THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF BIOLOGY

GENOME WIDE ANALYSES OF THE CAFFEIC ACID O-METHYLTRANSFERASE (COMT) GENE FAMILY IN *POPULUS*

ALEX CHOI
Spring 2010

A thesis
submitted in partial fulfillment
of the requirements
for baccalaureate degrees
in Biology and Psychology
with honors in Biology

Reviewed and approved* by the following:

Abdelali Barakat
Research Associate
Thesis Supervisor

John E. Carlson
Professor of Molecular Genetics
Director of the Schatz Center
Thesis Supervisor

Sarah Assmann
Waller Professor of Biology
Honors Adviser

* Signatures are on file in the Schreyer Honors College.

Abstract

Background

Monolignols are basic components of lignin; their synthesis involves a dozen enzymes. Caffeic acid o-methyltransferase (COMT) is one of the key enzymes involved in the synthesis of monolignols. The COMT gene family has been studied in several species, including model and non-model plants. Here we analyzed the organization, the phylogeny, and the expression of *COMT* genes in various tissues from non-stressed and stressed *Populus* plants.

Results

We identified 8 *COMT* genes from *Populus* and found that all of them are located in duplicated blocks. *COMT* genes are distributed in two phylogenetic classes with the real *COMT* gene distributed in Class I. Expression profiling showed that *COMT* genes present various expression profiles in leaves, bark, and xylem from poplar. One gene, which was previously described as a real *COMT* gene, showed preferential expression in xylem tissue. The others are either preferentially expressed in bark or show no differential expression.

Conclusions

COMT genes are highly duplicated and most of duplicates are still in conserved positions on homeologous blocks wthin the *Populus* genome. Expression profiling showed that *COMT* genes present different expression profiles in *Populus* tissues with only one gene differentially expressed in xylem. Our results suggest that *COMT* genes have evolved various expression profiles and may have evolved different functions following duplications.

TABLE OF CONTENTS

Abbreviations:

COMT, caffeic acid o-methyltransferase; nt, nucleotide; aa, amino acid; RT-PCR, reverse transcriptase polymerase chain reaction; CCoAOMT, caffeoyl CoA-methyltransferase; FAH, ferulic acid hydroxylase; SAM, S-adenosyl-L-methionine; AEOMT, hydroxycinnamic acids/hydroxycinnamoyl CoA esters o-methyltransferase; in, intron; ex, exon

Introduction

Lignin is a phenolic compound in plant cell walls which functions in maintaining the structural rigidity of plants and defending them against pests and pathogens. It is the second most abundant biopolymer on earth following cellulose (Raes et al., 2003). Lignin is composed mainly of three monolignols: *p*-coumaryl, coniferyl, and sinapyl alcohol (Ye and Varner, 1995) which are the basic units of G (guaiacyl), H (p-hydroxyphenyl), and S (syringyl) lignins respectively. Lignin varies in content and composition between gymnosperms and angiosperms. In the former, lignin contains G and H units while in angiosperms, it comprises G-units, H-units, (Whetten et al., 1998), and S-units. The synthesis of S-units is performed by methylation of 5-hydroxyconiferyl alcohol into sinapaldehyde or sinapyl alcohol (Ye et al., 1994; Zhong et al., 1998) under the action of the COMT enzymes. COMT is an S-adenosyl-L-Met-dependent O-methyltransferase that includes both COMT and caffeoyl CoA o-methyltransferase (CCoAOMT) gene families. While both enzymes are involved in lignin biosynthesis, CCoAOMT intervenes in earlier steps of the lignin biosynthesis pathway by transforming caffeoyl CoA to Feruloyl CoA (Davin and Lewis, 1992; Ye et al., 1994). COMT acts at the end of the pathway to produce sinapyl alcohol which is the main component of S lignin.

Previous studies (Anterola et al., 2002; Raes et al., 2003; Tuskan et al., 2006) enabled the identification of *COMT* genes from several model and non-model plant species. An early study (Collendavelloo et al., 1981) on COMT in tobacco allowed the identification of three distinct enzymes (COMT I, COMT II, COMT III) based on substrate specificity. Further analyses classified *COMT* genes into two functional classes. Class I genes include tobacco COMT I while Class II includes COMT II and III (Jaeck et al., 1992; Pellegrini et al., 1993). Class I genes are

involved in lignin biosynthesis whereas Class II genes are involved in plant defense against biotic and abiotic stresses (Pellegrini et al., 1993). This distribution was confirmed by another study (Joshi and Chiang, 1998) that showed that classified *OMT* genes are into two distinct groups, PL-OMT I and PL-OMT II. The PL-OMT I group is involved in lignin biosynthesis and uses only a pair of substrates. The members of the PL-OMT II group use a variety of substrates, such as caffeic acid (CA), 5-hydroxyferulic acid (5HFA), caffeoyl-CoA ester, 5-hydroxyferuloylester, *myo*-inositol, chalcones and scoulerine (Joshi and Chiang, 1998). Three conserved sequence motifs specific to S-adenosyl-L-methionine (SAM) were discovered in several plants (Joshi and Chiang, 1998). Further studies classified sequences based on sequence similarity to real *COMT* genes. For instance, in *Arabidopsis*, one real *COMT* and 13 *COMT-like* genes were reported (Raes et al., 2003). In *Populus*, two *COMT* genes and two *COMT-like* genes were reported with the initial publication of the genome sequence (Tuskan et al., 2006).

In *Arabidopsis*, *Populus*, and *Medicago*, *COMT* genes have been shown to act as a bi-functional enzyme (Vander Mijnsbrugge et al., 2000; Raes et al., 2003; Kota et al., 2004). Indeed, while COMT proteins can convert *in vitro* caffeic acid to sinapoylmalate (Hoffmann et al., 2004), their primary role is mainly in lignin biosynthesis. COMT proteins methylate 5-hydroxyconiferaldehyde and/or 5- hydroxyconiferyl alcohol to sinapaldehyde and/or sinapyl alcohol, respectively (Osakabe et al., 1999; Li et al., 2000; Parvathi et al., 2001). Many studies (Parvathi, 2001; Maury, 1999; Meng, 1998; Inoue et al., 2000; Zubieta et al., 2002) have investigated the substrate specificity of COMT proteins. In Medicago (Inoue et a.l, 2000), a previously characterized COMT I gene showed higher substrate preference towards 5HFA than CA, whereas COMT II, a novel OMT expressed predominantly is young internodes, showed the

opposite substrate preference. Moreover, in transgenic *Medicago*, down regulation of COMT resulted in reducing the activity against the non-esterified substrates in the following order: 5-hydroxyconiferyl acid, 5HFA and caffeoyl alcohol, caffeoyl aldehyde, caffeic acid, and 5-hydroxyconiferaldehyde, suggesting that COMT is unlikely to methylate caffeic acid and providing evidence of an alternate pathway to S lignin biosynthesis (Parvathi et al., 2001; Guo, 2001). This can be due to the fact that *Medicago* COMT protein has an unusually large catalytic site, accommodating a broad range of substrates (Zubieta et al., 2002).

Functional analyses of *COMT* genes showed that mutation of *COMT* genes can affect the monolignol type and the content of lignin as well as plant phenotypes (Tsai et al., 1995; Lapierre et al., 1999; Jouanin et al., 2000). Down-regulation of *COMT* genes in a *Populus* hybrid (*Populus tremula* x *Populus alba*) resulted in a nearly complete suppression of COMT activity and a 17% decrease in lignin composition compared to control plants (Jouanin et al., 2000). Similar results were obtained by down-regulating *COMT* genes in *Medicago* (Guo et al., 2001) and maize (Piquemal et al., 2002) where the lignin amount was decreased by 30%. In other studies (Van Doorsselaere et al., 1995a; Lapierre et al., 1999), down-regulation of *COMT* genes led to a change in the G/S lignin ratio. Moreover, down-expression of *COMT* genes using an antisense strategy in tobacco (Atanassova et al., 1995), maize (Piquemal et al., 2002), *Medicago* (Guo et al., 2001; Weeks et al., 2008), and *Populus* (Jouanin et al., 2000) showed decreased or near loss of S lignin units, accompanied by either a decrease or increase in 5-hydroxy-G units. A recent study on expression profiling of hundreds of genes in the COMT natural mutant bm3 and in transgenic plants expressing a COMT antisense construct (*AS225*) (Guillaumie et al., 2008) showed a disturbance in cell wall assembly. In sorghum, the bmr-12 mutant showed a 22%

decrease in the amount of lignin at the whole plant level (Vogler et al., 2009). The bmr-12 mutant also presented features of down-regulated *COMT* genes such as a decrease in the lignin content, decrease of the S/G lignin ratio, and the occurrence of unusual 5-OH guaiacyl units (Akin et al., 1986 ; Pillonel et al., 1992; Halpin et al., 1994 ; Lam et al., 1996; Hatfield et al., 1999; Bout and Vermerris, 2003).

COMT genes were reported as being involved in plant defense against various biotic and abiotic stresses (Toquin et al., 2003; Lee et al., 1997). Monolignol biosynthesis is a crucial process for cell wall apposition, one of the first lines of plant defense against invading fungi. For instance, a study showed that a *COMT* gene (*TmCAOMT*) is involved in wheat plant defense against powdery mildew invasion (Bhuiyan et al., 2008). The silencing of this gene was very effective in compromising the penetration resistance of both host and non-host pathogens (*Blumeria graminis f. sp. Tritici* and *B. gramini f. sp. Hordei*, respectively) (Bhuiyan et al., 2008). In *Arabidopsis*, treatment of plants with green leafy volatiles or isoprenoids such as (E)-2-gexenal and (Z)-3-hexenal treatment, which induce several resistance genes including *COMT*, resulted in slower rate of disease development when inoculated with *Botrytis cinerea* (Kishimoto et al., 2005). In tobacco, down-regulation of COMTI (real COMT) and CCoAOMT resulted in plants with large necrotic lesions following tobacco mosaic virus (TMV) infection (Maury et al., 1999; Hoffmann et al., 2000). The expression of a reporter gene (GUS) that was under the control of OMT promoters confirmed the reduction of Class II tobacco COMT gene (COMTII) in response to biotic and abiotic stresses (Toquin et al., 2003).

An earlier phylogenetic analysis of plant OMT sequences showed that they clustered within a monophyletic group derived from non-plant genes (Ibrahim et al., 1998). This study

showed the presence of two groups clustering according to a functional trait that reflects their substrate specificity. Other phylogenetic analysis showed a distribution of *COMT* genes in two groups (COMT and COMT-*like*), of which one included a real *COMT* gene (Raes et al., 2003). The COMT-*like* group included *COMT* genes clustering with hydroxycinnamic acids/hydroxycinnamoyl CoA ester o-methyltransferase (AEOMT) (Li et al., 1997 ; Li et al., 1999). Similarly, other study (Li et al., 2006) showed evidence of two groups of *COMT* genes. While these studies provided some insight into the phylogenetic relationship between *OMT* genes, they all used a limited set of data including mainly monocots and eudicots.

In this study we retrieved and annotated COMT sequences from a variety of plants covering all land plant lineages including *Physcomitrella* and *Selaginella*. We used this set to analyze the phylogeny of the *COMT* genes. In this study we investigated the gene structure, the genome organization, and the expression of *COMT* genes in *Populus*.

Methods

Plant materials

Leaves, cortex, and xylem were collected from young hybrid *Populus* OGY (*P. deltoides* x *P. nigra*) trees grown in a culture chamber at 25°C and 18°C in the day and night, respectively. The plants were grown at 16h/8h day/night regime and at 60% humidity. Herbivory stress treatment using *Lymantria dispar*s (*Lymantria dispar*) larvae was described previously (Barakat et al., 2010). Tissues were harvested and frozen in liquid nitrogen and stored at -80°C until use.

RNA isolation and cDNA synthesis

Total RNA was isolated using a CTAB method (Chang et al., 1993) with minor modifications. The RNA quality and concentration was assessed using an Agilent 2100 Bioanalyzer (Agilent Technologies). cDNA synthesis was performed as described previously (Barakat et al., 2009).

COMT sequences from model species

COMT sequences used in phylogenetic analyses include sequences from plants with fully sequenced genomes as well as other taxons representing key positions on the angiosperm phylogenetic tree. COMT sequences from *Arabidopsis*, *Oryza*, and *Populus* (Appendix A) were retrieved from TAIR, TIGR (Craig Venter Center), and Joint Genome Institute (JGI). COMT sequences from various other genomes *Carica papaya*, *Vitis vinifera*, and *Medicago truncatula*, *Sorghum bicolor*, *Physcomitrella patens*, *Selaginella moellendorffii* were identified by querying the genome sequences. COMT sequences from various non model species including gymnosperms were retrieved from TIGR Plant Genomics databases, Genbank, and a previous

study (Lam et al., 2007). Sequences were carefully inspected and corrected for annotation errors before use. Only sequences that present motifs A ((V/I/L)(V/L)(D/K)(V/I)GGXX(G/A)), B ((V/I/F)(A/P/E)X(A/P/G)DAXXXK(W/Y/F)), and C ((A/P/G/S)(L/I/V)(A/P/G/S)XX(A/P/G/S)(K/R)(V/I)(E/I)(L/I/V)) described previously (Joshi and Chiang, 1998) were considered in this study.

Intron-exon structure and promoter analysis of *COMT* genes

The intron (in)-exon (ex) structure of *COMT* genes was retrieved from the Joint Genome Institute http://www/jgi.doe.gov web site. For genes having complementary DNA (cDNA) sequences available, the structure was checked by aligning genomic and cDNA sequences. Promoter analysis was done by querying all *COMT* genes against TRANSFAC (Wingender et al., 1996) and PlantCARE (Lescot et al., 2002).

COMT sequence alignment and phylogenetic analyses

COMT nucleotide (nt) sequences were translated into protein sequences. The inferred protein sequences were then aligned using Muscle with default parameters (Edgar, 2004), and manually adjusted. Phylogenetic analyses were performed on the aligned amino acid (aa) sequences, as well as on the nt sequences that were aligned to match the aas. The WAG model (Whelan and Goldman, 2001), assuming among site rate heterogeneity (WAG+G), was used for the aa sequences. Maximum Likelihood (ML) analyses were implemented in PHYML v. 2.4.4 (Guindon and Gascuel, 2003) using 100 bootstrap replicates to estimate branch support.

COMT expression analysis using quantitative real time RT-PCR

Quantitative real-time reverse transcriptase polymerase chain reactions (RT-PCR) were performed in an Applied Biosystems 7500 Fast Real-Time PCR system (Applied Biosystems) with default parameters as described previously (Barakat et al., 2009). Primers used in this study were designed using Primer Express® software (Applied Biosystems). We used the gene encoding the 18S rRNA as an endogenous control to normalize for template quantity. For each gene, three biological replicates (three different trees) and three experimental replicates were used. Data was evaluated using the 7500 Fast System SDS software procedures (Applied Biosystems). Statistical analyses were performed using Statistica 6.0 software (StatSoft Poland Inc., Tulsa, OH, USA).

Results

COMT gene family organization

Sequence similarity and motifs described previously (Joshi and Chiang, 1998) allowed us to annotate 176 OMT genes in several model and non-model plants. Utilizing additional selection methods from a previous study (Lam et al., 2007), 66 of the 176 OMT sequences were described as *COMT* genes. 66 (100%) sequences include motif A and B. There are four (6.06%) sequences that showed one aa insertion inside motif B. Motif C was found in 66 (100%) of the sequences. 4 sequences (6.06%) showed insertions ranging from one to two aa within this motif. In model species for which the genome is completely sequenced, 66 *COMT* genes have been identified so far (Appendix A): 13 from *Arabidopsis*, 1 from *Carica*, 7 from *Medicago*, 4 from *Oryza*, 8 from *Populus*, 6 from *Selaginella*, 5 from *Sorghum*, 7 from *Vitis*, and 2 from *Physcomitrella*. Furthermore, we identified 8 *COMT* genes from gymnosperms. Three sequences (*LpeCOMT1*, *ZelCOMT1*, *ZmaCOMT1*) from (Lam et al., 2007) were included. *COMT* gene names in Appendix A include the name of species (Poptr for *Populus trichocarpa* for example), the protein name (COMT), and a number to indicate members of the family, accession numbers, and database sources.

Analysis of the gene distribution in the *Populus*, the *Arabidopsis*, and the *Oryza* genomes showed that most *COMT* genes are located on duplicated blocks. In *Populus*, the eight genes (100%) were mapped to 6 chromosomes: I, II, XI, XII, XIV, and XV (Fig. 1). All eight of the mapped genes (100%) were located on duplicated blocks. Duplicate gene pairs (*PoptrCOMT3 - PoptrCOMT8; PoptrCOMT3/PoptrCOMT5 – PoptrCOMT2/PoptrCOMT4; PoptrCOMT1 – PoptrCOMT6*) were still located on conserved positions on homeologous duplicated blocks. Two

pairs of *Populus COMT* genes (*PoptrCOMT3- PoptrCOMT5; PoptrCOMT2-PoptrCOMT4*) were distributed one after the other and seem to have been generated through tandem duplication.

A similar situation was found for *Arabidopsis* where all *COMT* genes (100%) were either duplicated in tandem or located on large duplicated blocks. Three sets of *Arabidopsis COMT* genes (*AthCOMT1, AthCOMT2, AthCOMT3, AthCOMT4*) (*AthCOMT9, AthCOMT10*) and (*AthCOMT12* and *AthCOMT13*) were distributed in tandem on chromosome I, chromosome II, and chromosome V respectively. In *Oryza*, none of the four identified *COMT* genes were located on duplicated blocks.

Intron-exon structure of *COMT* genes

Gene structure analysis of *Populus COMT* genes (Fig. 2) showed the existence of two patterns of intron-exon structure: Pattern 1 (4ex/3in) and pattern 2 (4ex/3in). While genes within these patterns showed similar size of exons, introns showed significant variations in length that could be associated with transposable element insertions. Pattern 2 was found in *Populus*, *Arabidopsis*, and rice. Pattern 1 exists only in *Populus* and seems to be derived from Pattern 2. One homeologous duplicate pair (*PoptrCOMT7 – PoptrCOMT8*) showed similar intron-exon structures (Fig. 2). Another pair (*PoptrCOMT1 – PoptrCOMT6PoptrCOMT6*) showed divergent structures. The third pair (*PoptrCOMT3/PoptrCOMT5 – PoptrCOMT2/*PoptrCOMT4) shared the same structure except *PoptrCOMT4*. One tandem duplicated gene pair (*PoptrCOMT3-PoptrCOMT5*) also showed similar intron-exon structures, while the other pair (*PoptrCOMT2-PoptrCOMT4*) showed divergent intron-exon structures.

Evolution of *COMT* genes

Maximum Likelihood (ML) phylogenetic tree using protein sequences (Fig. 6) showed that COMT genes were distributed in two major clades (Classes). These clades are supported by high bootstrap value (100). Class I includes two sub-clades. The first group includes eudicot, monocot, and gymnosperm sequences, whereas the second group includes eudicot and monocot sequences. Class II includes monocot sequences. The first sub-clade of Class I includes the previously characterized *Populus* (*PoptrCOMT7* and *PoptrCOMT8*) and *Arabidopsis* (*AthCOMT13*) real *COMT* genes (Raes et al., 2003; Tuskan et al., 2006). The second sub-clade includes the previously characterized *Arabidopsis* COMT-like genes (*AthCOMT1*, *AthCOMT2*, *AthCOMT3*, *AthCOMT4*, *AthCOMT5*, *AthCOMT6*, *AthCOMT7*, *AthCOMT8*, *AthCOMT9*, *AthCOMT10*, *AthCOMT11*, and *AthCOMT12*). The phylogeny shared evidence of extensive duplication events within each species. A pair of *Sorghum* and *Physcomitrella* sequences represent two OMT sequences used for rooting purpose.

Expression analysis of *Populus COMT* genes

The expression of all eight *COMT Populus* genes was analyzed using quantitative real-time RT-PCR. These 8 genes were chosen from different classes (clades) and include both duplicates generated either by tandem or segmental duplications to address various evolutionary questions. The results showed that all the *COMT* genes studied are expressed in leaves, bark, and xylem (Fig. 3). Based on the expression in the tissues studied, the expression patterns of *COMT* genes were classified into five different groups. Group 1 gene (*PoptrCOMT8*) was preferably expressed in xylem. *PoptrCOMT8* is eight times more highly expressed in the xylem than the

other tissues. ANOVA statistical analysis showed that the expression differences between tissues were statistically significant. Group 2 (*PoptrCOMT2* and *PoptrCOMT4*) includes genes that show preferential expression in bark tissue. The expression level of these genes in leaves is four to six times higher than the ones in xylem and bark. Group 3 (*PoptrCOMT7, PoptrCOMT3, PoptrCOMT6,* and *PoptrCOMT5*) did not show any differential expression in any tissue. The duplicate genes (*PoptrCOMT7 - PoptrCOMT8; PoptrCOMT1 – PoptrCOMT6; PoptrCOMT3/PoptrCOMT5 – PoptrCOMT2/PoptrCOMT4)* showed divergent profiles. Indeed, *PoptrCOMT7* is highly expressed in leaves and bark, while *PoptrCOMT8* is highly expressed in xylem. *PoptrCOMT1* showed high expression in bark tissue while *PoptrCOMT6* showed no differential expression. *PoptrCOMT3* and *PoptrCOMT5* showed high expression in bark tissue, whereas *PoptrCOMT2* and *PoptrCOMT4* showed no differential expression. In herbivory stressed plants, both the *PoptrCOMT7* and *PoptrCOMT8* genes showed significant changes in their expression. *PoptrCOMT7* showed an increase in expression within leaf tissue, whereas *PoptrCOMT8* showed a decrease in expression within xylem tissue.

Discussion

Sequence analyses allowed us to identify 26 *OMT* genes from *Populus*, of which eight were *COMT* genes. A relatively low number of genes were identified from *Arabidopsis* (12), rice (4) and other model species. It is also in accordance with previous studies on Apple (*Malus* × *domestica*) reporting seven *COMT* genes (Han et al., 2007). 28 and 30 OMTs have been annotated in the *Arabidopsis* and rice genomes, respectively. The high number of *OMT* genes in *Populus* is in accordance with the high number of genes identified from *Arabidopsis* and rice. This indicates that the *OMT* gene family is encoded by a small gene family that was generated via various duplication-retention events. Genome organization showed that all of the *COMT* genes were located on duplicated blocks (100%). A similar situation was found in *Arabidopsis* where all *COMT* genes were located on duplicated blocks. Tandem duplication seems to have contributed to the duplication of 2 *COMT* genes (20%). This is in accordance with data published previously (Tuskan et al., 2006) showing that 4839 genes representing 15.6% of *Populus* genome were tandemly duplicated. A similar situation was found for *Arabidopsis* where 1366 (27%) genes were distributed in tandem (Tuskan et al., 2006). Phylogenetic distribution of tandem duplicates showed that some tandem duplication events preceded the split of *Arabidopsis* and *Populus*. For instance, duplicate pair *PoptrCOMT7* and *PoptrCOMT8* seems to be generated by a duplication that may have occurred in the ancestor of Eurosid (Tuskan et al., 2006). This duplication mechanism seems to have contributed significantly to the duplication of *COMT* genes in other species such as Apple (*Malus* × *domestica*) where five *COMT* genes duplicated in tandem were reported (Han et al., 2007). In contrast, rice presents a lower number of duplicate genes; which may be due to a higher evolution rate of the rice genome and the *Poaceae*.

Phylogenetic distribution showed that *COMT* genes are distributed in two classes. Class I includes all previously described real *COMT* genes (*PoptrCOMT7*, *PoptrCOMT8*, *AthCOMT13*) (Raes et al., 2003; Tuskan et al., 2006). Class II contains monocot, eudicot, and gymnosperm sequences. This indicates that these Classes were generated by duplication in the ancestor of land plants. In Class I, two sub-clades exist. The first sub-clade includes monocot, eudicot, and gymnosperm sequences and all three previously described real COMT genes (*PoptrCOMT7*, *PoptrCOMT8*, *AthCOMT13*) (Raes et al., 2003; Tuskan et al., 2006). The second sub-clade includes eudicot and monocot sequences, including the previously described *Arabidopsis COMT-like* genes (Raes et al., 2003). They could correspond to derived sequences that have evolved after the split of eudicots and monocots. The disproportionate distribution of *Arabidopsis COMT* genes is supported by previous study (Raes et al., 2003). Unfortunately, none of the *Arabidopsis COMT-like* genes have been characterized in terms of biochemical conversions and expression profiles (Costa et al., 2003). The phylogenetic tree shows evidence of extensive duplication within each plant species. Moreover, there is evidence of several species-specific duplications revealed by comparing the genome organization of COMT in *Arabidopsis*, rice, *Medicago*, V*itis* etc. It also showed that several duplication events were genus or clade-specific. For instance, there is a duplication event that seems to be shared by all species from Eurosid I.

Intron-exon structure showed that *COMT* genes present two major patterns (Pattern 1 and Pattern 2). Pattern 2 was found in *Arabidopsis*, *Populus* and rice, suggesting it could be ancestral. Indeed, Pattern 2 is also found in a *Physcomitrella* sequence (*PpaCOMT1*). Another *Physcomitrella* sequence (*PpaCOMT2*) shows a 3ex/2in where the length of exon 1 almost

equals the sum of exon 1 and 2 of Pattern 2 (Fig. 2). These results support the possibility of Pattern 2 being the ancestral intron-exon pattern of *COMT* genes. Phylogenetic distribution shows that genes with Pattern 2 and Pattern 3 belong to Class II while genes with the remaining patterns belong to Class I. Pattern 1 was found only in *Populus* and seems to be a derived pattern that evolved recently. A similar situation was found in *Arabidopsis* and rice where several recently derived patterns were found.

Expression analyses showed that *COMT* genes present different expression patterns among *Populus* tissues with only one gene (*PoptrCOMT8*) preferentially expressed in xylem. *PoptrCOMT8* corresponds to one of the real *COMT* genes reported previously (Tuskan et al., 2006). These results suggest that *COMT* genes have evolved different expression profiles in different tissues and under different physiological conditions. This is in accordance with previous studies on *COMT* genes from *P. kitakamiensis* (Hayakawa et al., 1996) and apple (Han et al., 2007) where differences in gene expression were observed in the different plant tissues studied. While real *COMT* genes are involved in monolignol biosynthesis, the other genes may be involved in defense against biotic and abiotic stresses. Indeed, *OMT* genes (phenol OMT and flavonoid OMTs) retrieved from a previous study (Lam et al., 2007) were distributed exclusively in Class I along with several *OMT* genes that were differentially expressed in bark and leaf tissues. The large number of genes differentially expressed in leaves and bark is an indication that most of the *COMT-like* genes may be involved in stress defense. This is in agreement with previous studies that suggested that *COMT-like* genes from Class I were involved in plant defense (Stintzi et al., 1993; Fritig et al., 1998). The differential expression of *COMT* genes in various plant tissues is thought to be mainly due to regulation at the promoter level (Toquin et al.,

2003). This is in accordance with a previous study suggesting that *COMT* gene expression is regulated at the transcriptional level during development as well as defense under stress conditions (Ni et al., 1996).

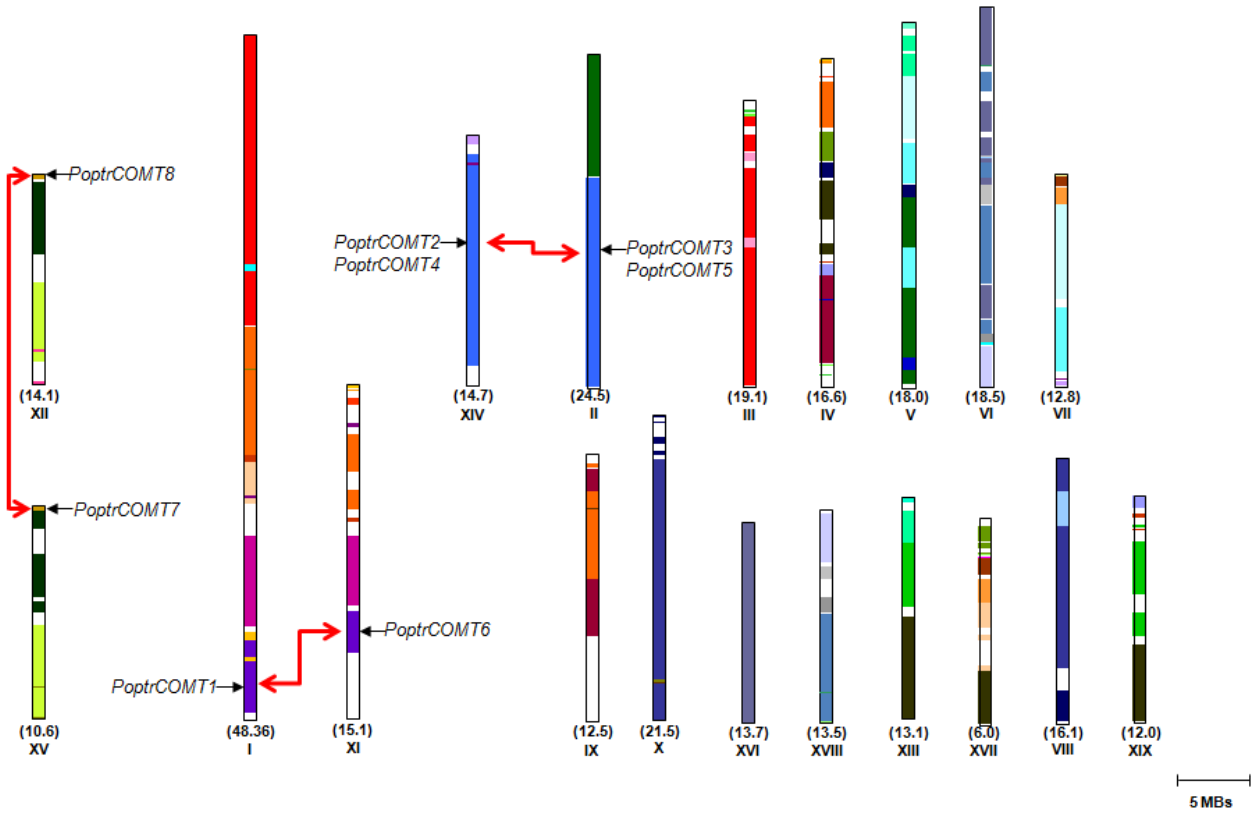

Figure 1. Distribution of *COMT* genes on *Populus* chromosomes. All *COMT* genes not shown in the figure are located on scaffolds that were not mapped yet on *Populus* chromosomes. The names of the chromosomes and their sizes (Mb) are indicated below each chromosome. Segmental duplicated homeologous blocks are indicated with the same color. The position of genes is indicated with an arrowhead.

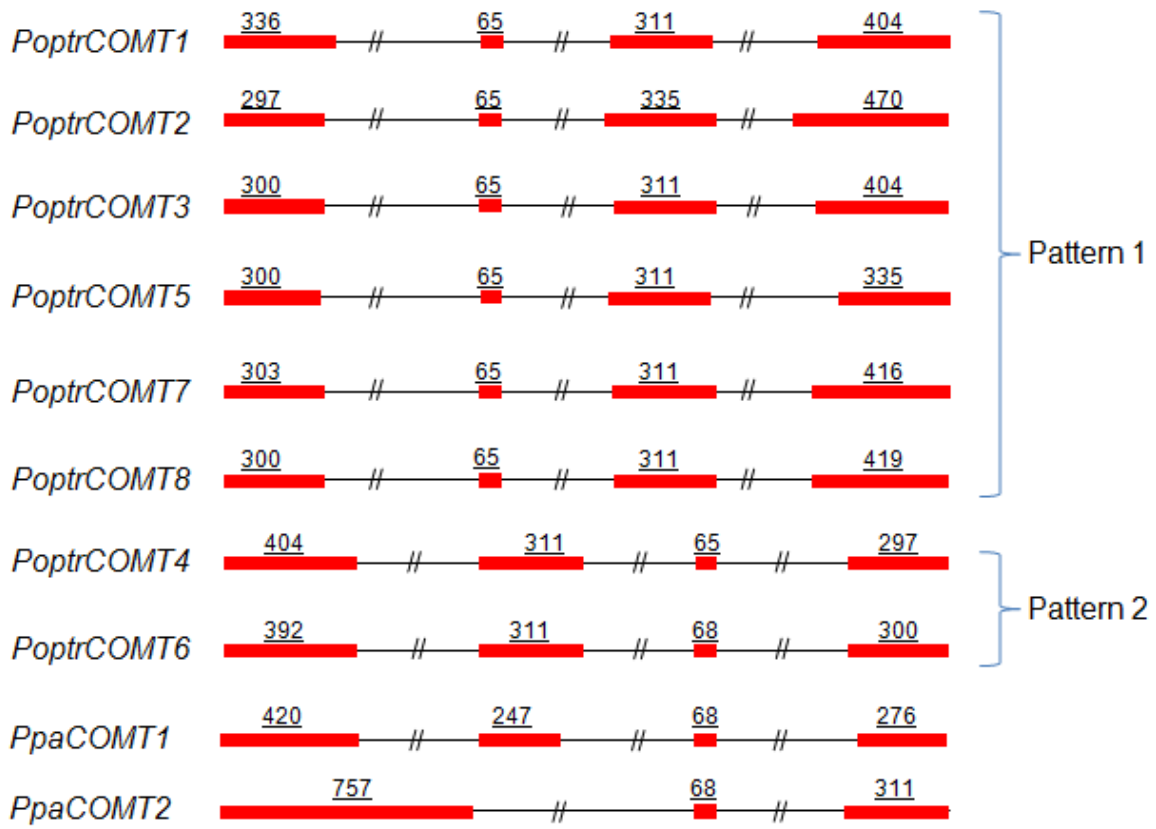

Figure 2. Intron-exon structures of *COMT* genes from *Populus*. Exons and introns are represented by boxes and lines, respectively. Numbers above boxes indicate the exon size. The intron sizes are not to scale. The names of *COMT* genes and intron-exon structure pattern are indicated at the left and right sides respectively. Two *Physcomitrella* intron-exon structures (*PpaCOMT1* and *PpaCOMT2*) indicate the ancestral structure (Pattern 2).

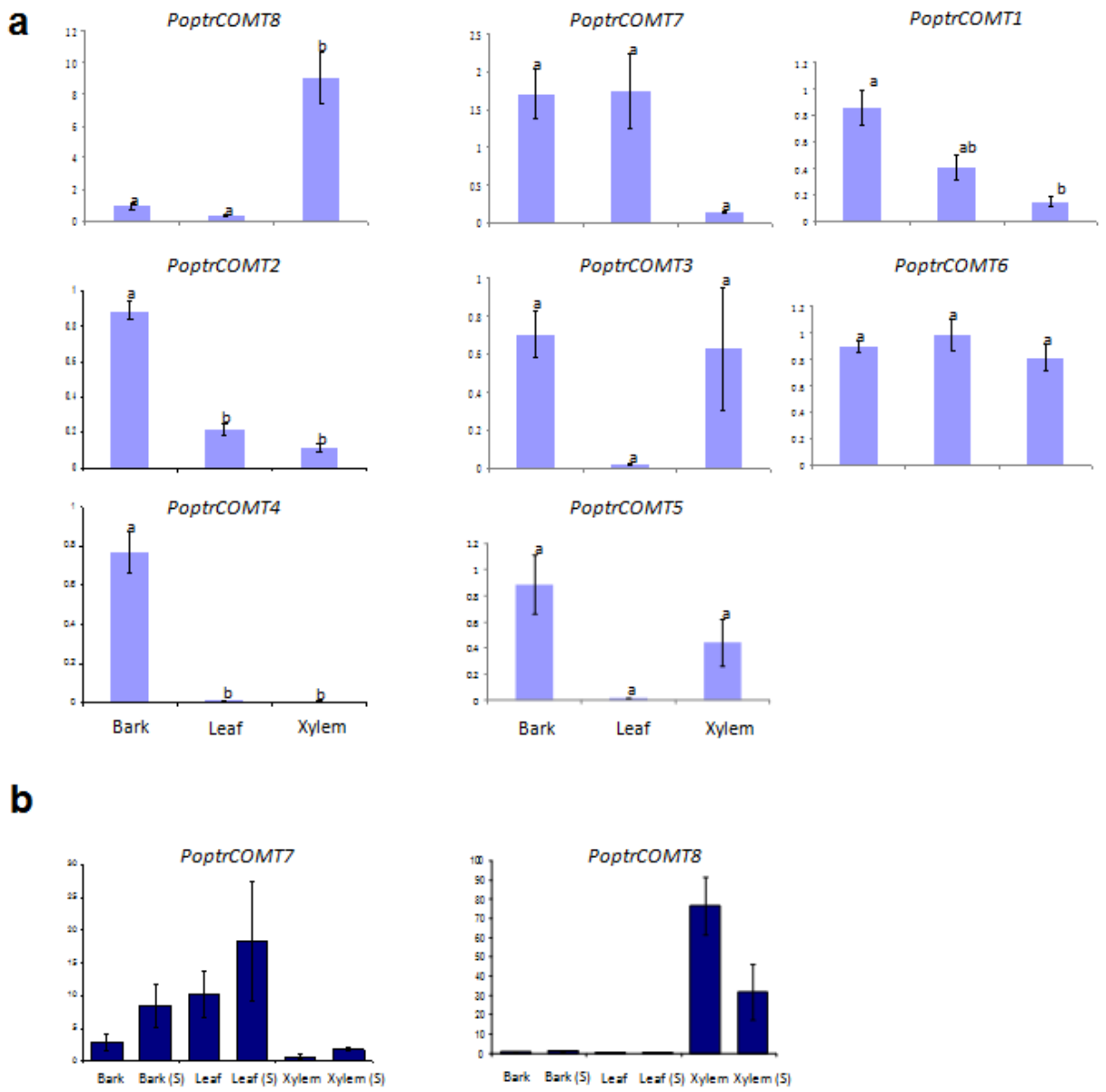

Figure 3. Quantitative expression of *Populus COMT* genes in non-stressed and stressed *Populus*. The name of each gene is indicated at the top of each histogram. Tissues studied are shown at the bottom of the diagrams. Means designated by the same letter do not differ significantly according to Tukey's HSD test; $P < 0.05$.

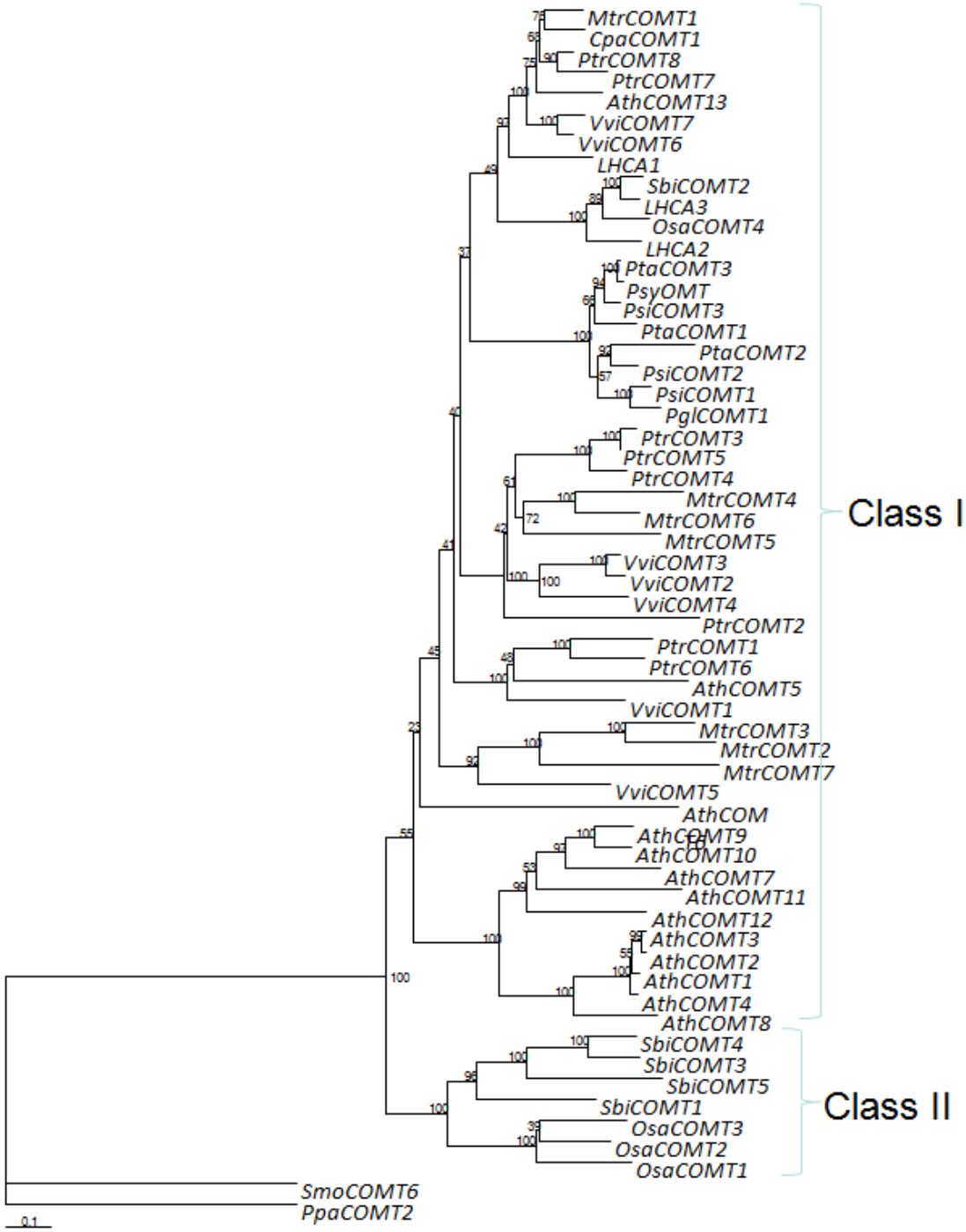

Figure 4. Maximum Likelihood bootstrap tree phylogeny based on amino acid sequences of *COMT* genes in various land plants. Numbers above branches refer to NJ bootstrap values.

Acknowledgements

I thank The Joint Genome Institute for providing access to *Populus trichocarpa* genome sequences. I thank Dr. Dawn Luthe for access to the real time RT-PCR machine. I thank Joseph Park for his help with RT-PCR experiments and for assisting in preparing the figures. I thank Dr. Chris Frost for performing the herbivore treatments. This work was supported by The Schatz Center for Tree Molecular Genetics at Penn State. Last, I thank the wonderful friends I have at Pennsylvania State University and my mother and father, who have sent unconditional support in all my endeavors.

Author Contribution

AB retrieved, curated, annotated, and aligned the COMT nucleotide and protein sequences. He prepared RNA, analyzed the gene structure, ran the phylogenetic analyses, supervised AC, and NB, and wrote the manuscript. AC contributed to the gene structure, curating protein and nucleotide sequences, the expression analyses, and helped drafting the manuscript. NB contributed to the RNA preparation and the expression analyses. ZS helped with the expression analyses. This project was initiated by AB and JC. JC directs The Schatz Center for Tree Molecular Genetics at Penn State which funded the project. He funded the project and assisted in the preparation of the manuscript.

References

Akin D, Hanna W, Snook M, Himmelsbach D, Barton F, Windham W (1986) Normal-12 and brown midrib-12 sorghum. II. Chemical variations and digestibility Agron J **78** 832-837

Anterola AM, Jeon JH, Davin LB, Lewis NG (2002) Transcriptional control of monolignol biosynthesis in Pinus taeda: factors affecting monolignol ratios and carbon allocation in phenylpropanoid metabolism. J Biol Chem. **277:** 18272-18280

Atanassova R, Favet N, Martz F, Chabbert B, Tollier M-T, Monties B, Fritig B, Legrand M (1995) Altered lignin composition in transgenic tobacco expressing O-methyl-transferase sequences in sense and antisense orientation. . Plant J. **8:** 465-477

Barakat A, Bagniewska-Zadworna A, Choi A, Plakkat U, DiLoreto DS, Yellanki P, Carlson JE (2009) The cinnamyl alcohol dehydrogenase gene family in *Populus*: phylogeny, organization, and expression. BMC Plant Biol **9:** 26

Bhuiyan NH, Selvaraj G, Wei Y, King J (2008) Gene expression profiling and silencing reveal that monolignol biosynthesis plays a critical role in penetration defence in wheat against powdery mildew invasion. J Exp Bot

Bout S, Vermerris W (2003) A candidate-gene approach to clone the sorghum Brown midrib gene encoding caffeic acid O-methyltransferase. Mol Genet Genomics **269:** 205-214

Bugos RC, Chiang VL, Campbell WH (1991) cDNA cloning, sequence analysis and seasonal expression of lignin-bispecific caffeic acid/5-hydroxyferulic acid O-methyltransferase of aspen. Plant Mol Biol **17:** 1203-1215

Chang S, Puryear J, Cairney J (1993) A simple and efficient method for isolating RNA from pine trees Plant Molecular Biology Reporter **11:** 113-116.

Chiron H, Drouet A, Claudot AC, Eckerskorn C, Trost M, Heller W, Ernst D, Sandermann H, Jr. (2000) Molecular cloning and functional expression of a stress-induced multifunctional O-methyltransferase with pinosylvin methyltransferase activity from Scots pine (Pinus sylvestris L.). Plant Mol Biol **44:** 733-745

Collendavelloo J, Legrand M, Geoffroy P, Barthelemy J, Fritig B (1981) Purification and properties of the three o-diphenol-O-methyltransferases of tobacco leaves. Phytochemistry **20:** 611-616

Davin LB, Lewis NG (1992) Phenylpropanoid metabolism: Biosynthesis of monolignols, lignans and neolignans, lignins and suberins. . Plenum Press New York:

Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics **5:** 113

Fritig B, Heitz T, Legrand M (1998) Antimicrobial proteins in induced plant defense. Curr Opin Immunol **10:** 16-22

Guillaumie S, Goffner D, Barbier O, Martinant JP, Pichon M, Barriere Y (2008) Expression of cell wall related genes in basal and ear internodes of silking brown-midrib-3, caffeic acid O-methyltransferase (COMT) down-regulated, and normal maize plants. BMC Plant Biol **8:** 71

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Systematic Biology **52:** 696-704

Guo D, Chen F, Inoue K, Blount JW, Dixon RA (2001) Downregulation of caffeic acid 3-O-methyltransferase and caffeoyl CoA 3-O-methyltransferase in transgenic alfalfa. impacts on lignin structure and implications for the biosynthesis of G and S lignin. Plant Cell **13:** 73-88

Halpin C, Knight ME, Foxon GA, Campbell MM, Boudet AM, Boon JJ, Chabbert B, Tollier M, Schuch W (1994) Manipulation of lignin quality by down regulation of cinnamyl alcohol dehydrogenase. Plant Journal **6:** 339–350

Han Y, Gasic K, Korban SS (2007) Multiple-copy cluster-type organization and evolution of genes encoding O-methyltransferases in the apple. Genetics **176:** 2625-2635

Hatfield RD, Grabber J, Ralph J, Brei K (1999) Using the acetyl bromide assay to determine lignin concentrations in herbaceous plants: some cautionary notes. J Agric Food Chem **47:** 628-632

Hayakawa T, Nanto K, Kawai S, Katayama Y, Morohoshi N (1996) Molecular cloning and tissue-specific expression of two genes that encode caffeic acid O-methyltransferases from *Populus* kitakamiensis. Plant Sci **113:** 157–165

Higuchi K, Kanazawa K, Nishizawa N, Chino M, Mori S (1994) Purification and characterization of nicotianamine synthase from Fe-deficient barley roots Plant Soil **165** 173–179

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

Hoffmann L, Maury S, Pinc¸on G, Geoffroy P, Fritig B, Legrand M (2000) Impacts of the repression of phenylpropanoid synthesis on development and disease resistance of transgenic tobacco. 6th International Congress of Plant Molecular Biology

Inoue K, Parvathi K, Dixon RA (2000) Substrate preferences of caffeic acid/5-hydroxyferulic acid 3/5-O-methyltransferases in developing stems of alfalfa (Medicago sativa L.). Arch Biochem Biophys **375:** 175-182

Inoue K, Sewalt VJ, Murray GB, Ni W, Sturzer C, Dixon RA (1998) Developmental expression and substrate specificities of alfalfa caffeic acid 3-O-methyltransferase and caffeoyl coenzyme A 3-O-methyltransferase in relation to lignification. Plant Physiol **117:** 761-770

Jaeck E, Dumas B, Geoffroy P, Favet N, Inze D, Van Montagu M, Fritig B, Legrand M (1992) Regulation of enzymes involved in lignin biosynthesis: induction of O-methyltransferase mRNAs during the hypersensitive reaction of tobacco to tobacco mosaic virus. Mol Plant Microbe Interact **5:** 294-300

Jorgenson L (1931) Brown midrib in maize and its lignage relations J Am Soc Agron **23** 549-557

Joshi C, Chiang V (1998) Conserved sequence motifs in plant S-adenosyl-L-methionine-dependent methyltransferase. Plant Mol Biol **37:** 663-674

Jouanin L, Goujon T, de Nadai V, Martin MT, Mila I, Vallet C, Pollet B, Yoshinaga A, Chabbert B, Petit-Conil M, Lapierre C (2000) Lignification in transgenic poplars with extremely reduced caffeic acid O-methyltransferase activity. Plant Physiol **123:** 1363-1374

Kersey R, Inoue K, Schubert KR, Dixon RA (1999) Immunolocalization of two ligninO-methyltransferases in stems of alfalfa (Medicago sativa L.). Protoplasma **209:** 46-57

Kishimoto K, Matsui K, Ozawa R, Takabayashi J (2005) Volatile C6-aldehydes and Allo-ocimene activate defense genes and induce resistance against Botrytis cinerea in *Arabidopsis* thaliana. Plant Cell Physiol **46:** 1093-1102

Kota P, Guo D, Zubieta C, Noel J, Dixon RA (2004) O-Methylation of benzaldehyde derivatives by "lignin specific" caffeic acid 3-O-methyltransferase. Phytochemistry **65:** 837-846

Lam HM, Coschigano KT, Oliveira IC, Melo-Oliveira R, Coruzzi GM (1996) THE MOLECULAR-GENETICS OF NITROGEN ASSIMILATION INTO AMINO ACIDS IN HIGHER PLANTS. Annu Rev Plant Physiol Plant Mol Biol **47:** 569-593

Lapierre C, Pollet B, Petit-Conil M, Toval G, Romero J, Pilate G, Leple JC, Boerjan W, Ferret VV, De Nadai V, Jouanin L (1999) Structural alterations of lignins in transgenic poplars with depressed cinnamyl alcohol dehydrogenase or caffeic acid O-methyltransferase activity have an opposite impact on the efficiency of industrial kraft pulping. Plant Physiol **119:** 153-164

Legrand M, Fritig B, Hirth L (1978) o-Diphenol Omethyltransferases of healthy and tobacco-mosaic-virus-infected hypersensitive tobacco. Planta **144:** 101–108

Lescot M, Dehais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouze P, Rombauts S (2002) PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Res **30:** 325-327

Li HM, Rotter D, Hartman TG, Pak FE, Havkin-Frenkel D, Belanger FC (2006) Evolution of novel O-methyltransferases from the Vanilla planifolia caffeic acid O-methyltransferase. Plant Mol Biol **61:** 537-552

Li L, Osakabe K, Joshi C, Chiang V (1999) Secondary xylem-specific expression of caffeoyl-coenzyme A 3-Omethyltransferase plays an important role in the methylation pathway associated with lignin biosynthesis in loblolly pine. Plant Mol Biol **40** 555–565

Li L, Popko J, Zhang X, Osakabe K, Tsai C, Joshi C, Chiang V (1997) A novel multifunctional O-methyltransferase implicated in a dual methylation pathway associated with lignin biosynthesis in loblolly pine. Proc Natl Acad Sci USA **94** 5461–5466

Li L, Popko JL, Umezawa T, Chiang VL (2000) 5-hydroxyconiferyl aldehyde modulates enzymatic methylation for syringyl monolignol formation, a new view of monolignol biosynthesis in angiosperms. J Biol Chem **275:** 6537-6545

Ma QH, Xu Y, Lin ZB, He P (2002) Cloning of cDNA encoding COMT from wheat which is differentially expressed in lodging-sensitive and -resistant cultivars. J Exp Bot **53:** 2281-2282

Maury S, Geoffroy P, Legrand M (1999) Tobacco O-methyltransferases involved in phenylpropanoid metabolism. The different caffeoyl-coenzyme A/5-hydroxyferuloyl-coenzyme A 3/5-O-methyltransferase and caffeic acid/5-hydroxyferulic acid 3/5-O-methyltransferase classes have distinct substrate specificities and expression patterns. Plant Physiol **121:** 215-224

Ming R, Hou S, Feng Y, Yu Q, Dionne-Laporte A, Saw JH, Senin P, Wang W, Ly BV, Lewis KL, Salzberg SL, Feng L, Jones MR, Skelton RL, Murray JE, Chen C, Qian W, Shen J, Du P, Eustice M, Tong E, Tang H, Lyons E, Paull RE, Michael TP, Wall

K, Rice DW, Albert H, Wang ML, Zhu YJ, Schatz M, Nagarajan N, Acob RA, Guan P, Blas A, Wai CM, Ackerman CM, Ren Y, Liu C, Wang J, Na JK, Shakirov EV, Haas B, Thimmapuram J, Nelson D, Wang X, Bowers JE, Gschwend AR, Delcher AL, Singh R, Suzuki JY, Tripathi S, Neupane K, Wei H, Irikura B, Paidi M, Jiang N, Zhang W, Presting G, Windsor A, Navajas-Perez R, Torres MJ, Feltus FA, Porter B, Li Y, Burroughs AM, Luo MC, Liu L, Christopher DA, Mount SM, Moore PH, Sugimura T, Jiang J, Schuler MA, Friedman V, Mitchell-Olds T, Shippen DE, dePamphilis CW, Palmer JD, Freeling M, Paterson AH, Gonsalves D, Wang L, Alam M (2008) The draft genome of the transgenic tropical fruit tree papaya (Carica papaya Linnaeus). Nature **452:** 991-996

Ni W, Sewalt V, Korth K, Blount J, Ballance G, Dixon R (1996) Stress responses in alfalfa (Medicago sativa L.). XXI. Activation of caffeic acid 3-O-methyltransferase and caffeoyl CoA 3-O-methyltransferase genes does not contribute to changes in metabolite accumulation in elicitor-treated cell suspension cultures. Plant Physiol **112:** 717–726

Nicholson R, Hammerschmidt R (1992) Phenolic compounds and their role in disease resistance. Annual Review of Phytopathology **30:** 369–389

Osakabe K, Tsao CC, Li L, Popko JL, Umezawa T, Carraway DT, Smeltzer RH, Joshi CP, Chiang VL (1999) Coniferyl aldehyde 5-hydroxylation and methylation direct syringyl lignin biosynthesis in angiosperms. Proc Natl Acad Sci U S A **96:** 8955-8960

Palmer NA, Sattler SE, Saathoff AJ, Funnell D, Pedersen JF, Sarath G (2008) Genetic background impacts soluble and cell wall-bound aromatics in brown midrib mutants of sorghum. Planta **229:** 115-127

Park JH, Halitschke R, Kim HB, Baldwin IT, Feldmann KA, Feyereisen R (2002) A knock-out mutation in allene oxide synthase results in male sterility and defective wound signal transduction in *Arabidopsis* due to a block in jasmonic acid biosynthesis. Plant J **31:** 1-12

Parvathi K, Chen F, Guo D, Blount JW, Dixon RA (2001) Substrate preferences of O-methyltransferases in alfalfa suggest new pathways for 3-O-methylation of monolignols. Plant J **25:** 193-202

Pellegrini L, Geoffroy P, Fritig B, Legrand M (1993) Molecular cloning and expression of a new class of ortho-diphenol-O-methyltransferases induced in tobacco (Nicotiana tabacum L.) leaves by infection or elicitor treatment. Plant Physiol **103:** 509-517

Pilate G, Guiney E, Holt K, Petit-Conil M, Lapierre C, Leple JC, Pollet B, Mila I, Webster EA, Marstorp HG, Hopkins DW, Jouanin L, Boerjan W, Schuch W, Cornu D, Halpin C (2002) Field and pulping performances of transgenic trees with altered lignification. Nat Biotechnol **20:** 607-612

Pillonel C, Hunziker P, Binder A (1992) Multiple forms of the constitutive wheat cinnamyl alcohol dehydrogenase. J Exp Bot **43** 299-305

Pincon G, Chabannes M, Lapierre C, Pollet B, Ruel K, Joseleau JP, Boudet AM, Legrand M (2001) Simultaneous down-regulation of caffeic/5-hydroxy ferulic acid-O-methyltransferase I and cinnamoyl-coenzyme A reductase in the progeny from a cross between tobacco lines homozygous for each transgene. Consequences for plant development and lignin synthesis. Plant Physiol **126:** 145-155

Piquemal J, Chamayou S, Nadaud I, Beckert M, Barriere Y, Mila I, Lapierre C, Rigau J, Puigdomenech P, Jauneau A, Digonnet C, Boudet AM, Goffner D, Pichon M (2002)

Down-regulation of caffeic acid o-methyltransferase in maize revisited using a transgenic approach. Plant Physiol **130:** 1675-1685

Raes J, Rohde A, Christensen JH, Van de Peer Y, Boerjan W (2003) Genome-wide characterization of the lignification toolbox in *Arabidopsis*. Plant Physiol **133:** 1051-1071

Rodriguez FJ, Oliver JL, Marín A, Medina JR (1990) The general stochastic model of nucleotide substitution. Journal of Theoretical Biology **142:** 485-501

Stintzi A, Heitz T, Prasad V, Wiedemann-Merdinoglu S, Kauffmann S, Geoffroy P, Legrand M, Fritig B (1993) Plant 'pathogenesis-related' proteins and their role in defense against pathogens. Biochimie **75:** 687-706

Stintzi A, Weber H, Reymond P, Browse J, Farmer EE (2001) Plant defense in the absence of jasmonic acid: the role of cyclopentenones. Proc Natl Acad Sci U S A **98:** 12837-12842

Tiimonen H, Aronen T, Laakso T, Saranpaa P, Chiang V, Ylioja T, Roininen H, Haggman H (2005) Does lignin modification affect feeding preference or growth performance of insect herbivores in transgenic silver birch (Betula pendula Roth)? Planta **222:** 699-708

Toquin V, Grausem B, Geoffroy P, Legrand M (2003) Structure of the tobacco caffeic acid O-methyltransferase (COMT) II gene: identification of promoter sequences involved in gene inducibility by various stimuli. Plant Mol Biol **52:** 495-509

Tsai CJ, Podila GK, Chiang VL (1995) Nucleotide sequence of a *Populus* tremuloides gene encoding bispecific caffeic acid/5-hydroxyferulic acid O-methyltransferase. Plant Physiol **107:** 1459

Tsai CJ, Popko JL, Mielke MR, Hu WJ, Podila GK, Chiang VL (1998) Suppression of O-methyltransferase gene by homologous sense transgene in quaking aspen causes red-brown wood phenotypes. Plant Physiol **117:** 101-112

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, Van de Peer Y, Rokhsar D (2006) The genome of black cottonwood, *Populus* trichocarpa (Torr. & Gray). Science **313:** 1596-1604

Van Doorsselaere J, Baucher M, Chognot E, Chabbert B, Tollier M-T, Petit-Conil M, Leplé J-C, Pilate G, Cornu D, Monties B, Van Montagu M, Inzé D, Boerjan W, Jouanin L (1995a) A nove1 lignin in poplar trees with a reduced caffeic acid/5-hydroxyferulic acid O-methyltransferase activity. Plant J **8**

Vander Mijnsbrugge K, Meyermans H, Van Montagu M, Bauw G, Boerjan W (2000) Wood formation in poplar: identification, characterization, and seasonal variation of xylem proteins. Planta **210:** 589-598

Velasco R, Zharkikh A, Troggio M, Cartwright DA, Cestaro A, Pruss D, Pindo M, Fitzgerald LM, Vezzulli S, Reid J, Malacarne G, Iliev D, Coppola G, Wardell B, Micheletti D, Macalma T, Facci M, Mitchell JT, Perazzolli M, Eldredge G, Gatto P, Oyzerski R, Moretto M, Gutin N, Stefanini M, Chen Y, Segala C, Davenport C, Dematte L, Mraz A, Battilana J, Stormo K, Costa F, Tao Q, Si-Ammour A, Harkins T, Lackey A, Perbost C, Taillon B, Stella A, Solovyev V, Fawcett JA, Sterck L, Vandepoele K, Grando SM, Toppo S, Moser C, Lanchbury J, Bogden R, Skolnick M, Sgaramella V, Bhatnagar SK, Fontana P, Gutin A, Van de Peer Y, Salamini F, Viola R (2007) A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. PLoS ONE **2:** e1326

Vignols F, Rigau J, Torres MA, Capellades M, Puigdomenech P (1995) The brown midrib3 (bm3) mutation in maize occurs in the gene encoding caffeic acid O-methyltransferase. Plant Cell **7:** 407-416

Vogler RK, Tesso TT, Johnson KD, Ejeta G (2009) The effect of allelic variation on forage quality of brown midrib sorghum mutants with reduced caffeic acid O-methyl transferase activity African Journal of Biochemistry Research **3:** 70-76

Weeks JT, Ye J, Rommens CM (2008) Development of an in planta method for transformation of alfalfa (Medicago sativa). Transgenic Res **17:** 587-597

Whelan S, Goldman N (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Molecular Biology and Evolution **18:** 691-699

Whetten R, MacKay JJ, Sederoff RR (1998) Recent advances in understanding lignin biosynthesis. Annu. Rev. Plant Physiol. Plant Mol. Biol **49:** 585-609

Wingender E, Dietze P, Karas H, Knuppel R (1996) TRANSFAC: a database on transcription factors and their DNA binding sites. Nucleic Acids Res **24:** 238-241

Ye ZH, Kneusel RE, Matern U, Varner JE (1994) An alternative methylation pathway in lignin biosynthesis in Zinnia. Plant Cell **6:** 1427-1439

Ye ZH, Varner JE (1995) Differential expression of two O-methyltransferases in lignin biosynthesis in Zinnia elegans. Plant Physiol **108:** 459-467

Zhong R, Iii WH, Negrel J, Ye ZH (1998) Dual methylation pathways in lignin biosynthesis. Plant Cell **10:** 2033-2046

Zubieta C, Kota P, Ferrer JL, Dixon RA, Noel JP (2002) Structural basis for the modulation of lignin monomer methylation by caffeic acid/5-hydroxyferulic acid 3/5-O-methyltransferase. Plant Cell **14:** 1265-1277

Appendix A. List of plant genes used in COMT gene phylogenetic analyses. The gene names used in this study, the accession number, species, the database source, and names of previously published genes are indicated.

Gene	Accession N	Species	Data source	Previous names	Previous accession codes
AthCOMT1	*AT1G21100*	*Arabidopsis thaliana*	*GenBank*		
AthCOMT10	*AT1G77530*	*Arabidopsis thaliana*	*GenBank*		
AthCOMT11	*AT5G37170*	*Arabidopsis thaliana*	*GenBank*		
AthCOMT12	*AT5G53810*	*Arabidopsis thaliana*	*GenBank*		
AthCOMT13	*AT5G54160*	*Arabidopsis thaliana*	*GenBank*	*COMT*	*AAB96879, U70424*
AthCOMT2	*AT1G21110*	*Arabidopsis thaliana*	*GenBank*		
AthCOMT3	*AT1G21120*	*Arabidopsis thaliana*	*GenBank*		
AthCOMT4	*AT1G21130*	*Arabidopsis thaliana*	*GenBank*		
AthCOMT5	*AT1G33030*	*Arabidopsis thaliana*	*GenBank*		
AthCOMT6	*AT1G51990*	*Arabidopsis thaliana*	*GenBank*		
AthCOMT7	*AT1G63140*	*Arabidopsis thaliana*	*GenBank*	*COMT*	*AK221717*
AthCOMT8	*AT1G76790*	*Arabidopsis thaliana*	*GenBank*		
AthCOMT9	*AT1G77520*	*Arabidopsis thaliana*	*GenBank*		
CpaCOMT1	*supercontig_3.161*	*Carica papaya*	*Ming et al., 2008*		
LpeCOMT1	*AF010291*	*Lolium perenne*	*Lam et al., 2007*	*LoliHCA*	
MtrCOMT1	*AC119415_2.4*	*Medicago truncatula*	*MIPS*	*MsCOMT*	*P28002, M63853, AAB46623.1*
MtrCOMT2	*AC144341_5.5*	*Medicago truncatula*	*MIPS*		
MtrCOMT3	*AC144341_6.5*	*Medicago truncatula*	*MIPS*		
MtrCOMT4	*AC145512_17.3*	*Medicago truncatula*	*MIPS*		
MtrCOMT5	*AC150443_24.5*	*Medicago truncatula*	*MIPS*		
MtrCOMT6	*AC150443_26.5*	*Medicago truncatula*	*MIPS*		
MtrCOMT7	*AC167710_15.3*	*Medicago truncatula*	*MIPS*		
OsaCOMT1	*Os04g01470*	*Oryza sativa*	*TIGR*		
OsaCOMT2	*Os04g09604*	*Oryza sativa*	*TIGR*		
OsaCOMT3	*Os04g09654*	*Oryza sativa*	*TIGR*		
OsaCOMT4	*Os08g06100*	*Oryza sativa*	*TIGR*		
PglCOMT1	*TA15300_3330*	*Picea glauca*	*TIGR*		
PpaCOMT1	*e_gw1.116.68.1*	*Physcomitrella patens*	*JGI*		
PpaCOMT2	*estExt_gwp_gw1.C_560052*	*Physcomitrella patens*	*JGI*		
PsiCOMT1	*TA20657_3332*	*Picea sitchensis*	*TIGR*		
PsiCOMT2	*TA20659_3332*	*Picea sitchensis*	*TIGR*		
PsiCOMT3	*TA20659_3332*	*Picea sitchensis*	*TIGR*		
PsyOMT	*directly added*	*Pinus sylvestris*	*TIGR*		
PtaCOMT1	*TA5057_3352*	*Pinus taeda*	*TIGR*		
PtaCOMT2	*TA5058_3352*	*Pinus taeda*	*TIGR*		
PtaCOMT3	*TA5054_3352*	*Pinus taeda*	*TIGR*		
PtrCOMT1	*gw1.I.5581.1*	*Populus trichocarpa*	*JGI*		
PtrCOMT2	*gw1.XIV.1942.1*	*Populus trichocarpa*	*JGI*		
PtrCOMT3	*gw1.12518.3.1*	*Populus trichocarpa*	*JGI*		
PtrCOMT4	*estExt_Genewise1_v1.C_LG_XIV1942*	*Populus trichocarpa*	*JGI*		
PtrCOMT5	*fgenesh4_pm.C_LG_II000840*	*Populus trichocarpa*	*JGI*		
PtrCOMT6	*fgenesh4_pm.C_LG_XI000417*	*Populus trichocarpa*	*JGI*		
PtrCOMT7	*estExt_fgenesh4_pg.C_LG_XV0035*	*Populus trichocarpa*	*JGI*	*OMT*	*U50522*
PtrCOMT8	*estExt_fgenesh4_pm.C_LG_XII0129*	*Populus trichocarpa*	*JGI*	*CAOMT*	*M73431, U13171, X62096*
SbiCOMT1	*Sb01g042900.1*	*Sorghum bicolor*	*JGI*		
SbiCOMT2	*Sb07g003860.1*	*Sorghum bicolor*	*JGI*	*COMT*	*AF387790*

SbiCOMT3	*Sb07g004680.1*	*Sorghum bicolor*	*JGI*		
SbiCOMT4	*Sb07g004690.1*	*Sorghum bicolor*	*JGI*		
SbiCOMT5	*Sb07g004710.1*	*Sorghum bicolor*	*JGI*		
SmoCOMT1	*e_gw1.130.28.1*	*Selaginella moellendorffii*	*JGI*		
SmoCOMT2	*estExt_fgenesh2_pg.C_30578*	*Selaginella moellendorffii*	*JGI*		
SmoCOMT3	*e_gw1.1.450.1*	*Selaginella moellendorffii*	*JGI*		
SmoCOMT4	*e_gw1.2.377.1*	*Selaginella moellendorffii*	*JGI*		
SmoCOMT5	*e_gw1.7.273.1*	*Selaginella moellendorffii*	*JGI*		
SmoCOMT6	*gw1.73.10.1*	*Selaginella moellendorffii*	*JGI*		
VviCOMT1	*GSVIVP00000479001*	*Vitis vinifera*	*Velasco et al., 2007*		
VviCOMT2	*GSVIVP00026177001*	*Vitis vinifera*	*Velasco et al., 2007*		
VviCOMT3	*GSVIVP00026179001*	*Vitis vinifera*	*Velasco et al., 2007*		
VviCOMT4	*GSVIVP00026182001*	*Vitis vinifera*	*Velasco et al., 2007*		
VviCOMT5	*GSVIVP00031799001*	*Vitis vinifera*	*Velasco et al., 2007*		
VviCOMT6	*GSVIVP00037163001*	*Vitis vinifera*	*Velasco et al., 2007*		
VviCOMT7	*GSVIVP00037164001*	*Vitis vinifera*	*Velasco et al., 2007*		
ZelCOMT1	*U19911*	*Zinnia elegans*	*Lam et al., 2007*	*ZinnHCA*	
ZmaCOMT1	*M73235*	*Zea mays*	*Lam et al., 2007*	*ZeaHCA*	

Alex Choi

814-482-0362 • choi0625@gmail.com • 532 E. College Ave APT10, State College, PA, 16801

EDUCATION

THE PENNYLVANIA STATE UNIVERSITY — University Park, PA, USA
Biology B.S. - Vertebrate Physiology Option — Eberly College of Science
Psychology B.A. — College of Liberal Arts
Estimate Graduation Date: May 2010

HONORS

Schreyer's Honors College — **Fall 2006 - Present**
Dean's List — **Fall 2006 – Present**

WORK EXPERIENCE

FOREST RESOURCES LAB – PSU (15hours/week) — **Spring 2008 - Present**

- Participate in experiments investigating role of key enzymes in lignin biosynthesis
- Utilize **databases** (TAIR, TIGR, GenBank, etc.) to gather sequences, **bioinformatics software** (BioEdit, Phyml2.4.4, etc.) to conduct computational analyses, **lab** (centrifuges, scales, liquid nitrogen, realtime RT-PCR, autoclave, etc.) to conduct expression analyses
- Participated in poster constructions, of which one during Summer 2008 was funded with a $750 grant from the Pre-Eminence in Honors Education Fund.
- Received $375 for summer internship during Summer 2009.

TEACHING ASSISTANT – PSU Department of Biology — **Fall 2008 – Spring 2009**

- Taught and graded lab portion of Biology core courses (BIO110, BIO 240W) for one or more section(s) (20 to 25 students each)
- Prepared class material: PowerPoint slides, Quizzes, Overheads, Lab Manuals
- Graded assignments: Quizzes, In-class assignments, Homework, Lab Reports
- Attended TAs meeting to discuss curriculum and modification

PSYCHOLOGY LAB – PSU (10hours/week) — **Spring 2007 – Fall 2007**

- Submitted data from experiments investigating child fear development
- Gathered data of potential subjects in past birth records on microfilms
- Participated in experiments as: recorder, "stranger," and scene organizer

PUBLICATIONS

1. Barakat A, Bagniewska-Zadworna A, **Choi A**, Plakkat U, DiLoreto DS, Yellanki P, Carlson JE. 2009. The cinnamyl alcohol dehydrogenase gene family in *Populus*: phylogeny, organization, and expression. *BMC Plant Biol*. 9:26.

- **POSTERS**
  - The Pan American Congress on Plants & BioEnergy. Mérida, Mexico, June, 22-25, 2008.
  - Plant & Animal Genome XIV Conference. San Diego, CA, January, 10-14, 2009.
  - Schatz Tree Genetics Colloquium. Monto Alto, PA, June, 8, 2009.
  - Undergraduate Exhibition 2009, Pennyslvania State University, University Park, PA, April, 2009.

2. Barakat A, **Choi A**, Yassin NB, Sun Z, Carlson JE. (2009). Genome wide analyses of the Caffeic Acid O-Methyltransferase (COMT) gene family in *Populus* (Manuscript and Honors Thesis in preparation)

- **POSTERS**
  - Undergraduate Exhibition 2009, Pennyslvania State University, University Park, PA, April, 2009.

3. Barakat A, **Choi A**, Carlson JE. (2009). Genome wide organization and evolution of Caffeoyl CoA Reductase (CCR) genes in *Populus*, *Arabidopsis*, and rice. (Manuscript in preparation)

EXTRA-CURRICULAR ACTIVITIES

SCHOOL SINFONIETTA **Spring 2007 – Spring 2010**

- First violin chair after audition
- Recital at school auditorium at end of every semester

HOSPITAL VOLUNTEER at ER (300hrs) **Summer 2007, 2008, 2009**

- Shadowed doctor in surgeries
- Helped nurses care patients
- Stayed overnight shifts biweekly
- Total 300 hours