N+ and N- treated leaves, transcript levels of the *Hyb4CL1-a* and *Ptr4CL2* transgenes were 2- to 11-fold higher than the

levels of endogenous *Ptr4CL2*, while the transcript levels of *Hyb4CL1*-b were up to 32-fold higher. In the xylem tissue, endogenous *Ptr4CL1* and the transgenes were expressed at similar levels.

Desalted crude protein extracts were obtained from LPI 5 lamina of N+ plants in order to assay 4CL enzyme activity (Fig. 3-11). Total 4CL activity was significantly increased in the Hyb4CL1-b_E26 transgenic line, trended higher in Hyb4CL1-b_E27 and Ptr4CL2_G8 lines, and did not change in other transgenic lines, compared to wild type. Hydroxycinnamate substrate utilization patterns differed from the wild type in Hyb4CL1-a and Ptr4CL2 transgenic lines but were not altered in Hyb4CL1-b transgenics. Extracts from HYb4CL1-a_A7 and Ptr4CL2_G9 transgenic leaves exhibited stronger activity toward PA and weaker activity toward CA. The PA-CoA/CA-CoA ratio was significantly higher in Hyb4CL1-a_A3, Hyb4CL1-a_A7, and Ptr4CL2_G9 transgenic leaf extracts than in wild type extracts.

Metabolic changes of wild-type and mutant 4CL transgenic *Populus*

Metabolite changes in young leaf, bark and xylem extracts from the various treatment × line combinations were obtained by GC-MS and HPLC-MS/TOF. Genotype and treatment effects on phenylpropanoid metabolism were readily observed (Fig. 3-12). Among transgenic lines, the metabolic response of Hyb4CL1-a was least altered compared to wild types. On the other hand, Hyb4CL1-b and Ptr4CL2 transgenic lines exhibited considerable metabolite changes from the wild type, with the largest differences

occurring in bark. The bark content of caffeoyl-quinate/shikimate esters was higher, and of *p*-coumaric and *o*-coumaric acid glucose esters, lower, in Hyb4CL1-b and Ptr4CL2 transgenic lines, regardless of N-status (Fig. 3-12). The decrease in caffeic acid glucose ester was most severe in the Hyb4CL1-b lines. The reduction of caffeic acid glucose ester levels was also observed in leaf and xylem tissues of Hyb4CL1-b transgenic lines. Catechin was higher in leaf and bark tissues of Hyb4CL1-b and Ptr4CL2 transgenics than in wild types, but only in N+ plants. Fewer soluble phenylpropanoid metabolites were detected in xylem than in leaf and bark tissues. The amounts of *p*-coumaric glucose ester in xylem of Hyb4CL1-b and Ptr4CL2 transgenic lines were higher compared to WT in N+ plants, and lower compared to WT under N- conditions.

Lignin analysis of the transgenic *Populus*

Stem wood from pot-grown wild type and transgenic plants was used for lignin content and syringyl-to-guaiacyl (S/G) lignin ratio analysis. Soluble and insoluble Klason lignin content exhibited no statistical difference between genotypes (Fig 3-13). S/G lignin ratio as determined by Pyrolysis Molecular Beam Mass Spectrometry (PyMBMS) also did not differ between wild type and transgenic lines.

Discussion

New 4CL isoforms revealed correlations between residue 302 and substrate selectivity

A higher resolution phylogenetic analysis of 4CL proteins than before was made possible by inclusion of new genes from *Populus* and various genome-sequenced taxa.

Class I isoforms of *Populus* were divided into subclades Class I-A (*Ptr4CL5* with *Ptr4CL1*) and Class I-B (*Ptr4CL3*, *Ptr4CL4*, and *Ptr4CL6*). Recombinant Ptr4CL5, like its Class I-A paralog Ptr4CL1, exhibited lower K_m toward CA than PA, and featured lysine at position 302 (Hu et al., 1998; Harding et al., 2002; Lindermayr et al., 2002). In contrast, substrate utilization patterns differed among Class I-B 4CLs, although Ptr4CL3/6 and Ptr4CL4 were also derived from whole-genome duplication. The low substrate selectivity of recombinant Ptr4CL3 was unique among all *Populus* isoforms and similar to that of Class III loblolly pine Pta4CL1 (Harding et al., 2002). Although phylogenetically distinct, both Ptr4CL3 and Pta4CL1 feature methionine at the 302 position. Class I-B Ptr4CL4 exhibited a substrate utilization pattern like that of Class II Ptr4CL2, with PA as the preferred substrate (Ehlting et al., 1999; Harding et al., 2002; Lindermayr et al., 2002). Interestingly, like Class II 4CLs, Ptr4CL4 also features a leucine residue at position 302. With these newly characterized 4CLs, Class I features a larger breadth of catalytically distinct 4CL isoforms than previously recognized. The evolutionary implications of this will be interesting to pursue in future studies. The findings also reinforced the notion that 4CL substrate utilization patterns may be strongly conditioned by the amino acid residue at position 302.

Proposed model highlighting the role of residue 302 in substrate binding

Based on the crystal structure of Pto4CL1, the orientation of Ptr4CL1 substrate binding pocket (SBP) residues K302, S239, and Q209 would permit hydrogen bonding with the 4-hydroxyl group of PA (Fig. 12A; Hu et al., 2010). K302 is the least conserved among the three residues, and H-bonding across its 3.52 Å distance from the 4-hydroxyl

group of the substrate would be weak (Hu et al., 2010; Saballos et al., 2012). K302 was therefore predicted not to be essential for substrate positioning in the SBP but to be able to influence substrate selectivity (Hu et al., 2010; Saballos et al., 2012). Our findings are consistent with the hypothesis that substrate-specificity is indeed conferred by the residue occupying position 302. In the Pto4CL1 structural model, a cavity near S239 at one end of the SBP would accommodate the 4-hydroxyl group of the hydroxycinnamate substrate, and in the case of CA, would facilitate H-bonding between the 4-hydroxyl group and S239, and between the 3-hydroxyl group and K302 (Fig. 3-14B). With two H-bonding interactions, CA would be more stabilized than PA in the SBP of Ptr4CL1. Replacement of K302 with L or V (Fig. 3-14C & E) would eliminate the H-bonding between residue 302 and the 3-hydroxyl group of CA. Consistent with expectation, Ptr4CL1 K302L and K302V mutations decreased CA catalysis while reducing K_m with PA as the substrate. Like L and V, M at position 302 also cannot H-bond with CA (Fig. 3-14D). However, the flexible side chain of M may allow space for a water molecule to enter and form a stabilizing H-bond with the 3-hydroxyl group of CA in the SBP. A large hydrophobic cavity close to residue 302 may accommodate the M side chain, and permit such movement. This M302 model offers a scenario consistent with the low substrate selectivity observed for Ptr4CL1-K302M (Table 3). Similar substrate utilization patterns were also observed in Pta4CL1 and Ptr4CL3, both of which feature M302. The least preferred substrate of Ptr4CL1, FA, carries a bulkier 3-methoxyl group that would be expected to encounter steric interference in the cavity near K302. The lowest K_m and the highest catalytic efficiency toward FA among the three Ptr4CL1 mutants were exhibited in the K302V mutant. The side chain of V is small, and compared to K, L, and M, would

be expected to offer minimal steric interference to SBP accommodation of FA. The flexibility of the side chain of M may negate steric interference from that bulky side chain of FA and enable K302M catalysis.

Factors other than residue 302 that affect 4CL substrate selectivity

In contrast to the Ptr4CL1 mutants, the L302M and L302V mutants of Ptr4CL2 exhibited no change in substrate selectivity, indicating that other factors are also involved. The hydroxycinnamate SBP in 4CL is mainly formed by hydrophobic groups, and hydrophobic interaction other than H-bonding is the largest contributor to substrate binding (Hu et al., 2010). Therefore, SBP size in addition to H-bonding could affect substrate selectivity (Schneider et al., 2003; Hu et al., 2010). In addition, our alignment revealed a potentially important difference at position 235 when comparing Ptr4CL1 and Ptr4CL2. This position was occupied by Y in Ptr4CL1 and Pto4CL1, and by phenylalanine (F) in Ptr4L2. Conversion of Y235 of Pto4CL1 to Y235F led to a decrease in activity toward CA, while its activity toward PA was not changed (Hu et al., 2010). It was proposed that F235 made the binding of bulky substrates less thermodynamically favorable. F235 disrupted the hydrogen bond between Y235 and V276 which could destabilize (or simply weaken) substrate binding (Hu et al., 2010). This may explain the inability of Ptr4CL2 L302M and L302V mutants to utilize bulkier (CA and FA) substrates than PA.

4CL backbone may affect catalysis and *in vivo* function

4CL product PA-CoA/CA-CoA ratios were higher in crude protein assays of leaf extracts from Hyb4CL1-a transgenic lines than wild-type or Hyb4CL1-b transgenic lines. This enhancement of PA-CoA activation in Hyb4CL1-a transgenic lines was not expected based on *in vitro* assays of recombinant Hyb4CL1-a, and could imply an interaction of some kind in the crude extracts that preferentially facilitates the CoA-activation of PA. Interestingly in view of the crude protein *in vitro* assay results, metabolic profiles were significantly changed in Hyb4CL1-b transgenics, but not Hyb4CL1-a plants relative to wild-type. Specifically, there was a reduction of caffeoyl glucose ester levels and an increase of caffeoyl-quinate levels in bark of Hyb4CL1-b transgenic lines. The biosynthesis of caffeoyl glucose ester and caffeoyl-quinate are thought to be mediated by glucose transferases (Mock and Strack, 1993) and 4CL (Niggeweg et al., 2004; Sonnante et al., 2010), respectively. Therefore, the metabolic changes in Hyb4CL1-b transgenics may be attributed to the relatively stronger Hyb4CL1-b affinity toward CA than PA. However, this is not a strong argument in light of the very poor *in vitro* activity of Hyb4CL1-b. Other hypotheses invoking pathway competition and metabolon interactions may have to be tested in order to resolve the finding.

It is possible that the backbone structure itself may contribute to protein-protein interactions with other phenylpropanoid enzymes to form metabolons that might influence metabolic outcomes. To assess this possibility, 3D models of the 4CLs were generated by SWISS-MODEL (Arnold et al., 2006) based on the crystal structure of Pto4CL1 (Hu et al., 2010). Alignment of the Ptr4CL2 and Pta4CL1 3D structures showed

that the major difference lies in the N-terminal region where Ptr4CL2 features a large loop. Such structures can be involved in protein-protein interactions and may vary greatly between isoforms and species (Dunker et al., 2008; Bellay et al., 2011). Several enzymes in the phenylpropanoid pathway, including phenylalanine ammonia-lyase (PAL), cinnamic acid 4-hydroxylase (C4H), and cinnamic acid 3-hydroxylase (C3H), were found to form protein complexes (Achnine et al., 2004; Chen et al., 2011). A recent study showed that 4CL can interact with C3H (Bassard et al., 2012). In the presence of C3H, cytosolic 4CL was relocated toward endoplasmic reticulum (ER). Thus, 4CL isoforms may interact with specific isoforms of other phenylpropanoid to form metabolons with specific roles depending on tissue type or plant condition. Metabolon interactions with altered 4CLs could conceivably result in activity losses which could complicate interpretation of metabolic effects such as those observed in the case of Hyb4CL1-b.

Specific metabolon interactions may also be considered in order to understand the expression ratios in shoots and roots of two closely related Class I isoforms. Ptr4CL5 can CoA-activate CA with high efficiency *in vitro* as has been reported for Class I 4CLs that are already known to be associated with lignin biosynthesis (Lee and Douglas, 1996; Ehlting et al., 1999; Harding et al., 2002; Lindermayr et al., 2002). However, unlike those 4CLs, *Ptr4CL5* is predominantly expressed in root (Anino, 2008), where biosynthesis of products besides lignin but including suberin is important (Soler et al., 2007; Chen et al., 2013). Whether Ptr4CL5 has a different *in vivo* function than Ptr4CL1 remains unclear. In light of the growing number of reports on phenylpropanoid metabolons for compartmentalizing 4CL functions (Achnine et al., 2004; Chen et al., 2011; Bassard et al., 2012), it is possible that Ptr4CL1 and Ptr4CL5 in roots have metabolically distinct roles.

Summary

While Class I 4CL appears to contain a number of functionally distinct members, all Class II isoforms are thought to be involved in non-lignin phenylpropanoid biosynthesis and to prefer PA (Hu et al., 1998; Ehlting et al., 1999; Harding et al., 2002; Lindermayr et al., 2002). These observations are consistent with the idea of a correlation between 4CL function/role *in vivo* and substrate preference. Residue 302 in 4CL has been considered to be non-essential for substrate binding (Hu et al., 2010) but was suggested to be responsible for 4CL substrate specificity by Saballos et al. (2012). Results from this study support the argument that residue 302 is involved in 4CL substrate selectivity. Over-expressing Class I-A-like and Class II 4CL isoforms in *Populus* led to metabolic profile changes that reflected the importance of 4CL substrate selectivity to *in vivo* function. The association of 4CL with C3H was reported recently by Bassard (2012). Findings from the present investigation suggested that structural features of 4CL may contribute to its participation in phenylpropanoid metabolic channeling. Findings from transgenic *Populus* harboring both pine and domain-swapped 4CLs in the current study were consistent with effects of 4CL protein structure on protein-protein interactions that modulate its function *in vivo*.

References

Achnine L, Blancaflor EB, Rasmussen S, Dixon RA (2004) Colocalization of L-phenylalanine ammonia-lyase and cinnarnate 4-hydroxylase for metabolic channeling in phenylpropanoid biosynthesis. Plant Cell **16:** 3098-3109

Anino EO (2008) Characterization of Populus tremuloides Caffeic acid O-methyltransferase and 4-coumarate:CoA ligase gene promoters to identify regulatory elements. Michigan Technology University **Ph. D. dissertation**

Anisimova M, Gascuel O (2006) Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. Systematic Biology **55:** 539-552

Anterola AM, Lewis NG (2002) Trends in lignin modification: a comprehensive analysis of the effects of genetic manipulations/mutations on lignification and vascular integrity. Phytochemistry **61:** 221-294

Arnold K, Bordoli L, Kopp Jr, Schwede T (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics **22:** 195-201

Babbitt PC, Kenyon GL, Martin BM, Charest H, Slyvestre M, Scholten JD, Chang KH, Liang PH, Dunawaymariano D (1992) Ancestry of the 4-chlorobenzoate dehalogenase - analysis of amino acid sequence identities among families of acyl-adenyl ligase, enoyl-CoA hydratases isomerase, and acyl-CoA thioesteraes. Biochemistry **31:** 5594-5604

Bassard JE, Richert L, Geerinck J, Renault H, Duval F, Ullmann P, Schmitt M, Meyer E, Mutterer J, Boerjan W, De Jaeger G, Mely Y, Goossens A, Werck-Reichhart D (2012) Protein-protein and protein-membrane associations in the lignin pathway. Plant Cell **24:** 4465-4482

Bellay J, Michaut M, Kim T, Han S, Colak R, Myers CL, Kim PM (2011) An omics perspective of protein disorder. Molecular BioSystems **8:** 185-193

Bravo L (1998) Polyphenols: chemistry, dietary sources, metabolism, and nutritional significance. Nutrition Reviews **56:** 317-333

Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Molecular Biology and Evolution **17:** 540-552

Chen H-C, Li Q, Shuford CM, Liu J, Muddiman DC, Sederoff RR, Chiang VL (2011) Membrane protein complexes catalyze both 4- and 3-hydroxylation of cinnamic acid derivatives in monolignol biosynthesis. Proceedings of the National Academy of Sciences **108:** 21253-21258

Chen H-C, Song J, Williams CM, Shuford CM, Liu J, Wang JP, Li Q, Shi R, Gokce E, Ducoste J, Muddiman DC, Sederoff RR, Chiang VL (2013) Monolignol pathway 4-coumaric acid:coenzyme A ligases in Populus. trichocarpa: Novel specificity, metabolic regulation, and simulation of coenzyme A ligation fluxes. Plant Physiology **161:** 1501-1516

Chen H-Y (2008) Characterization of two distinct 4-coumarate:CoA ligase (4CL) genes in loblolly pine (*Pinus taeda*). Unpublished master thesis

Cheung WY, Hubert N, Landry BS (1993) A simple and rapid DNA microextraction method for plant, animal, and insect suitable for RAPD and other PCR analyses. Pcr-Methods and Applications **3:** 69-70

Conti E, Stachelhaus T, Marahiel MA, Brick P (1997) Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. Embo Journal **16:** 4174-4183

Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard JF, Guindon S, Lefort V, Lescot M, Claverie JM, Gascuel O (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Research **36:** W465-W469

Dunker AK, Silman I, Uversky VN, Sussman JL (2008) Function and structure of inherently disordered proteins. Current Opinion in Structural Biology **18:** 756-764

Ebel J, Schaller B, Knobloch KH, Wellman E, Grisebach H, Hahlbrock K (1974) Coordinated changes in enzyme-activities of phenylpropanoid metabolism during growth of soybean cell-suspension cultures. Biochimica Et Biophysica Acta **362:** 417-424

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research **32:** 1792-1797

Ehlting J, Buttner D, Wang Q, Douglas CJ, Somssich IE, Kombrink E (1999) Three 4-coumarate : coenzyme A ligases in Arabidopsis thaliana represent two evolutionarily divergent classes in angiosperms. Plant Journal **19:** 9-20

Ehlting J, Shin JJK, Douglas CJ (2001) Identification of 4-coumarate : coenzyme A ligase (4CL) substrate recognition domains. Plant Journal **27:** 455-465

Fulda M, Heinz E, Wolter FP (1994) The fadD gene of Escherichia coli K12 is located closeto rnd at 39.6 min of the cgromosomal map and is a new member of the AMP-binding protein family. Molecular & General Genetics **242:** 241-249

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS (2011) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Research **40:** D1178-D1186

Gross GG, Zenk MH (1966) Darstellung und eigenschaften von coenzym A-thioestern substituierter Zimtsäuren. Z. Naturforsch **21b:** 683

Gui JS, Shen JH, Li LG (2011) Functional characterization of evolutionarily divergent 4-coumarate: coenzyme A ligases in rice. Plant Physiology **157:** 574-586

Guindon Sp, Dufayard J-Fo, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. Systematic Biology **59:** 307-321

Hahlbrock K, Grisebach H (1979) Enzymic controls in the biosynthesis of lignin and flavonoids. Annual Review of Plant Physiology and Plant Molecular Biology **30:** 105-130

Hamberger B, Hahlbrock K (2004) The 4-coumarate : CoA ligase gene family in Arabidopsis thaliana comprises one rare, sinapate-activating and three commonly occurring isoenzymes. Proceedings of the National Academy of Sciences of the United States of America **101:** 2209-2214

Hamberger Br, Ellis M, Friedmann M, de Azevedo Souza C, Barbazuk B, Douglas CJ (2007) Genome-wide analyses of phenylpropanoid-related genes in Populus trichocarpa, Arabidopsis thaliana, and Oryza sativa: the Populus lignin toolbox and conservation and diversification of angiosperm gene familiesThis article is one of a selection of papers published in the Special Issue on Poplar Research in Canada. Canadian Journal of Botany **85:** 1182-1201

Harding SA, Jarvie MM, Lindroth RL, Tsai CJ (2009) A comparative analysis of phenylpropanoid metabolism, N utilization, and carbon partitioning in fast- and slow-growing Populus hybrid clones. Journal of Experimental Botany **60:** 3443-3452

Harding SA, Leshkevich J, Chiang VL, Tsai C-J (2002) Differential substrate inhibition couples kinetically distinct 4-coumarate:coenzyme A ligases with spatially distinct metabolic roles in quaking aspen. Plant Physiol. **128:** 428-438

Hu WJ, Harding SA, Lung J, Popko JL, Ralph J, Stokke DD, Tsai CJ, Chiang VL (1999) Repression of lignin biosynthesis promotes cellulose accumulation and growth in transgenic trees. Nature Biotechnology **17:** 808-812

Hu WJ, Kawaoka A, Tsai CJ, Lung JH, Osakabe K, Ebinuma H, Chiang VL (1998) Compartmentalized expression of two structurally and functionally distinct 4-coumarate : CoA ligase genes in aspen (Populus tremuloides). Proceedings of the National Academy of Sciences of the United States of America **95:** 5407-5412

Hu Y, Gai Y, Yin L, Wang X, Feng C, Feng L, Li D, Jiang X-N, Wang D-C (2010) Crystal structures of a Populus tomentosa 4-coumarate:CoA ligase shed light on its enzymatic mechanisms. Plant Cell **22:** 3093-3104

Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. Computer Applications in the Biosciences **8:** 275-282

Kajita S, Katayama Y, Omori S (1996) Alterations in the biosynthesis of lignin in transgenic plants with chimeric genes for 4-coumarate:coenzyme a ligase. Plant and Cell Physiology **37:** 957-965

La Camera S, Gouzerh G, Dhondt S, Hoffmann L, Fritig B, Legrand M, Heitz T (2004) Metabolic reprogramming in plant innate immunity: the contributions of phenylpropanoid and oxylipin pathways. Immunological Reviews **198:** 267-284

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. Bioinformatics **23:** 2947-2948

Lee D, Douglas CJ (1996) Two divergent members of a tobacco 4-coumarate:coenzyme A ligase (4CL) gene family - cDNA structure, gene inheritance and expression, and properties of recombinant proteins. Plant Physiology **112:** 193-205

Lee D, Meyer K, Chapple C, Douglas CJ (1997) Antisense suppression of 4-coumarate:coenzyme A ligase activity in Arabidopsis leads to altered lignin subunit composition. The Plant Cell **9:** 1985-1998

Lindermayr C, Mollers B, Fliegmann J, Uhlmann A, Lottspeich F, Meimberg H, Ebel J (2002) Divergent members of a soybean (Glycine max L.) 4-coumarate : coenzyme A ligase gene family - Primary structures, catalytic properties, and differential expression. European Journal of Biochemistry **269:** 1304-1315

Mark PY (2002) Identification of amino acid sequences that determine substrate utilization by 4-coumarate:CoA ligase (4CL) proteins in aspen (Populus tremuloides Michx.) M.S. Thesis. Michigan Technological University, Houghton, MI

Pietrowska-Borek M, Stuible HP, Kombrink E, Guranowski A (2003) 4-Coumarate : coenzyme A ligase has the catalytic capacity to synthesize and reuse various (di)adenosine polyphosphates. Plant Physiology **131:** 1401-1410

Saballos A, Sattler SE, Sanchez E, Foster TP, Xin Z, Kang C, Pedersen JF, Vermerris W (2012) Brown midrib2 (Bmr2) encodes the major 4-coumarate:coenzyme A ligase involved in lignin biosynthesis in sorghum (Sorghum bicolor (L.) Moench). The Plant Journal **70:** 818-830

Schneider K, Hovel K, Witzel K, Hamberger B, Schomburg D, Kombrink E, Stuible HP (2003) The substrate specificity-determining amino acid code of 4-coumarate : CoA ligase. Proceedings of the National Academy of Sciences of the United States of America **100:** 8601-8606

Silber MV, Meimberg H, Ebel J (2008) Identification of a 4-coumarate:CoA ligase gene family in the moss, Physcomitrella patens. Phytochemistry **69:** 2449-2456

Soler Ma, Serra O, Molinas M, Huguet G, Fluch S, Figueras M (2007) A genomic approach to suberin biosynthesis and cork differentiation. Plant Physiology **144:** 419-431

Stuible HP, Kombrink E (2001) Identification of the substrate specificity-conferring amino acid residues of 4-coumarate : coenzyme A ligase allows the rational design of mutant enzymes with new catalytic properties. Journal of Biological Chemistry **276:** 26893-26897

Vogt T (2010) Phenylpropanoid Biosynthesis. Molecular Plant **3:** 2-20

Voo KS, Whetten RW, O'Malley DM, Sederoff RR (1995) 4-coumarate:coenzyme A ligase from loblolly pine xylem (isolation, characterization, and complementary DNA cloning). Plant Physiol. **108:** 85-97

Wagner A, Donaldson L, Kim H, Phillips L, Flint H, Steward D, Torr K, Koch G, Schmitt U, Ralph J (2009) Suppression of 4-Coumarate-CoA Ligase in the Coniferous Gymnosperm Pinus radiata. Plant Physiol. **149:** 370-383

Zhang XH, Chiang VL (1997) Molecular cloning of 4-coumarate:coenzyme A ligase in loblolly pine and the roles of this enzyme in the biosynthesis of lignin in compression wood. Plant Physiol. **113:** 65-74

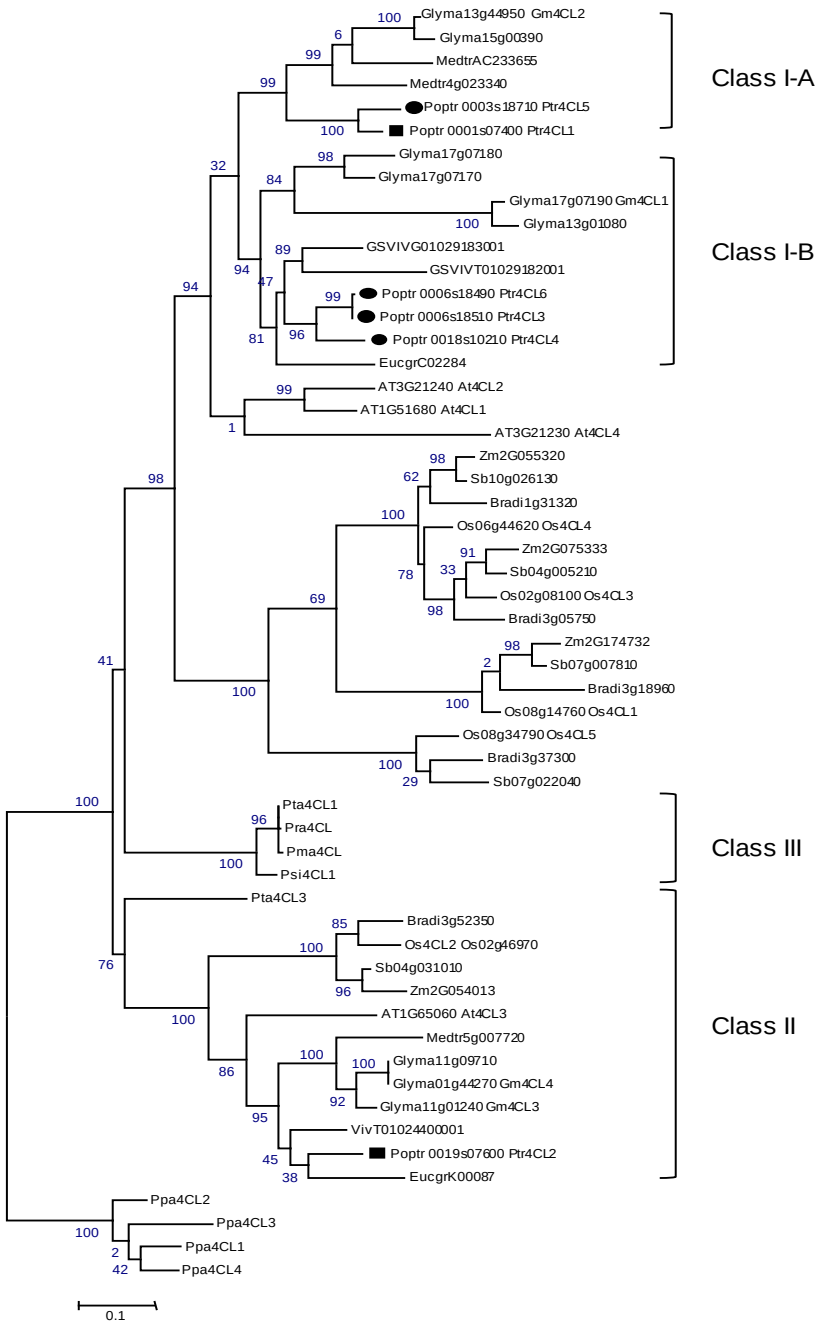


Figure 3-1. Phylogenetic analysis of 4CL isoforms from various species. Full length amino acid sequences were used for the PhyML analysis. Approximate Likelihood-Ratio Test was used to calculate the branch support. Jones-Taylor-Thornton matrix was used as the substitution model. Ptr4CL3, 4, 5, and 6 are marked with closed circles, and Ptr4CL1 and 2 are marked in closed squares. The isoforms presented with locus names were obtained from Phytozome, and the others were derived from GenBank.

Table 3-1. Primer sequences. Sequences of primers used for cloning the wild type and mutant *4CL* genes and real-time RT-PCR.

Gene/Clone	Primer sequence
Gene	
Ptr4CL1	F: ATGAATCCACAAGAATTCAT
	R: GAAGTTGCCAGGCATATAA
Ptr4CL2	F: ATGATGTCCGTGGCCACGGTTGAGC
	R: GCCACAGCCACCACCATGTCCTAG
Ptr4CL3	F: AAGGTACCGAGGCGGAAAATGATCA
	R: TTAATGAGGAAGATCGCCAG
Ptr4CL4	F: AAGGTACCGAGGCCAATAAGGATCAAG
	R: CTAATGCTGGTTTTGCATATATGTCG
Ptr4CL5	F: GGGGTACCGATACAATAACAAAGCAAAAAGAAG
	R: TTACTTTTGCAAACCACCTG
Ptr4CL1 K302L	F: CTGTCTGGAGGGGCTCCATTG
	R: GGAGCCCCTCCAGACAGTATCATCCTCAAAGAAGACAGG
Ptr4CL1 K302M	F: ATGTCTGGAGGGGCTCCATTG
	R: GGAGCCCCTCCAGACATTATCATCCTCAAAGAAGACAGG
Ptr4CL1 K302V	F: GTATCTGGAGGGGCTCCATTG
	R: GGAGCCCCTCCAGATACTATCATCCTCAAAGAAGACAGG
Ptr4CL2 L302K	F: AAGTCAGGGGCTGCGCCACTG
	R: GTGGCGCAGCCCCTGACTTGACTACCCTGATCGAACTC
Ptr4CL2 L302M	F: ATGTCAGGGGCTGCGCCACTG
	R: GTGGCGCAGCCCCTGACATGACTACCCTGATCGAACTC
Ptr4CL2 L302V	F: GTATCAGGGGCTGCGCCACTG
	R: GTGGCGCAGCCCCTGATACGACTACCCTGATCGAACTC
Real-time RT-PCR	
Ptr4CL1	F: CCAGGCATATAACTGAAGA
	R: TTGAATTACTCCATATGGCGACAGGAG
Ptr4CL2 (WT)	F: AATTCATTCCGTTAAATCTGCATT
	R: CAGAATGATGGGTTTGTAGTAATT
Hyb4CL1-a	F: AACCAGGTGCATGTGGAACTGTAG
	R: GGAGCCACCTGGAAGCCCTTATATTT
Hyb4CL1-b	F: AACCAGGTGCATGTGGAACTGTAG
	R: CGGCACCTGGAAGCCTTTGAATTT
Ptr4CL2	F: AGAAACGCAGAGCTCAAGGTC
	R: CTAGGACATGGTGGTGGC
Aspen ubiquitin-conjugating enzyme E2	F: CTGAAGAAGGAGATGACARCMCCA
	R: GCATCCCTTCAACACAGTTTCAMG
Aspen elongation factor 1-β	F: AAGAGGACAAGAAGGCAGCA
	R: CTAACCGCCTTCTCCAACAC

Table 3-2. Enzyme kinetic properties of recombinant Ptr4CL3, 4, and 5. The indicated values are the mean ± error range of two protein purification batches, each with three technical replicates.

Enzyme	Substrate	K_m μM	k_{cat} S^{-1}	k_{cat}/K_m
Ptr4CL3	**cinnamate**	1266.83±28.99	4.91±0.78	0.0039
	4-coumarate	45.80±5.90	3.56±0.36	0.0780
	caffeate	16.23±0.88	1.76±0.16	0.1084
	ferulate	48.9±6.37	4.90±0.20	0.1002
Ptr4CL4	**cinnamate**	1627.83±108.19	29.69±1.43	0.0180
	4-coumarate	84.18±16.54	52.86±17.73	0.6279
	caffeate	142.63±43.65	56.02±12.47	0.3928
	ferulate	189.53±34.98	69.33±23.08	0.3658
Ptr4CL5	**cinnamate**	-	-	-
	4-coumarate	206.73±9.10	3.72±0.41	0.0180
	caffeate	29.77±10.37	2.23±0.78	0.0749
	ferulate	473.55±29.86	4.76±1.42	0.0101

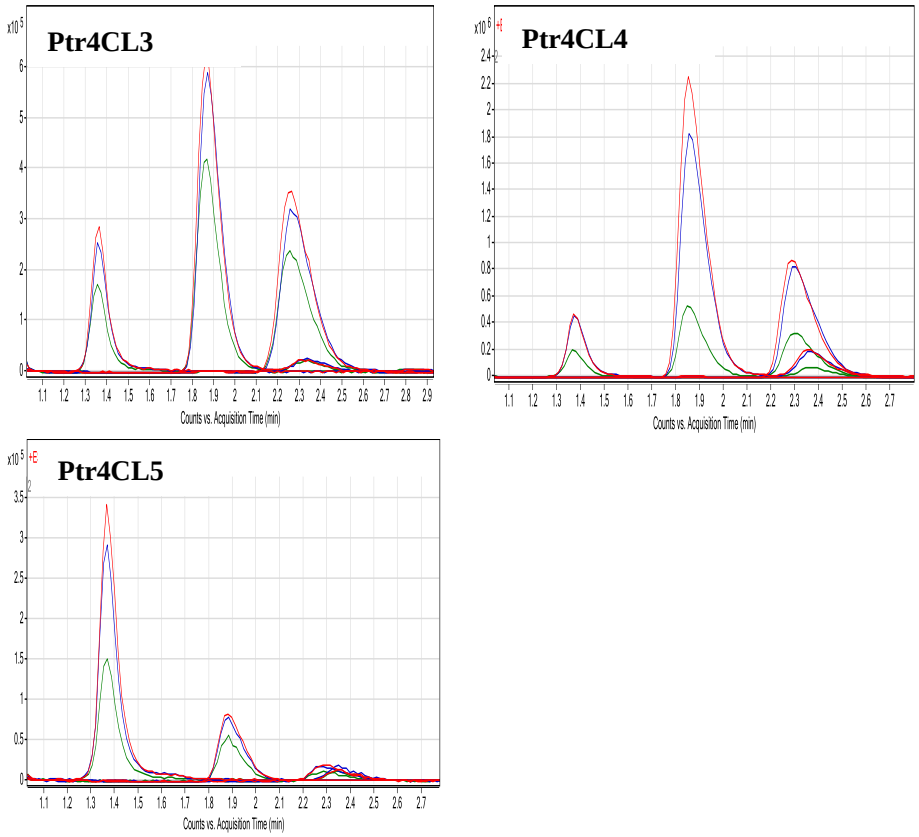


Figure 3-2. Mixed substrate enzyme assays of Ptr4CL3, 4, and 5. Shown are EIC chromatograms of CoA thioester products from mixed substrate enzyme assay with 50 µM (green), 100 µM (blue), and 500 µM (red) of equal molar substrates. CA-CoA: *m/z* 930.16, RT: 1.35 min; PA-CoA: *m/z* 914.16, RT: 1.88 min; FA-CoA: *m/z* 944.16, RT: 2.28 min.

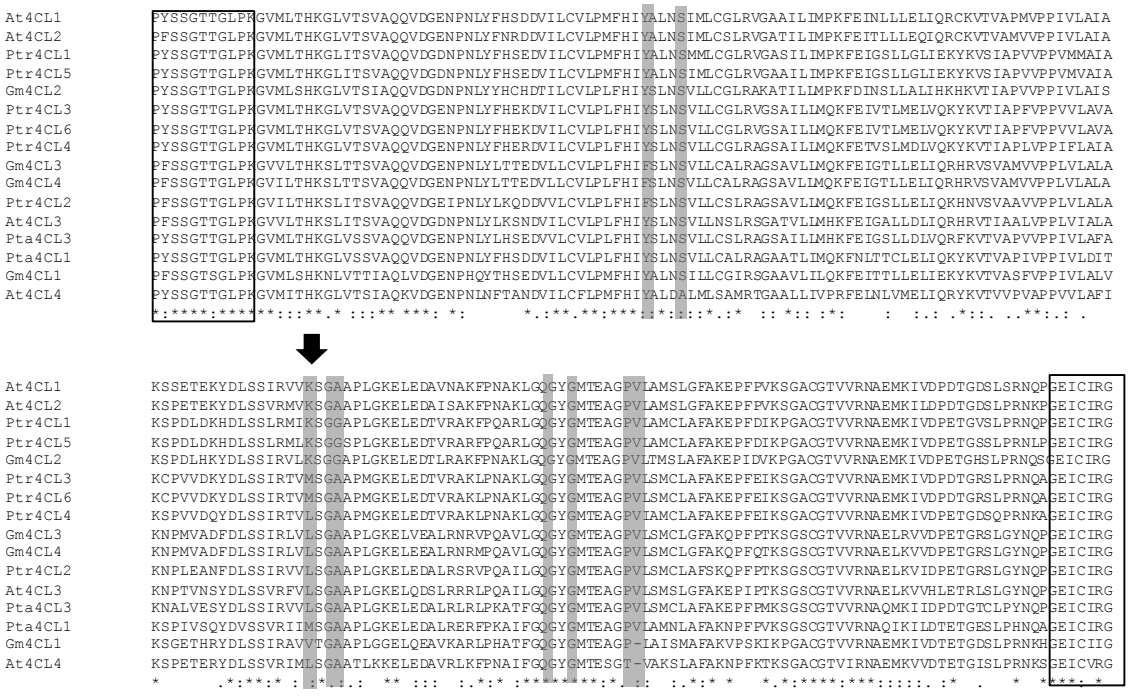


Figure 3-3. 4CL amino acid sequence alignment using the region between Box I and Box II (open boxes). The shaded positions are predicted substrate binding residues based on Pto4CL1 homology modeling. The arrow indicates the residues at position 302.

Table 3-3. Enzyme kinetic properties of recombinant Ptr4CL1, Ptr4CL2, and Ptr4CL1 mutant proteins. The indicated values are the mean ± error range of two protein purification batches each with three technical replicates. N/A indicated the values were below the detection level of the UV-spectrometry.

Enzyme	Substrate	K_m μM	k_{cat} S^{-1}	k_{cat}/K_m 1/ S*μM
Ptr4CL1	**4-coumarate**	120.69 ± 24.62	3.30±0.30	0.027
	caffeate	52.43 ± 10.22	2.33±0.33	0.044
	ferulate	N/A	N/A	N/A
Ptr4CL2	**4-coumarate**	45.45 ± 5.32	2.03±0.24	0.045
	caffeate	328.31 ± 33.01	5.71±1.33	0.017
	ferulate	N/A	N/A	N/A
Ptr4CL1 K302L	**4-coumarate**	80.91 ± 3.48	2.18±0.29	0.027
	caffeate	121.72 ± 1.40	2.40±0.42	0.020
	ferulate	210.09 ±19.70	2.47±0.66	0.012
Ptr4CL1 K302M	**4-coumarate**	128.75 ± 17.85	7.26±2.11	0.056
	caffeate	63.46 ± 2.76	3.05±0.42	0.048
	ferulate	93.89 ±7.69	3.47±0.27	0.037
Ptr4CL1 K302V	**4-coumarate**	33.42 ± 1.82	3.24±0.31	0.097
	caffeate	229.52 ± 67.13	6.89±2.10	0.030
	ferulate	41.26 ± 1.20	2.60±0.19	0.063

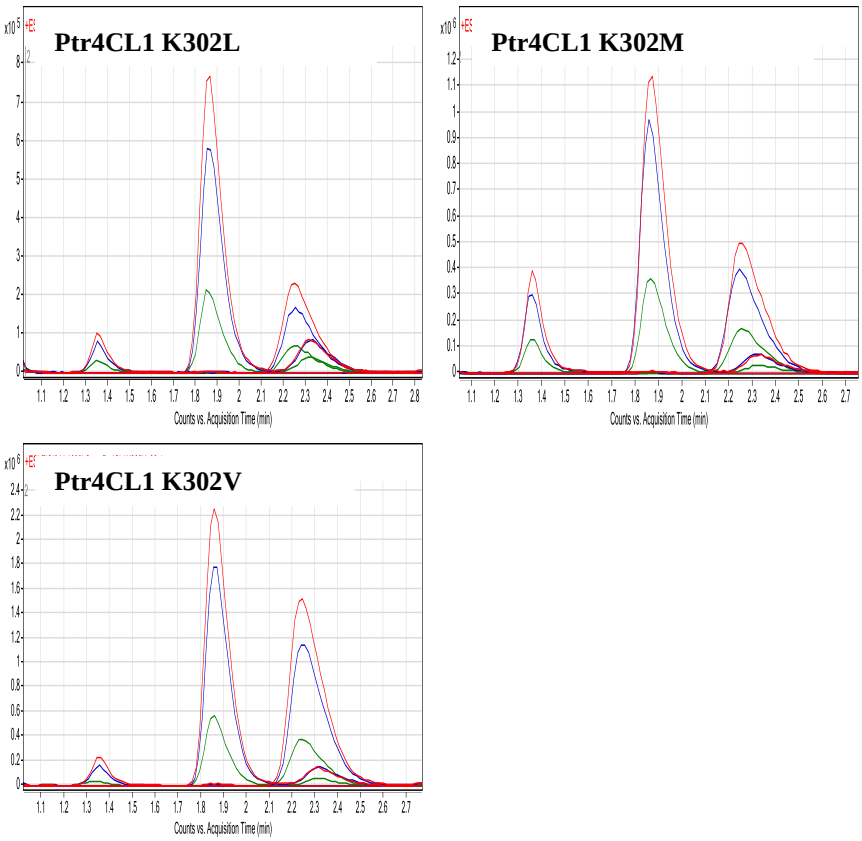


Figure 3-4. Mixed substrate enzyme assays of Ptr4CL1mutants. Shown are EIC chromatograms of CoA thioester products from mixed substrate enzyme assay with 50 µM (green), 100 µM (blue), and 500 µM (red) of equal molar substrates. CA-CoA: *m/z* 930.16, RT: 1.35 min; PA-CoA: *m/z* 914.16, RT: 1.88 min; FA-CoA: *m/z* 944.16, RT: 2.28 min.

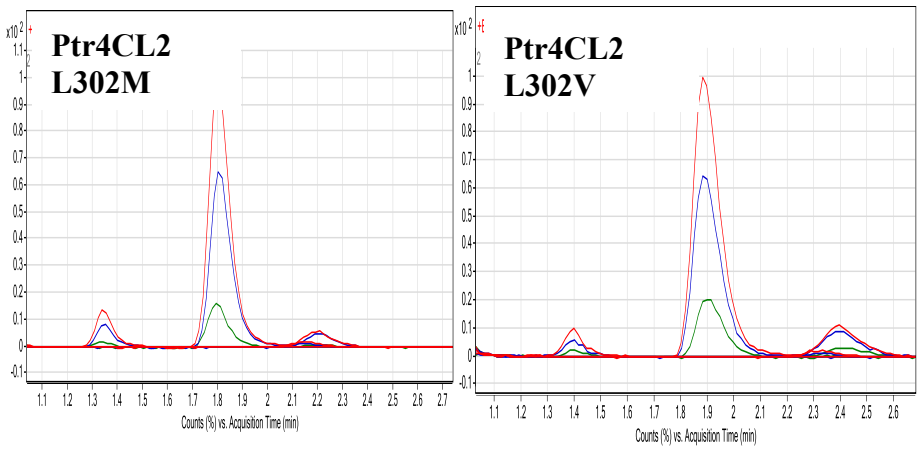


Figure 3-5. Mixed substrate enzyme assays of Ptr4CL2 mutants. Shown are EIC chromatograms of CoA thioester products from mixed substrate enzyme assay with 50 µM (green), 100 µM (blue), and 500 µM (red) of equal molar substrates. CA-CoA: *m/z* 930.16, RT: 1.35 min; PA-CoA: *m/z* 914.16, RT: 1.88 min; FA-CoA: *m/z* 944.16, RT: 2.28 min.

Table 3-4. Enzyme kinetic properties of recombinant Hyb4CL1-a and Hyb4CL1-b. The indicated values are the mean ± error range of two protein purification batches each with three technical replicates. N/A indicated the values were below the detection level of the UV-spectrometry.

Enzyme	Substrate	K_m μM	k_{cat} S^{-1}	k_{cat}/K_m
Hyb4CL1-a	**4-coumarate**	181.93 ± 4.57	5.41±1.36	0.030
	caffeate	43.05 ± 5.53	1.84±0.13	0.043
	ferulate	165.24 ± 64.23	2.41±0.90	0.015
Hyb4CL1-b	**4-coumarate**	N/A	N/A	N/A
	caffeate	235.43 ± 15.06	12.89 ± 0.98	0.055
	ferulate	N/A	N/A	N/A

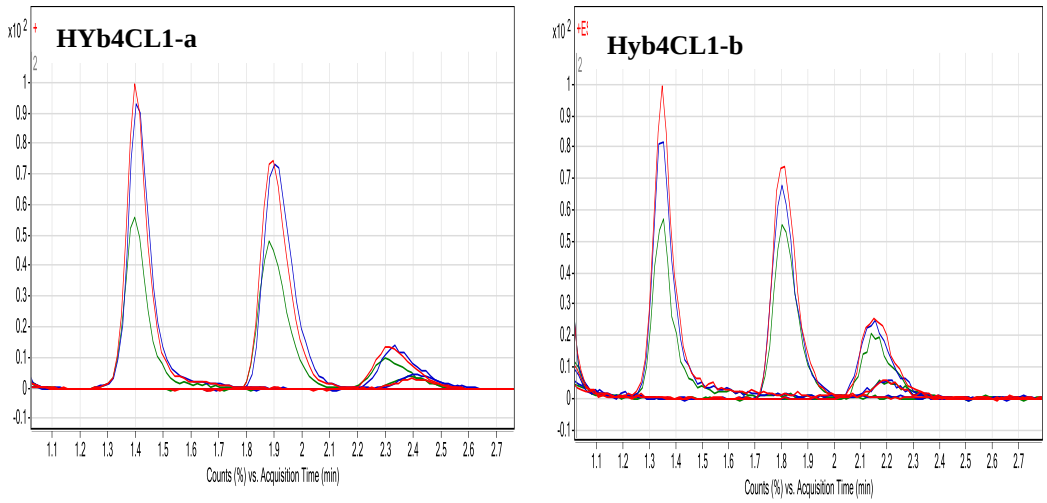


Figure 3-6. Mixed substrate enzyme assays of domain-swapped 4CLs. Shown are EIC chromatograms of CoA thioester products from mixed substrate enzyme assay with 50 µM (green), 100 µM (blue), and 500 µM (red) of equal molar substrates. CA-CoA: *m/z* 930.16, RT: 1.35 min; PA-CoA: *m/z* 914.16, RT: 1.88 min; FA-CoA: *m/z* 944.16, RT: 2.28 min.

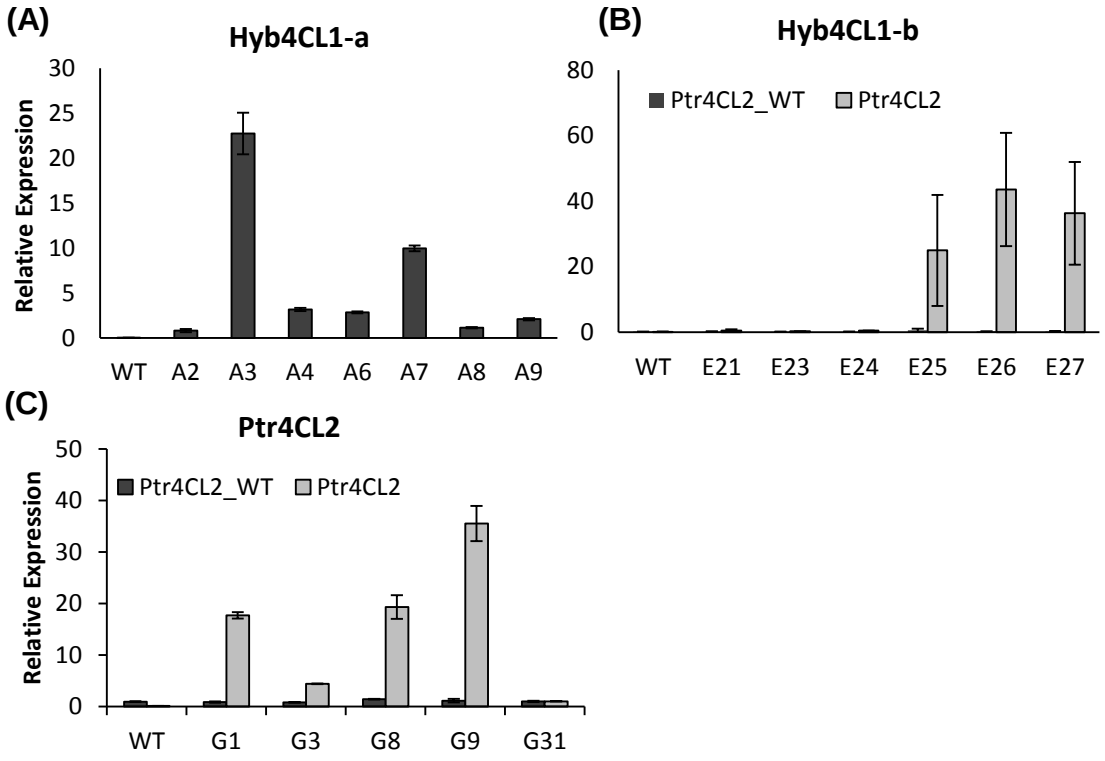


Figure 3-7. The pre-screening of the transgenic lines by using qRT-PCR. The transcript levels of the transgenes in the transgenic lines Hyb4CL1-a (A), Hyb4CL1-b (B), and Ptr4CL2 (C) were analyzed in the xylem tissues. For Hyb4CL-b and Ptr4CL2 panels, Ptr4CL2 _WT primers (black) amplified the endogenous *Ptr4CL2* only, and Ptr4CL2 primers amplified both endogenous and transgenic *Ptr4CL2* sequences. *Pta4CL1* primers were used to amplify the *Pta4CL1* sequence on the 3' end of *Hyb4CL1-a*.

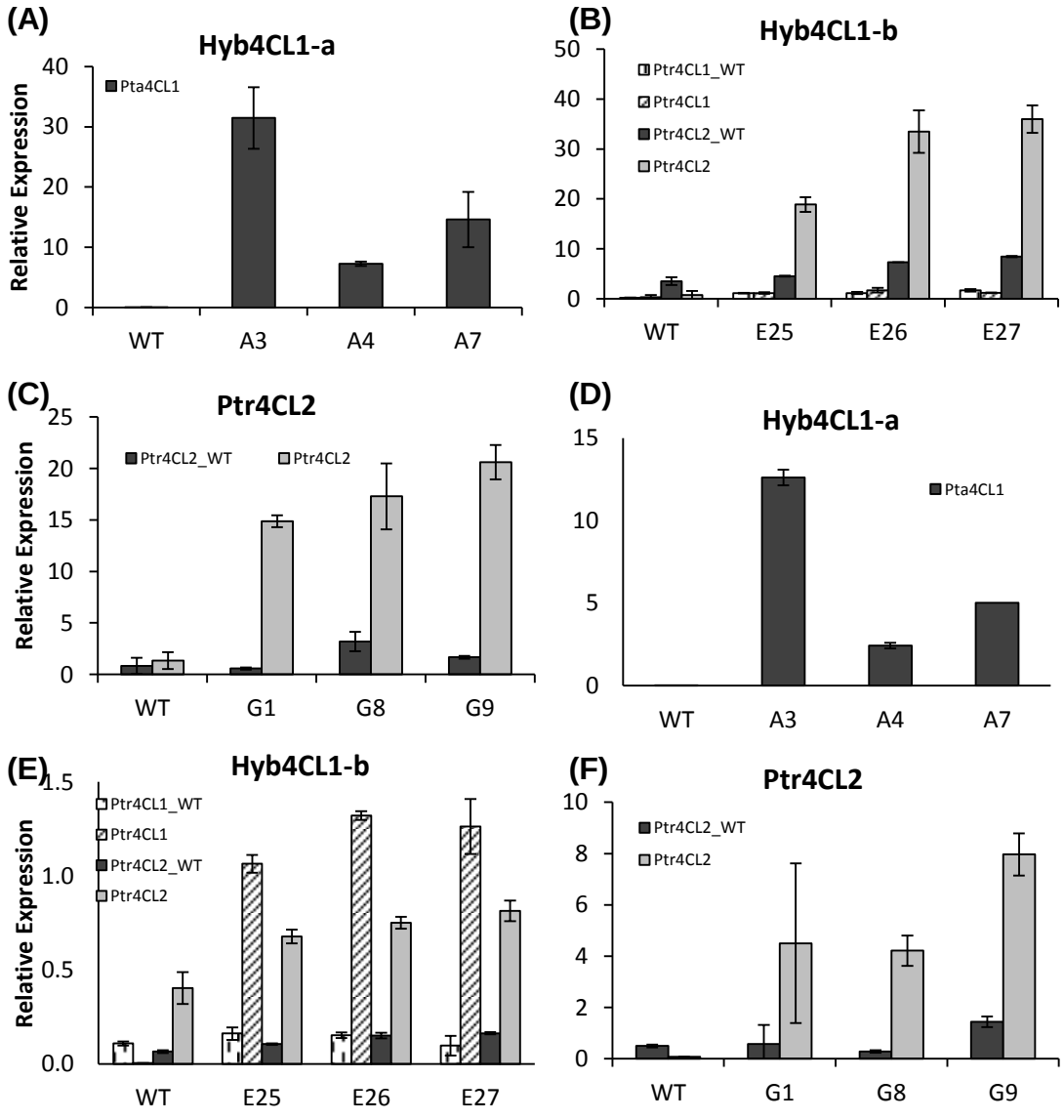


Figure 3-8. The pre-screening of the transgenic lines by using qRT-PCR. The transcript levels of the transgenes in the transgenic lines Hyb4CL1-a, Hyb4CL1-b, and Ptr4CL2 were analyzed in the leaf LPI 5 (A to C) and bark (D to F)tissues. Ptr4CL1_WT and Ptr4CL2 _WT primers were endogenous *Ptr4CL1* and *Ptr4CL2* specific. *Ptr4CL1* and *Ptr4CL2* primers amplified both endogenous and transgenic 4CL sequences. *Pta4CL1* primers were used to amplify the *Pta4CL1* sequence on the 3' end of *Hyb4CL1-a*.

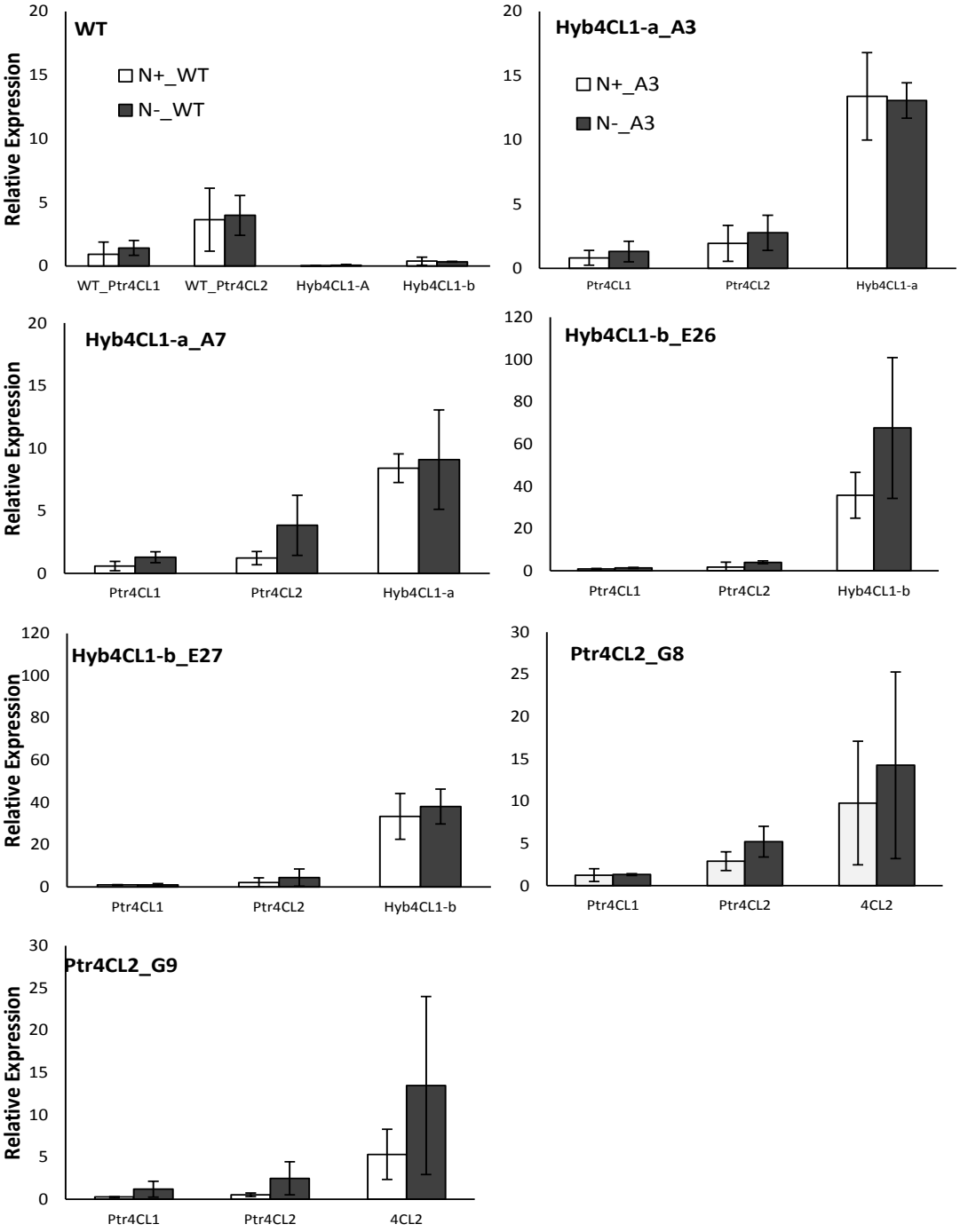


Figure 3-9. The transcription levels of endogenous Ptr4CLs and transgenic 4CLs in leaf LPI 5 of wild type and transgenic *Populus*. Two replicates of the 4CL2-G8 group and three replicates of all the other groups were used for the real-time RT-PCR analysis. The expression was normalized to the geometric mean of *elongation factor 1-β* (EF1B) and *ubiquitin-conjugating enzyme E2* (UBCc) using the delta-CT method. Error bars indicate the standard deviation of the biological replicates. Opened bars, N+ treated samples; Closed bars, N- treated samples.

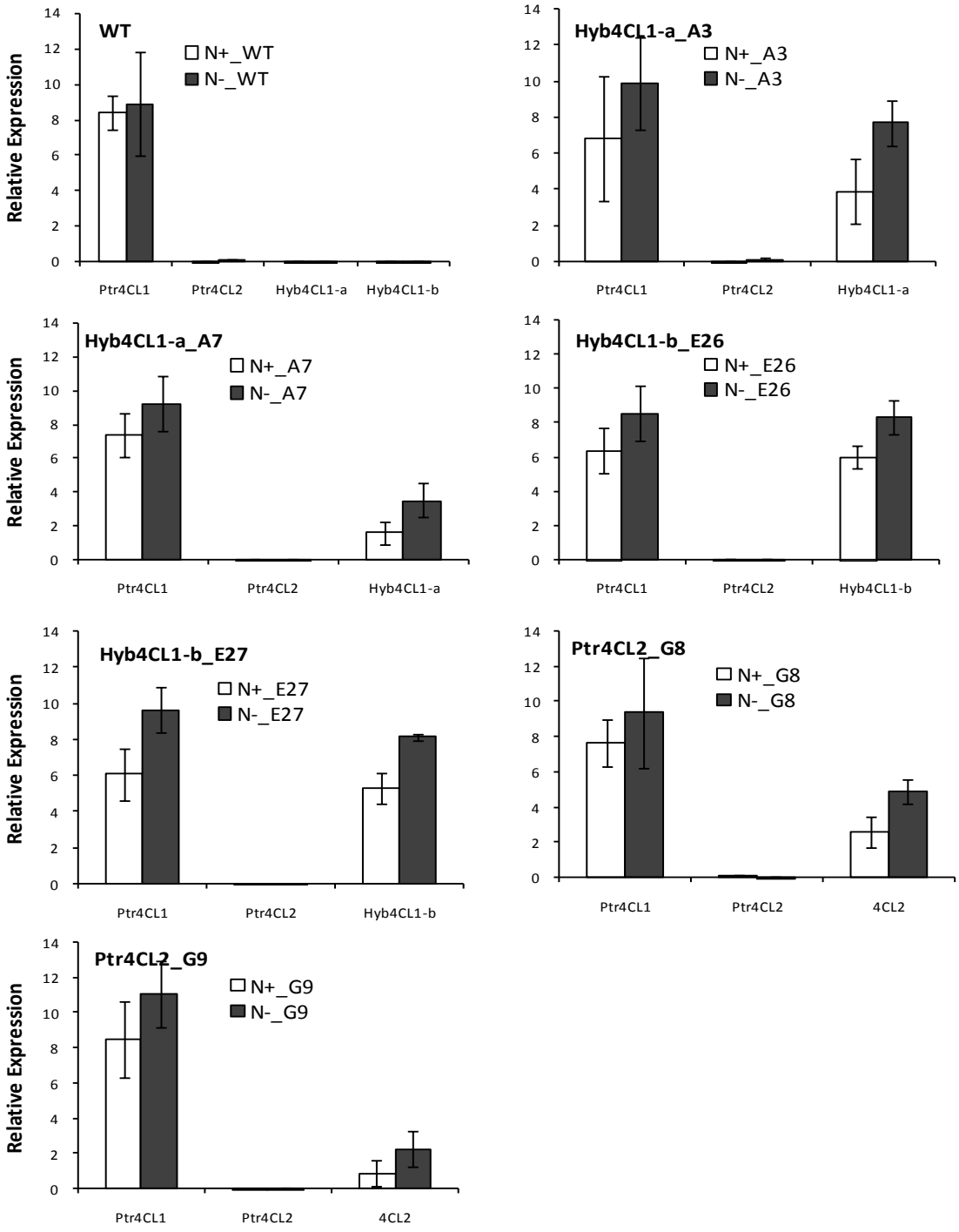


Figure 3-10. The transcription levels of endogenous Ptr4CLs and transgenic 4CLs in xylem of wild type and transgenic *Populus*. Two replicates of the 4CL2-G8 group and three replicates of all the other groups were used for the real-time RT-PCR analysis. The expression was normalized to the geometric mean of *elongation factor 1-β* (EF1B) and *ubiquitin-conjugating enzyme E2* (UBCc) using the delta-CT method. Error bars indicate the standard deviation of the biological replicates. Opened bars, N+ treated samples; Closed bars, N- treated samples.

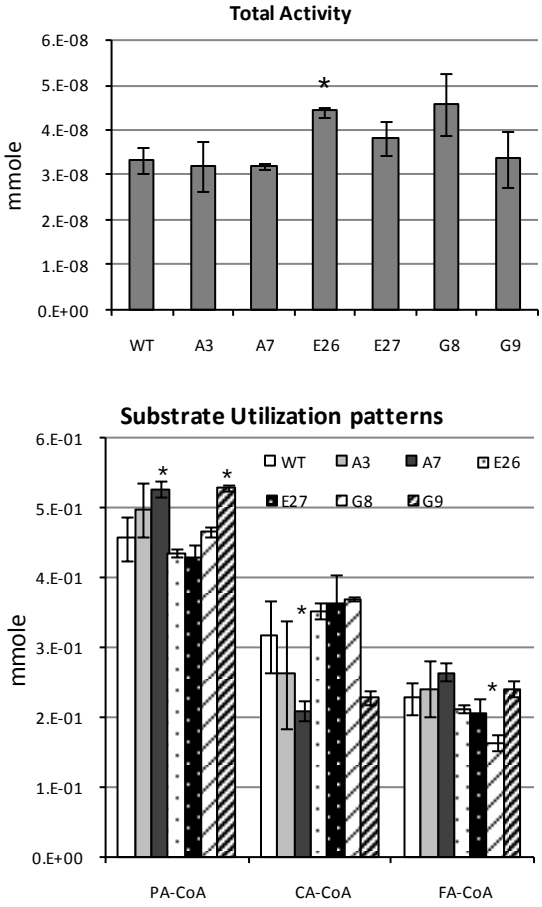


Figure 3-11. 4CL mixed substrate assays using crude protein extracted from leaf LPI 5. Total 4CL activity (A) was the sum of all hydroxycinnamoyl-CoA in each reaction. The substrate utilization pattern of each was presented by the ratio of each product amount over total product amount. The asterisk marks indicated statistical significant ($p< 0.05$) difference between the transgenic line and the wild type control.

LPI 2

	N+						N-					
	Hyb4CL1-a/WT		Hyb4CL1-b/WT		Ptr4CL2/WT		Hyb4CL1-a/WT		Hyb4CL1-b/WT		Ptr4CL2/WT	
	A3	A7	E26	E27	G8	G9	A3	A7	E26	E27	G8	G9
4-coumarate	0.1	0.1	-0.5	-0.2	-0.1	0	0.2	-0.2	-0.2	-0.1	-0.3	-0.1
caffeate	-	-	-	-	-	-	0.6	0.1	0.2	0	-0.1	-0.2
4HBA glucos ester	0	0.3	0.5	**1.1**	-0.1	0.4	0	-0.1	-0.2	0	-0.1	-0.1
4-coumaroyl glucose ester	0	**0.5**	0.3	**0.4**	0.5	0.2	0.3	-0.4	-0.7	-0.9	-0.1	**-1.4**
2-Coumaric acid glucose ester	-	-	-	-	-	-	-	-	-	-	-	-
caffeic acid glucose ester (1)	0	0.7	-1.6	-1.6	1.4	-0.8	0.3	0	**-1.9**	**-1.5**	-0.2	-0.7
caffeic acid glucose ester (2)	-	-	-	-	-	-	-	-	-	-	-	-
ferulic acid glucose ester (1)	0.1	0.2	0.7	**0.9**	**0.8**	0.4	0.1	-0.2	**-0.8**	-0.1	-0.4	-0.3
ferulic acid glucose ester (2)	0	0.4	-1	-0.6	1.2	-0.9	0	-0.1	-0.5	0	-2.1	0
5-*O*-caffeoyl quinic acid	0.1	0.1	0.2	0.5	1.1	0.1	0.3	-0.3	0	0.1	0.6	0.3
3-*O*-caffeoyl quinic acid (1)	0.1	0.3	0.2	0.4	1.1	0.2	0	-0.2	0.1	0.1	-0.1	0
3-*O*-caffeoyl quinic acid (2)	0	0.2	0.3	0.3	0.7	0	0	-0.2	0	0	0	0
5-O-Caffeoyl shikimic acid	-	-	-	-	-	-	-	-	-	-	-	-
kaempferol	1	0.6	0.4	0.3	0.5	0.3	0.3	-0.3	-0.3	0	-1	-0.5
catechin	-0.1	0.3	1.2	0.5	**1.1**	0	0.3	-0.5	0.4	0	0.6	-0.6
coniferyl alcohol	0.3	0.4	0	0	0.5	-0.2	-	-	-	-	-	-

Bark

	N+						N-					
	Hyb4CL1-a/WT		Hyb4CL1-b/WT		Ptr4CL2/WT		Hyb4CL1-a/WT		Hyb4CL1-b/WT		Ptr4CL2/WT	
	A3	A7	E26	E27	G8	G9	A3	A7	E26	E27	G8	G9
4-coumarate	0.2	-0.7	-0.9	-0.8	-0.4	-0.4	-0.1	-0.7	-1.6	-1.9	-0.8	-1.7
caffeate	-	-	-	-	-	-	0.2	0.4	-0.9	0.1	-0.1	-0.3
4HBA glucos ester	-0.1	-0.3	0	-0.4	-0.3	-0.5	-0.2	-0.3	-0.3	-0.5	-0.4	**-0.8**
4-coumaroyl glucose ester	0	-0.1	**-0.7**	**-0.9**	-0.8	**-1.3**	-0.1	-0.1	-0.6	-1	-1.6	**-1.5**
2-Coumaric acid glucose ester	0	-0.2	-0.9	**-1.3**	-1.5	**-1.5**	-0.8	-0.3	**-1.3**	**-1.7**	**-2.3**	**-2**
caffeic acid glucose ester (1)	0	-0.5	-1.1	**-1.4**	0	-0.4	-0.4	-0.8	**-1.1**	**-1.5**	-0.8	-0.5
caffeic acid glucose ester (2)	0	-0.2	**-0.9**	**-1.1**	-0.5	**-0.5**	-0.3	-0.3	**-1**	**-0.9**	-0.2	-0.6
ferulic acid glucose ester (1)	0	-0.1	-0.6	**-0.7**	0.2	-0.2	-0.2	-0.3	**-1**	**-1**	**-0.7**	**-0.5**
ferulic acid glucose ester (2)	**-0.5**	-0.1	0.5	0.3	0	0.4	-0.5	-0.1	0.4	-0.4	0	0.4
5-*O*-caffeoyl quinic acid	0	-0.2	-0.4	0	0.2	0.4	0	-0.7	-0.3	0.1	0.6	0.3
3-*O*-caffeoyl quinic acid (1)	-0.3	-0.3	**1**	0.4	**0.6**	0.9	**-0.1**	0	1.2	0.7	0.7	1
3-*O*-caffeoyl quinic acid (2)	-0.2	0	**1**	0.5	**0.6**	**1**	**-0.3**	-0.1	0.9	0.4	0.5	0.8
5-O-Caffeoyl shikimic acid	-	-	-	-	-	-	0.2	0.4	-0.9	0.1	-0.1	-0.3
kaempferol	-	-	-	-	-	-	-	-	-	-	-	-
catechin	0.4	0.1	1.2	0.9	**1.2**	0.6	0	-0.6	-0.3	-0.5	-0.5	-0.6
coniferyl alcohol	-	-	-	-	-	-	-	-	-	-	-	-

Figure 3-12. The metabolic profile results of *Hybrid 4CL1-a*, *Hyb4CL1-b*, and *Ptr4CL2* transgenic lines presented in heat maps. Three tissues, including young leaf , bark, and xylem were analyzed by LC-MS/TOF and GC-MS. The fold change between transgenic events and wild-type control of each compound was Log 2 transformed. Statistical significance of metabolite level between the transgenic events and control was determined pair wise comparison. Bold fonts and underlines indicate significant difference at $p < 0.05$. Metabolite in grey color indicates the predicted results by searching KNApSAcK database using the *m/z* or calculated molecular formula.

	Stem											
	N+						N-					
	Hyb4CL1-a/WT		Hyb4CL1-b/WT		Ptr4CL2/WT		Hyb4CL1-a/WT		Hyb4CL1-b/WT		Ptr4CL2/WT	
	A3	A7	E26	E27	G8	G9	A3	A7	E26	E27	G8	G9
4-coumarate	-	-	-	-	-	-	-	-	-	-	-	-
caffeate	-	-	-	-	-	-	-	-	-	-	-	-
4HBA glucos ester	-	-	-	-	-	-	-	-	-	-	-	-
4-coumaroyl glucose ester	0	0.2	**1.2**	**1**	0.9	**0.6**	0.7	-0.3	-1.1	-1.3	-0.6	**-1.5**
2-Coumaric acid glucose ester		-	-	-	-							
caffeic acid glucose ester (1)	0.5	-0.7	-0.1	-0.7	-0.3	0.9	-0.1	0	**-1**	**-0.3**	-0.3	-0.1
caffeic acid glucose ester (2)	-	-	-	-	-		-	-	-	-	-	-
ferulic acid glucose ester (1)	-	-	-	-	-		-	-	-	-	-	-
ferulic acid glucose ester (2)	-	-	-	-	-		-	-	-	-	-	-
5-*O*-caffeoyl quinic acid	0	0	0	0	0	0	-0.9	-1.3	-0.7	-0.1	-0.1	0.1
3-*O*-caffeoyl quinic acid (1)	-0.2	0	**-0.6**	**-0.7**	-0.5	**-0.4**	-0.9	-0.8	-0.8	0	-0.3	0.3
3-*O*-caffeoyl quinic acid (2)	0.4	0.2	0	-0.4	0	0	**-0.8**	-0.5	-0.3	0	0	0.4
5-O-Caffeoyl shikimic acid	0	0.3	0.6	0.2	0	0	-0.4	-0.2	-0.3	0.2	0	0.4
kaempferol	-	-	-	-	-	-	-	-	-	-	-	-
catechin	-	-	-	-	-	-	-	-	-	-	-	-
coniferyl alcohol	0.2	**0.5**	0	0.4	0.3	0.6	-0.1	0	-0.4	0	-0.2	0.3

Figure 3-12. The metabolic profile results of *Hybrid 4CL1-a*, *Hyb4CL1-b*, and *Ptr4CL2* transgenic lines presented in heat maps. Three tissues, including young leaf , bark, and xylem were analyzed by LC-MS/TOF and GC-MS. The fold change between transgenic events and wild-type control of each compound was Log 2 transformed. Statistical significance of metabolite level between the transgenic events and control was determined by pair wise comparison. Bold fonts and underlines indicate significant difference at $p < 0.05$. Metabolite in grey color indicates the predicted results by searching KNApSAcK database using the *m/z* or calculated molecular formula.

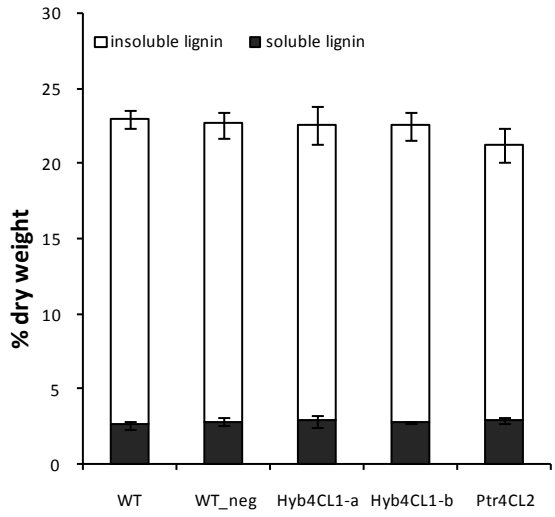


Figure 3-13. The soluble lignin, insoluble lignin, and total lignin content of wild-type (WT717), negative transformant (WT_neg), Hyb4CL1-a, Hyb4CL1-b, and Ptr4CL2 transgenic lines in percent dry weight.

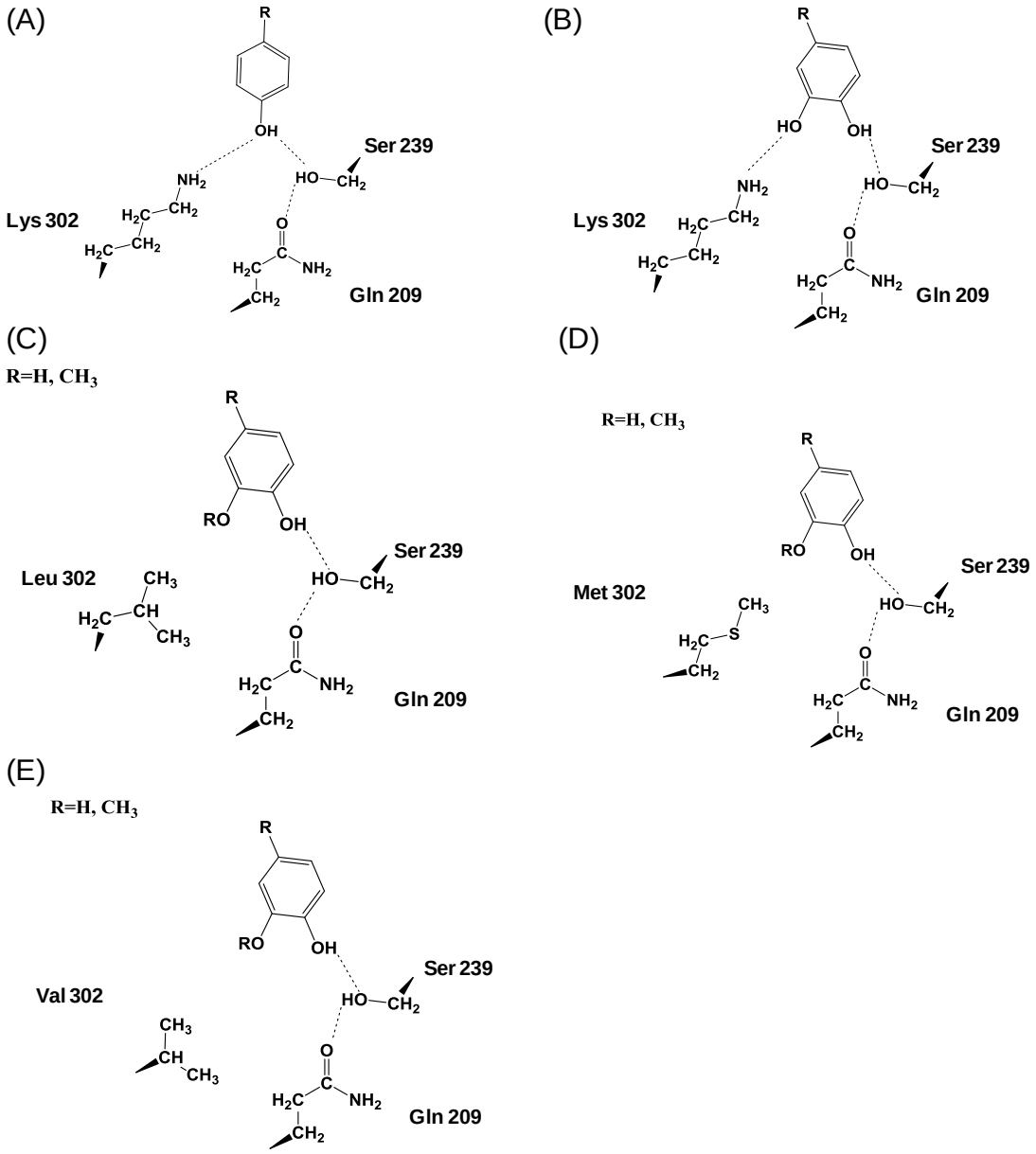


Figure 3-14. The proposed model of different substrate binding mechanisms in the substrate binding pocket of distinct Ptr4CL1 mutants. Based on the crystal model, the 4-hydroxyl group of PA bonds with the hydrogen bond partners (dash lines), Lys 302, Ser 239, and Gln 209 (A). Caffeate can also form H-bonds with these three residues (B). In (C), (D), and (E), different mutation and substrates (caffeate, R=H; ferulate, R=OH) resulted in distinct binding mechanisms are illustrated.

CHAPTER 4

CONCLUSIONS AND FUTURE WORK

Conclusions

Lignin biosynthesis-associated 4CLs in angiosperm and gymnosperm species form two phylogenetically distinct clades known as Class I and Class III, respectively (Ehlting et al., 1999; Chen, 2008). Based on the sequences of the newly characterized aspen isoforms, Ptr4CL3-6, angiosperm Class I now contains two subclades. Class I-A and I-B. In addition, loblolly pine Pta4CL3 becomes the first Class II isoform characterized from a gymnosperm (Fig. 2-1). The *in vivo* roles of the new Class I-B isoforms, Ptr4CL3-6, remain uncharacterized, but there is variation in the *in vitro* substrate preference among Class I-B members and between Class I-B and I-A. In addition, I-A and I-B isoforms exhibit spatially distinct gene expression patterns. Relative to lignin biosynthesis-associated 4CLs, Class II isoforms are more conserved between the seed plants, and there has been less expansion of the family. (Kajita et al., 1996; Lee et al., 1997; Zhang and Chiang, 1997; Ehlting et al., 1999; Lindermayr et al., 2002; Hamberger and Hahlbrock, 2004; Chen, 2008; Wagner et al., 2009). However, functional characterization of Class II 4CLs is limited, especially in gymnosperm species (Hu et al., 1998; Ehlting et al., 1999; Lindermayr et al., 2002). The isoforms in this group are thought to support the biosynthesis of nonstructural phenylpropanoids rather than that of lignin (Harding et al., 2002; Kao et al., 2002). Based on the identification of pine

Pta4CL3 as a Class II isoform, 4CL divergence was underway before the evolutionary split of the angiosperm and gymnosperm lineages. The recently published Norway spruce (*Picea abies*) and white spruce (*Pieca glauca*) genome sequences (Birol et al., 2013; Nystedt et al., 2013) should facilitate further investigations of phenylpropanoid pathway gene evolution of gymnosperms, as shown for 4CL in this study.

The structural basis for differing substrate selectivities among 4CL isoforms, and the relevance of substrate utilization to phenylpropanoid metabolism were explored using *in vitro* and *in vivo* approaches. A very large component of the *in vitro* substrate selectivity bias was found to depend on a single binding pocket amino acid residue at position 302. Manipulation of this residue by site-directed mutagenesis or via domain swapping led to changes in substrate utilization patterns *in vitro*, especially in Ptr4CL1. Amino acid substitutions at position 302 of Ptr4CL2 had little effect on substrate specificity. Binding pocket size as well as a difference at position 235 between 4CL1 and 4CL2 may have contributed to the differential effectiveness of 302 substitutions in these two isoforms. Over-expressing wild-type or mutant 4CLs in transgenic *Populus* affected phenylpropanoid metabolism in ways that differed from expectations based on their *in vitro* substrate utilization biases. Minor metabolic changes *in vivo* were observed in the transgenic *Populus* over-expressing loblolly pine Pta4CL1 and Hyb4CL1-a, which is a domain-swapped mutant of loblolly pine Pta4CL1 in which the substrate binding domain was replaced by that of aspen Ptr4CL1. These results are consistent with the argument that the 4CL substrate binding domain is not the sole determinant of 4CL function *in vivo*. An emerging concept is that protein-protein interactions of various kinds modulate the roles of enzymatic proteins *in vivo* (Winkel, 2004; Jørgensen et al., 2005). A recent study

identified a productive protein-protein interaction between 4CL and *p*-coumarate 3-hydroxylase (C3H; Bassard et al., 2012). The findings are consistent with 4CL participation in multi-enzyme metabolons that can modulate the efficiency and output of metabolic pathways. Such metabolic channeling has been reported in several recent studies (Achnine et al., 2004; Chen et al., 2011; Bassard et al., 2012). One interpretation of the weak metabolic effect of over-expressing the hybrid 4CL is that there was a poor association with the hypothetical metabolon. The backbone sequence may be essential for successful integration with the phenylpropanoid metabolon. Due to the sequence divergence between Pta4CL1 and *Populus* 4CLs (Voo et al., 1995; Hu et al., 1998), there may be structural impediments to association of Pta4CL1 with the *Populus* metabolons.

While there were limited metabolic changes in Pta4CL1 transgenic lines, there were metabolic changes in the Pta4CL3 transgenic lines. The unusual metabolite cinnamoyl-quinate was detected at much higher levels in Pta4CL3 transgenic than in wild-type plants. The biosynthesis of cinnamoyl-quinate requires the CoA-activation of cinnamate (Liu et al., 1995), a poor *in vitro* substrate of all 4CLs. Pta4CL3 exhibited poor, but better *in vitro* affinity with cinnmate than did the other 4CL proteins we tested. This finding raises the possibility that certain 4CL isoforms utilize cinnamate *in vivo*. The *in vivo* effect of 4CL CoA-activating cinnamate may reflect high levels of 4CL enzyme, since the crude protein assay results indicated an elevated 4CL preference toward *p*-coumarate in Pta4CL3 transgenics. However, Pta4CL3 exhibits low affinity *in vitro* toward cinnamate, and cinnamate CoA-activation requires high substrate concentration that is not observed in *Populus* tissues. Alternatively, the accumulation of cinnamoyl-quinate may be considered in the context of metabolic channeling involving Pta4CL3 and

the *Populus* metabolon. It is at least reasonable to suggest that with greater sequence conservation among Class II 4CLs than among Class I and III 4CLs, the loblolly pine Pta4CL3 was able to interact in a productive way with the *Populus* metabolon. The metabolons may lower the kinetic constraint of Pta4CL3 by preventing the dilution of cinnamate in the cytosol (Jørgensen et al., 2005).

Future Work

Our results suggested that the backbone sequence of 4CL may affect its function *in vivo*. The interaction between 4CL and C3H was previously observed (Bassard et al., 2012). To further test the hypothesis that the 4CL backbone sequence is involved in a hypothetical protein-protein interaction with C3H, wild-type and mutant 4CLs can be used for both *in vitro* and *in vivo* assays. With the published Pto4CL1 crystal structure (Hu et al., 2010), the mechanism of the interaction between 4CL and C3H may be illustrated by computational analysis. Phenylalanine ammonia-lyase (PAL), which catalyzes the entry step of the general phenylpropanoid pathway, physically interacts with C4H (Achnine et al., 2004), and the strength of the interaction differs between different PAL isoforms. It would be interesting to explore the strength of interactions between C3H and various 4CL native or mutant isoforms from different phylogenetic classes or with altered sequences. Since 4CL classification is correlated with differential gene expression in different tissues and *in vivo* functions, the result may further our understanding of phenylpropanoid metabolism in distinct tissues. For the 4CL substrate selectivity manipulation discussed in chapter three, position 235 may also play an

important role in Ptr4CL2 substrate binding, and double mutations at position 235 and 302 in Ptr4CL2 may exert different effects on the substrate selectivity than single mutations at position 302 alone. The proposed mutagenesis studies in Ptr4CL2 should illustrate the substrate binding pocket difference between Ptr4CL1 and Ptr4CL2, and enable the design of novel 4CLs for metabolic engineering.

References

Achnine L, Blancaflor EB, Rasmussen S, Dixon RA (2004) Colocalization of L-phenylalanine ammonia-lyase and cinnarnate 4-hydroxylase for metabolic channeling in phenylpropanoid biosynthesis. Plant Cell **16:** 3098-3109

Anino EO (2008) Characterization of Populus tremuloides Caffeic acid O-methyltransferase and 4-coumarate:CoA ligase gene promoters to identify regulatory elements. Michigan Technology University **Ph. D. dissertation**

Anisimova M, Gascuel O (2006) Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. Systematic Biology **55:** 539-552

Anterola AM, Lewis NG (2002) Trends in lignin modification: a comprehensive analysis of the effects of genetic manipulations/mutations on lignification and vascular integrity. Phytochemistry **61:** 221-294

Arnold K, Bordoli L, Kopp Jr, Schwede T (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics **22:** 195-201

Babbitt PC, Kenyon GL, Martin BM, Charest H, Slyvestre M, Scholten JD, Chang KH, Liang PH, Dunawaymariano D (1992) Ancestry of the 4-chlorobenzoate dehalogenase - analysis of amino acid sequence identities among families of acyl-adenyl ligase, enoyl-CoA hydratases isomerase, and acyl-CoA thioesteraes. Biochemistry **31:** 5594-5604

Bassard JE, Richert L, Geerinck J, Renault H, Duval F, Ullmann P, Schmitt M, Meyer E, Mutterer J, Boerjan W, De Jaeger G, Mely Y, Goossens A, Werck-Reichhart D (2012) Protein-protein and protein-membrane associations in the lignin pathway. Plant Cell **24:** 4465-4482

Becker-Andre M, Schulze-Lefert P, Hahlbrock K (1991) Structural comparison, modes of expression, and putative cis-acting elements of the two 4-coumarate: CoA ligase genes in potato. J. Biol. Chem. **266:** 8551-8559

Bellay J, Michaut M, Kim T, Han S, Colak R, Myers CL, Kim PM (2011) An omics perspective of protein disorder. Molecular BioSystems **8:** 185-193

Birol I, Raymond A, Jackman SD, Pleasance S, Coope R, Taylor GA, Yuen MMS, Keeling CI, Brand D, Vandervalk BP, Kirk H, Pandoh P, Moore RA, Zhao Y, Mungall AJ, Jaquish B, Yanchuk A, Ritland C, Boyle B, Bousquet J, Ritland K, MacKay J, Bohlmann Jr, Jones SJM (2013) Assembling the 20 Gb white spruce (Picea glauca) genome from whole-genome shotgun sequencing data. Bioinformatics

Boerjan W, Ralph J, Baucher M (2003) Lignin biosynthesis. Annual Review of Plant Biology **54:** 519-546

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Analytical Biochemistry **72:** 248-254

Bravo L (1998) Polyphenols: chemistry, dietary sources, metabolism, and nutritional significance. Nutrition Reviews **56:** 317-333

Burbulis IE, Winkel-Shirley B (1999) Interactions among enzymes of the Arabidopsis flavonoid biosynthetic pathway. Proceedings of the National Academy of Sciences **96:** 12929-12934

Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Molecular Biology and Evolution **17:** 540-552

Chen H-C, Li Q, Shuford CM, Liu J, Muddiman DC, Sederoff RR, Chiang VL (2011) Membrane protein complexes catalyze both 4- and 3-hydroxylation of cinnamic acid derivatives in monolignol biosynthesis. Proceedings of the National Academy of Sciences **108:** 21253-21258

Chen H-C, Song J, Williams CM, Shuford CM, Liu J, Wang JP, Li Q, Shi R, Gokce E, Ducoste J, Muddiman DC, Sederoff RR, Chiang VL (2013) Monolignol pathway 4-coumaric acid:coenzyme A ligases in Populus. trichocarpa: Novel specificity, metabolic regulation, and simulation of coenzyme A ligation fluxes. Plant Physiology **161:** 1501-1516

Chen H-Y (2008) Characterization of two distinct 4-coumarate:CoA ligase (4CL) genes in loblolly pine (*Pinus taeda*). Unpublished master thesis

Cheung WY, Hubert N, Landry BS (1993) A simple and rapid DNA microextraction method for plant, animal, and insect suitable for RAPD and other PCR analyses. Pcr-Methods and Applications **3:** 69-70

Conti E, Stachelhaus T, Marahiel MA, Brick P (1997) Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. Embo Journal **16:** 4174-4183

Cukovic D, Ehlting J, VanZiffle JA, Douglas CJ (2001) Structure and evolution of 4-coumarate : coenzyme A ligase (4CL) gene families. Biological Chemistry **382:** 645-654

Davin LB, Lewis NG (2005) Lignin primary structures and dirigent sites. Current Opinion in Biotechnology **16:** 407-415

de Wet JR, Wood KV, DeLuca M, Helinski DR, Subramani S (1987) Firefly luciferase gene: structure and expression in mammalian cells. Molecular and Cellular Biology **7:** 725-737

Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard JF, Guindon S, Lefort V, Lescot M, Claverie JM, Gascuel O (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Research **36:** W465-W469

Dieckmann R, PavelaVrancic M, Pfeifer E, VonDohren H, Kleinkauf H (1997) The adenylation domain of tyrocidine synthetase 1 - Structural and functional role of the interdomain linker region and the (S/T)GT(T/S)GXPKG core sequence. European Journal of Biochemistry **247:** 1074-1082

Dixon RA, Achnine L, Kota P, Liu CJ, Reddy MSS, Wang LJ (2002) The phenylpropanoid pathway and plant defence - a genomics perspective. Molecular Plant Pathology **3:** 371-390

Dunker AK, Silman I, Uversky VN, Sussman JL (2008) Function and structure of inherently disordered proteins. Current Opinion in Structural Biology **18:** 756-764

Ebel J, Schaller B, Knobloch KH, Wellman E, Grisebach H, Hahlbrock K (1974) Coordinated changes in enzyme-activities of phenylpropanoid metabolism during growth of soybean cell-suspension cultures. Biochimica Et Biophysica Acta **362:** 417-424

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research **32:** 1792-1797

Ehlting J, Buttner D, Wang Q, Douglas CJ, Somssich IE, Kombrink E (1999) Three 4-coumarate : coenzyme A ligases in Arabidopsis thaliana represent two evolutionarily divergent classes in angiosperms. Plant Journal **19:** 9-20

Ehlting J, Shin JJK, Douglas CJ (2001) Identification of 4-coumarate : coenzyme A ligase (4CL) substrate recognition domains. Plant Journal **27:** 455-465

Fillatti JJ, Sellmer J, McCown B, Haissig B, Comai L (1987) Agrobacterium mediated transformation and regeneration of Populus. Molecular & General Genetics **206:** 192-199

Fliegmann J, Schroder G, Schanz S, Britsch L, Schroder J (1992) Molecular analysis of chalcone and dihydropinosylvin synthase from scots pine (Pinus sylvestris), and differential regulation of these and related enzyme activities in stressed plants. Plant Molecular Biology **18:** 489-503

Franke R, Humphreys JM, Hemm MR, Denault JW, Ruegger MO, Cusumano JC, Chapple C (2002) The Arabidopsis REF8 gene encodes the 3-hydroxylase of phenylpropanoid metabolism. Plant Journal **30:** 33-45

Friedmann M, Ralph SG, Aeschliman D, Zhuang J, Ritland K, Ellis BE, Bohlmann J, Douglas CJ (2007) Microarray gene expression profiling of developmental transitions in Sitka spruce (Picea sitchensis) apical shoots. Journal of Experimental Botany **58:** 593-614

Fulda M, Heinz E, Wolter FP (1994) The fadD gene of Escherichia coli K12 is located closeto rnd at 39.6 min of the cgromosomal map and is a new member of the AMP-binding protein family. Molecular & General Genetics **242:** 241-249

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS (2011) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Research **40:** D1178-D1186

Gross GG, Mansell RL, Zenk MH (1975) Hydroxycinnamate-coenzyme A ligase from lignifying tissue of higher-plants - some properties and taxonomic distribution. Biochemie Und Physiologie Der Pflanzen **168:** 41-51

Gross GG, Zenk MH (1966) Darstellung und eigenschaften von coenzym A-thioestern substituierter Zimtsäuren. Z. Naturforsch **21b:** 683

Gross GG, Zenk MH (1974) Isolation and properties of hydroxycinnamate-CoA ligase from lignifying tissue of *Forsthia*. European Journal of Biochemistry **42:** 453-459

Gui JS, Shen JH, Li LG (2011) Functional characterization of evolutionarily divergent 4-coumarate: coenzyme A ligases in rice. Plant Physiology **157:** 574-586

Guindon Sp, Dufayard J-Fo, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. Systematic Biology **59:** 307-321

Hahlbrock K, Grisebach H (1970) Formation of coenzyme-A esters of cinnamic acids with an enzyme preparation cell suspension ceultures of parsley. Febs Letters **11:** 62-&

Hahlbrock K, Grisebach H (1979) Enzymic controls in the biosynthesis of lignin and flavonoids. Annual Review of Plant Physiology and Plant Molecular Biology **30:** 105-130

Hamberger B, Hahlbrock K (2004) The 4-coumarate : CoA ligase gene family in Arabidopsis thaliana comprises one rare, sinapate-activating and three commonly

occurring isoenzymes. Proceedings of the National Academy of Sciences of the United States of America **101:** 2209-2214

Hamberger Br, Ellis M, Friedmann M, de Azevedo Souza C, Barbazuk B, Douglas CJ (2007) Genome-wide analyses of phenylpropanoid-related genes in Populus trichocarpa, Arabidopsis thaliana, and Oryza sativa: the Populus lignin toolbox and conservation and diversification of angiosperm gene familiesThis article is one of a selection of papers published in the Special Issue on Poplar Research in Canada. Canadian Journal of Botany **85:** 1182-1201

Han KH, Meilan R, Ma C, Strauss SH (2000) An Agrobacterium tumefaciens transformation protocol effective on a variety of cottonwood hybrids (genus Populus). Plant Cell Reports **19:** 315-320

Harding SA, Jarvie MM, Lindroth RL, Tsai CJ (2009) A comparative analysis of phenylpropanoid metabolism, N utilization, and carbon partitioning in fast- and slow-growing Populus hybrid clones. Journal of Experimental Botany **60:** 3443-3452

Harding SA, Jiang HY, Jeong ML, Casado FL, Lin HW, Tsai CJ (2005) Functional genomics analysis of foliar condensed tannin and phenolic glycoside regulation in natural cottonwood hybrids. Tree Physiology **25:** 1475-1486

Harding SA, Leshkevich J, Chiang VL, Tsai C-J (2002) Differential substrate inhibition couples kinetically distinct 4-coumarate:coenzyme A ligases with spatially distinct metabolic roles in quaking aspen. Plant Physiol. **128:** 428-438

Hoffmann L, Besseau S, Geoffroy P, Ritzenthaler C, Meyer D, Lapierre C, Pollet B, Legrand M (2004) Silencing of hydroxycinnamoy-coenzyme A shikimate/quinate hydroxycinnamoyltransferase affects phenylpropanoid biosynthesis. Plant Cell **16:** 1446-1465

Hoffmann L, Maury S, Martz F, Geoffroy P, Legrand M (2003) Purification, cloning, and properties of an acyltransferase controlling shikimate and quinate ester intermediates in phenylpropanoid metabolism. Journal of Biological Chemistry **278:** 95-103

Holsters M, Waele D, Depicker A, Messens E, Montagu M, Schell J (1978) Transfection and transformation of Agrobacterium tumefaciens. Molecular and General Genetics MGG **163:** 181-187

Hu WJ, Harding SA, Lung J, Popko JL, Ralph J, Stokke DD, Tsai CJ, Chiang VL (1999) Repression of lignin biosynthesis promotes cellulose accumulation and growth in transgenic trees. Nature Biotechnology **17:** 808-812

Hu WJ, Kawaoka A, Tsai CJ, Lung JH, Osakabe K, Ebinuma H, Chiang VL (1998) Compartmentalized expression of two structurally and functionally distinct 4-coumarate : CoA ligase genes in aspen (Populus tremuloides). Proceedings of the National Academy of Sciences of the United States of America **95:** 5407-5412

Hu Y, Gai Y, Yin L, Wang X, Feng C, Feng L, Li D, Jiang X-N, Wang D-C (2010) Crystal structures of a Populus tomentosa 4-coumarate:CoA ligase shed light on its enzymatic mechanisms. Plant Cell **22:** 3093-3104

Humphreys JM, Chapple C (2002) Rewriting the lignin roadmap. Current Opinion in Plant Biology **5:** 224-229

Hutzler P, Fischbach R, Heller W, Jungblut TP, Reuber S, Schmitz R, Veit M, WeissenbÃ¶ck G, Schnitzler Jr-P (1998) Tissue localization of phenolic

compounds in plants by confocal laser scanning microscopy. Journal of Experimental Botany **49:** 953-965

Jørgensen K, Rasmussen AV, Morant M, Nielsen AH, Bjarnholt N, Zagrobelny M, Bak S, Møller BL (2005) Metabolon formation and metabolic channeling in the biosynthesis of plant natural products. Current Opinion in Plant Biology **8:** 280-291

Jansson S, Douglas CJ (2007) Populus: A model system for plant biology. *In* Annual Review of Plant Biology, Vol 58. Annual Reviews, Palo Alto, pp 435-458

Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. Computer Applications in the Biosciences **8:** 275-282

Kajita S, Katayama Y, Omori S (1996) Alterations in the biosynthesis of lignin in transgenic plants with chimeric genes for 4-coumarate:coenzyme a ligase. Plant and Cell Physiology **37:** 957-965

Kao YY, Harding SA, Tsai CJ (2002) Differential expression of two distinct phenylalanine ammonia-lyase genes in condensed tannin-accumulating and lignifying cells of quaking aspen. Plant Physiology **130:** 796-807

Knobloch K-H, Hahlbrock K (1975) Isoenzymes of p-coumarate: CoA ligase from cell suspension cultures of *Glycine max*. European Journal of Biochemistry **52:** 311-320

Knobloch KH, Hahlbrock K (1977) 4-coumarate - CoA ligase from cell- suspension cultures of Petroselinum hortense Hoffm - partial -purification, substrate-specificity, and further properties. Archives of Biochemistry and Biophysics **184:** 237-248

Koutaniemi S, Warinowski T, KÃ ¤rkÃ ¶nen A, Alatalo E, Fossdal C, SaranpÅ ¤Ã ¤ P, Laakso T, Fagerstedt K, Simola L, Paulin L, Rudd S, Teeri T (2007) Expression profiling of the lignin biosynthetic pathway in Norway spruce using EST sequencing and real-time RT-PCR. Plant Molecular Biology **65:** 311-328

La Camera S, Gouzerh G, Dhondt S, Hoffmann L, Fritig B, Legrand M, Heitz T (2004) Metabolic reprogramming in plant innate immunity: the contributions of phenylpropanoid and oxylipin pathways. Immunological Reviews **198:** 267-284

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. Bioinformatics **23:** 2947-2948

Lee D, Douglas CJ (1996) Two divergent members of a tobacco 4-coumarate:coenzyme A ligase (4CL) gene family - cDNA structure, gene inheritance and expression, and properties of recombinant proteins. Plant Physiology **112:** 193-205

Lee D, Meyer K, Chapple C, Douglas CJ (1997) Antisense suppression of 4-soumarate:coenzyme A ligase activity in Arabidopsis leads to altered lignin subunit composition. The Plant Cell **9:** 1985-1998

Levy CC, Zucker M (1960) Cinnamyl and coumaryl ester intermediates in the synthesis of chlorogenic aicd. Federation Proceedings **19:** 41-41

Li L, Zhou YH, Cheng XF, Sun JY, Marita JM, Ralph J, Chiang VL (2003) Combinatorial modification of multiple lignin traits in trees through multigene

cotransformation. Proceedings of the National Academy of Sciences of the United States of America **100:** 4939-4944

Lim KB, Walker CRB, Guo L, Pellett S, Shabanowitz J, Hunt DF, Hewlett EL, Ludwig A, Goebel W, Welch RA, Hackett M (2000) Escherichia coli Î±-Hemolysin (HlyA) Is Heterogeneously Acylated in Vivo with 14-, 15-, and 17-Carbon Fatty Acids. Journal of Biological Chemistry **275:** 36698-36702

Lindermayr C, Mollers B, Fliegmann J, Uhlmann A, Lottspeich F, Meimberg H, Ebel J (2002) Divergent members of a soybean (Glycine max L.) 4-coumarate : coenzyme A ligase gene family - Primary structures, catalytic properties, and differential expression. European Journal of Biochemistry **269:** 1304-1315

Lindl T, Kreuzale.F, Hahlbroc.K (1973) Synthesis of para-coumaroyl coenzyme A with a partially purified para-coumarate-CoA ligase from cell suspension cultures of soybean (*Glycine max*). Biochimica Et Biophysica Acta **302:** 457-464

Liu Q, Bonness MS, Liu M, Seradge E, Dixon RA, Mabry TJ (1995) Enzymes of B-ring-deoxy flavonoid biosynthesis in elicited cell cultures of "old man" cactus (Cephalocereus senilis). Archives of Biochemistry and Biophysics **321:** 397-404

Maria Koch B, Sibbesen O, Halkier BA, Svendsen I, Lindberg MÅ¸ller B (1995) The primary sequence of cytochrome P450tyr, the multifunctional N-hydroxylase catalyzing the conversion of L-tyrosine to p-hydroxyphenylacetaldehyde oxime in the biosynthesis of the cyanogenic glucoside dhurrin in Sorghum bicolor(L.) Moench. Archives of Biochemistry and Biophysics **323:** 177-186

Mark PY (2002) Identification of amino acid sequences that determine substrate utilization by 4-coumarate:CoA ligase (4CL) proteins in aspen (Populus tremuloides Michx.) M.S. Thesis. Michigan Technological University, Houghton, MI

Meng HB, Campbell WH (1997) Facile enzymic synthesis of caffeoyl CoA. Phytochemistry **44:** 605-608

Niggeweg R, Michael AJ, Martin C (2004) Engineering plants with increased levels of the antioxidant chlorogenic acid. Nature Biotechnology **22:** 746-754

Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin Y-C, Scofield DG, Vezzi F, Delhomme N, Giacomello S, Alexeyenko A, Vicedomini R, Sahlin K, Sherwood E, Elfstrand M, Gramzow L, Holmberg K, Hallman J, Keech O, Klasson L, Koriabine M, Kucukoglu M, Kaller M, Luthman J, Lysholm F, Niittyla T, Olson A, Rilakovic N, Ritland C, Rossello JA, Sena J, Svensson T, Talavera-Lopez C, Theiszen G, Tuominen H, Vanneste K, Wu Z-Q, Zhang B, Zerbe P, Arvestad L, Bhalerao R, Bohlmann J, Bousquet J, Garcia Gil R, Hvidsten TR, de Jong P, MacKay J, Morgante M, Ritland K, Sundberg B, Lee Thompson S, Van de Peer Y, Andersson B, Nilsson O, Ingvarsson PK, Lundeberg J, Jansson S (2013) The Norway spruce genome sequence and conifer genome evolution. Nature **advance online publication:** 1476-4687

Payyavula R, Babst B, Nelsen M, Harding S, Tsai C-J (2009) Glycosylation-mediated phenylpropanoid partitioning in Populus tremuloides cell cultures. Bmc Plant Biology **9:** 151

Pietrowska-Borek M, Stuible HP, Kombrink E, Guranowski A (2003) 4-Coumarate : coenzyme A ligase has the catalytic capacity to synthesize and reuse various (di)adenosine polyphosphates. Plant Physiology **131:** 1401-1410

Porter LJ, Hrstich LN, Chan BG (1985) The conversion of procyanidins and prodelphinidins to cyanidin and delphinidin. Phytochemistry **25:** 223-230

Ralston L, Yu O (2006) Metabolons involving plant cytochrome P450s. Phytochemistry Reviews **5:** 459-472

Rasmussen S, Dixon RA (1999) Transgene-mediated and elicitor-Induced perturbation of metabolic channeling at the entry point into the phenylpropanoid pathway. The Plant Cell Online **11:** 1537-1551

Saballos A, Sattler SE, Sanchez E, Foster TP, Xin Z, Kang C, Pedersen JF, Vermerris W (2012) Brown midrib2 (Bmr2) encodes the major 4-coumarate:coenzyme A ligase involved in lignin biosynthesis in sorghum (Sorghum bicolor (L.) Moench). The Plant Journal **70:** 818-830

Sakakibara N, Nakatsubo T, Suzuki S, Shibata D, Shimada M, Umezawa T (2007) Metabolic analysis of the cinnamate/monolignol pathway in Carthamus tinctorius seeds by a stable-isotope-dilution method. Organic & Biomolecular Chemistry **5:** 802-815

Saslowsky D, Winkel-Shirley B (2001) Localization of flavonoid enzymes in Arabidopsis roots. The Plant Journal **27:** 37-48

Schneider K, Hovel K, Witzel K, Hamberger B, Schomburg D, Kombrink E, Stuible HP (2003) The substrate specificity-determining amino acid code of 4-coumarate : CoA ligase. Proceedings of the National Academy of Sciences of the United States of America **100:** 8601-8606

Schoch G, Goepfert S, Morant M, Hehn A, Meyer D, Ullmann P, Werck-Reichhart D (2001) CYP98A3 from Arabidopsis thaliana is a 3 '-hydroxylase of phenolic esters, a missing link in the phenylpropanoid pathway. Journal of Biological Chemistry **276:** 36566-36574

Schoch GA, Morant M, Abdulrazzak N, Asnaghi C, Goepfert S, Petersen M, Ullmann P, Werck-Reichhart D (2006) The meta-hydroxylation step in the phenylpropanoid pathway: a new level of complexity in the pathway and its regulation. Environmental Chemistry Letters **4:** 127-136

Silber MV, Meimberg H, Ebel J (2008) Identification of a 4-coumarate:CoA ligase gene family in the moss, Physcomitrella patens. Phytochemistry **69:** 2449-2456

Sluiter JB, Ruiz RO, Scarlata CJ, Sluiter AD, Templeton DW (2010) Compositional analysis of lignocellulosic feedstocks. 1. Review and description of methods. Journal of Agricultural and Food Chemistry **58:** 9043-9053

Soler Ma, Serra O, Molinas M, Huguet G, Fluch S, Figueras M (2007) A genomic approach to suberin biosynthesis and cork differentiation. Plant Physiology **144:** 419-431

Soltani B, Ehlting Jr, Hamberger Br, Douglas C (2006) Multiple cis-regulatory elements regulate distinct and complex patterns of developmental and wound-induced expression of Arabidopsis thaliana 4CL gene family members. Planta **224:** 1226-1238

Sonnante G, D'Amore R, Blanco E, Pierri CL, De Palma M, Luo J, Tucci M, Martin C (2010) Novel hydroxycinnamoyl-coenzyme A quinate transferase genes from artichoke are involved in the synthesis of chlorogenic acid. Plant Physiology **153:** 1224-1238

Stafford HA (1974) The metabolism of aromatic compounds,

Stockigt J, Zenk MH (1975) Chemical synthesis and properties of hdroxycinnamoyl coenzyme A derivatives. Zeitschrift Fur Naturforschung C-a Journal of Biosciences **30:** 352-358

Stuible HP, Buttner D, Ehlting J, Hahlbrock K, Kombrink E (2000) Mutational analysis of 4-coumarate : CoA ligase identifies functionally important amino acids and verifies its close relationship to other adenylate-forming enzymes. Febs Letters **467:** 117-122

Stuible HP, Kombrink E (2001) Identification of the substrate specificity-conferring amino acid residues of 4-coumarate : coenzyme A ligase allows the rational design of mutant enzymes with new catalytic properties. Journal of Biological Chemistry **276:** 26893-26897

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution **28:** 2731-2739

Tsai CJ, Harding SA, Tschaplinski TJ, Lindroth RL, Yuan YN (2006) Genome-wide analysis of the structural genes regulating defense phenylpropanoid metabolism in Populus. New Phytologist **172:** 47-62

Tsuji Y, Chen F, Yasuda S, Fukushima K (2003) The behavior of deuterium-labeled monolignol and monolignol glucosides in lignin biosynthesis in angiosperms. Journal of Agricultural and Food Chemistry **52:** 131-134

Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao RR, Bhalerao RP, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen GL, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, Cunningham R, Davis J, Degroeve S, Dejardin A, Depamphilis C, Detter J, Dirks B, Dubchak I, Duplessis S, Ehlting J, Ellis B, Gendler K, Goodstein D, Gribskov M, Grimwood J, Groover A, Gunter L, Hamberger B, Heinze B, Helariutta Y, Henrissat B, Holligan D, Holt R, Huang W, Islam-Faridi N, Jones S, Jones-Rhoades M, Jorgensen R, Joshi C, Kangasjarvi J, Karlsson J, Kelleher C, Kirkpatrick R, Kirst M, Kohler A, Kalluri U, Larimer F, Leebens-Mack J, Leple JC, Locascio P, Lou Y, Lucas S, Martin F, Montanini B, Napoli C, Nelson DR, Nelson C, Nieminen K, Nilsson O, Pereda V, Peter G, Philippe R, Pilate G, Poliakov A, Razumovskaya J, Richardson P, Rinaldi C, Ritland K, Rouze P, Ryaboy D, Schmutz J, Schrader J, Segerman B, Shin H, Siddiqui A, Sterky F, Terry A, Tsai CJ, Uberbacher E, Unneberg P, Vahala J, Wall K, Wessler S, Yang G, Yin T, Douglas C, Marra M, Sandberg G, Peer Y, Rokhsar D (2006) The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). Science **313:** 1596 - 1604

Uhlmann A, Ebel J (1993) Molecular-cloning and expression of 4-coumarate-coenzyme A ligase, an enzyme involved in resistance reponse of soybean (Glycine Max L) against pathogen attack. Plant Physiology **102:** 1147-1156

Umezawa T (2010) The cinnamate/monolignol pathway. Phytochemistry Reviews **9:** 1-17

Vogt T (2010) Phenylpropanoid Biosynthesis. Molecular Plant **3:** 2-20

Voo KS, Whetten RW, O'Malley DM, Sederoff RR (1995) 4-coumarate:coenzyme A ligase from loblolly pine xylem (isolation, characterization, and complementary DNA cloning). Plant Physiol. **108:** 85-97

Wagner A, Donaldson L, Kim H, Phillips L, Flint H, Steward D, Torr K, Koch G, Schmitt U, Ralph J (2009) Suppression of 4-Coumarate-CoA Ligase in the Coniferous Gymnosperm Pinus radiata. Plant Physiol. **149:** 370-383

Walton E, Butt VS (1971) The demonstration of cinnamyl-CoA synthetase activity in leaf extracts. Phytochemistry **10:** 295-304

Wang H-Q, Tuominen LK, Tsai C-J (2010) SLIM: a sliding linear model for estimating the proportion of true null hypotheses in datasets with dependence structures. Bioinformatics **27:** 225-231

Weng J-K, Akiyama T, Bonawitz ND, Li X, Ralph J, Chapple C (2010) Convergent evolution of syringyl lignin biosynthesis via distinct pathways in the Lycophyte selaginella and fowering plants. Plant Cell **22:** 1033-1045

Winkel-Shirley B (1999) Evidence for enzyme complexes in the phenylpropanoid and flavonoid pathways. Physiologia Plantarum **107:** 142-149

Winkel-Shirley B (2001) Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. Plant Physiology **126:** 485-493

Winkel BSJ (2004) Metabolic channeling in plants. Annual Review of Plant Biology **55:** 85-107

Yamauchi K, Yasuda S, Fukushima K (2002) Evidence for the biosynthetic pathway from sinapic acid to syringyl lignin using labeled sinapic acid with stable isotope at both methoxy groups in Robinia pseudoacacia and Nerium indicum. Journal of Agricultural and Food Chemistry **50:** 3222-3227

Yamauchi K, Yasuda S, Hamada K, Tsutsumi Y, Fukushima K (2003) Multiform biosynthetic pathway of syringyl lignin in angiosperms. Planta **216:** 496-501

Zhang XH, Chiang VL (1997) Molecular cloning of 4-coumarate:coenzyme A ligase in loblolly pine and the roles of this enzyme in the biosynthesis of lignin in compression wood. Plant Physiol. **113:** 65-74

Zhong R, Morrison WH, Negrel J, Ye Z-H (1998) Dual methylation pathways in lignin biosynthesis. The Plant Cell Online **10:** 2033-2045