Comparative genomics and evolutionary analyses of the O-methyltransferase gene family in *Populus*

Abdelali Barakat, Alex Choi, Norzawani Buang M. Yassin, Joseph S. Park, Zichao Sun, John E. Carlson

**A R T I C L E   I N F O**

**Abstract**

S-adenosyl-l-methionine (SAM) dependent O-methyltransferases (OMTs) proteins are involved in the methylation of various secondary metabolites. The OMT genes have been studied in various plants, but these studies focused either on a single or a small set of genes. Moreover, no comprehensive study was published yet on the OMT gene family in a tree species. To investigate the evolutionary history of this gene family and the functional diversification of its members, phylogenetic and several comparative genomics analyses were performed. Phylogeny across land plant lineages showed that OMT genes were distributed in two main classes deeply rooted in the phylogeny of land plants, suggesting that they have evolved by a gene duplication that had happen in the ancestor of land plants. COMT and COMT-like genes were clustering with few flavonoid and multifunctional OMT genes in class II. Class I included flavonoid, simple phenol, and multifunctional OMT genes. All 26 *Populus* OMT genes were located in segmental duplication blocks and two third of them were tandem duplicated, indicating the role of duplication processes in the expansion of this gene family. Expression profiling of OMT genes in *Populus* showed that only *PoptrOMT25* was differentially expressed in xylem. The other genes were differentially expressed in leaves, bark, or both. Some OMT genes showed differential expression patterns under various biotic and abiotic stresses. The divergence of protein sequences, the phylogenetic distribution, and the expression of COMT and COMT-like genes suggest that they have evolved different functions or tissue specificities following duplications.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

S-adenosyl-l-methionine (SAM) dependent O-methyltransferases (OMTs) are proteins that catalyze the methylation of small molecules such as flavonoids, alkaloids, phytoalexins, lignin precursors, etc. (Lam et al., 2007). Plant methyltransferases were classified into three main classes based on their substrates: O-methyltransferases (OMTs), N-methyltransferases (NMTs), and C-methyltransferases (CMTs) (Roje, 2006). OMTs were further classified into five classes based on substrate specificity (Roje, 2006). Class A includes caffeoyl coenzyme A 3-O-methyltransferase (CCoAOMT) and caffeic acid 3-O-methyltransferase (COMT) proteins, which act on the hydroxyl groups of phenylpropanoids. Genes from class B, C, and D are involved in the methylation of hydroxyl groups within flavonoids, alkaloids, and myo-inositol O-methyltransferase, respectively (Rammesmayer et al., 1995; Roje, 2006). A fifth class of OMT proteins methylates the carboxyl group of various acids (Roje, 2006).

An early study (Joshi and Chiang, 1998) showed that OMT genes were distributed into two distinct groups (PL-OMT I and PL-OMT II) based on sequence similarity and protein signature motifs. Genes from the PL-OMT I group represented mainly by CCoAOMT sequences were involved in lignin biosynthesis and use only a pair of substrates (caffeoyl CoA and 5-hydroxyferuloyl CoA). PL-OMT II group was represented by COMT and other OMT proteins acting on a variety of substrates, such as caffeic acid (CA), 5-hydroxyferulic acid (5HFA), caffeoyl CoA ester, 5-hydroxyferuloyl ester, myo-inositol, chalcones and scoulerine (Joshi and Chiang, 1998). Conserved protein motifs involved in the binding of S-adenosyl methionine, which is a common methyl donor of a large variety of methyltransferases in plants, specific to SAM-OMT from PL-OMT I and PL-OMT II were reported (Joshi and Chiang, 1998).

Abbreviations: OMT, O-methyltransferase; 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; AOMT, hydroxycinnamoyl acids/hydroxycinnamoyl CoA esters O-methyltransferase; in, intron; ex, exon.

* Corresponding author at: Department of Genetics and Biochemistry, Clemson University, 100 Jordan Hall, Clemson, SC 29634, USA. Tel.: +1 864 656 3060; fax: +1 864 656 6879.

**E-mail address:** abarakat@clemson.edu (A. Barakat).

1 Current address: Department of Genetics and Biochemistry, Clemson University, 100 Jordan Hall, Clemson, SC 29634, USA.

2 Current address: University of Michigan School of Public Health, 1415 Washington Heights, 1700 SPH I, Ann Arbor, MI 48109-2029, USA.
CCoAOMT and COMT are crucial enzymes in the lignin biosynthesis pathway. While both enzymes are involved in lignin biosynthesis, CCoAOMT catalyzes an earlier step in the lignin biosynthesis pathway by transforming caffeoyl CoA to feruloyl CoA (Davin and Lewis, 1992; Ye et al., 1994). COMT proteins act at the end of the pathway to produce sinapyl alcohol which is the main component of S-type lignin. COMT proteins methylate mainly caffeic acid and 5-HFA (Hoffmann et al., 2004), but they could also methylate one or more of various other substrates such as caffeoyl CoA, caffeoyl aldehyde, caffeoyl alcohol, 5-hydroxyferuloyl CoA, 5-hydroxyconiferyl alcohol, and 5-hydroxyconiferyl alcohol (Roje, 2006). Functional analyses showed that mutations of real COMT genes can affect the monolignol type and the content of lignin as well as plant phenotypes (Tsai et al., 1995; Lapière et al., 1999; Jouanin et al., 2000). Down-regulation of COMT genes resulted in 17–30% decrease in lignin composition in the hybrid *Populus* ( *Populus tremula* × *Populus alba*) (Jouanin et al., 2000), Medicago (Guo et al., 2001), maize (Piquemal et al., 2002), and Sorghum (Vogler et al., 2009). Plants with suppressed or disturbed COMT gene expression (Atanassova et al., 1995; Jouanin et al., 2000; Guo et al., 2001; Pincon et al., 2001; Piquemal et al., 2002; Bout and Vermerris, 2003; Weeks et al., 2008) 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 the COMT natural mutant bm3 and in transgenic plants expressing a COMT antisense construct (AS225) (Guillaumie et al., 2008) showed that a mutation of a COMT gene could result in a disturbance of the whole cell wall assembly. COMT and COMT-like genes were reported being involved in plant defense against various biotic and abiotic stresses (Lee et al., 1997; Toquin et al., 2003). 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. *Triticum* and *B. graminis* f. sp. *Hordei*, respectively) (Bhuiyan et al., 2008). In *Arabidopsis*, treatment of plants with green leafy volatiles or isoprenoids such as (E)-2-gezenal and (Z)-3-hexenal treatment, which induce several resistance genes including COMT, resulted in a slower rate of disease development when inoculated with *Botrytis cinerea* (Kishimoto et al., 2005). In tobacco, down-regulation of COMT and CCoAOMT genes resulted in plants with large necrotic lesions following tobacco mosaic virus (TMV) infection (Maury et al., 1999; Hoffmann et al., 2000). An expression study using a reporter gene (GUS) that was under the control of OMT promoters confirmed the induction of class II tobacco COMT gene (COMTII) in response to biotic and abiotic stresses (Toquin et al., 2003).

A phylogenetic analysis of plant OMT sequences (Ibrahim et al., 1998) showed that they clustered within a monophyletic group derived from non-plant genes. This study showed the presence of two plant sequence groups clustering according to a functional trait that reflects their substrate specificity. Similar results were published by Joshi and Chiang (1998) who reported a dendrogram showing the distribution of SAM-OMT genes in two groups. Group I OMTs included proteins that use cateyl CoA derivatives as substrate and group II that include the rest of plant OMTs. A phylogenetic analysis using a small set of COMT and COMT-like sequences (Raes et al., 2003) showed that COMT genes cluster together in one group. COMT-like genes were distributed in a separate group including a multifunctional OMT (hydroxycinnamic acids/hydroxycinnamoyl CoA ester O-methyltransferase) (AEOMT) (Li et al., 1997; Li et al., 1999). A recent phylogenetic study using biochemically characterized OMTs from various species (Lam et al., 2007) reported that OMT sequences were distributed in two main lineages. One included CCoAOMT and carboxylic OMT genes, while the other included sequences encoding simple phenols, flavonoids, alkaloids, and COMT proteins (Lam et al., 2007). While these studies and several others not cited in this manuscript provided some insight into the phylogenetic relationship between OMT genes, they all used a partial dataset or sequences not covering all land plant lineages and could not reflect the evolutionary history of this gene family. Moreover, no comprehensive study on the OMT gene family was yet published in a tree species.

In this study, OMT sequences from a variety of plants covering all land plant lineages including *Physcomitrella* and *Selaginella* were retrieved and annotated. The large set of OMT genes was used to analyze the phylogeny of this gene family. Several other comparative genomics analyses including protein motif characterization, gene structure, genome organization, and expression of OMT genes in *Populus* were studied. The evolution of the function of COMT genes is discussed.

2. Materials and methods

2.1. Plant materials

Leaves, cortex, and xylem were collected from young hybrid *Populus* OGY (*P. deltoides* × *P. nigra*), trees grown in a culture chamber at 25 and 18 °C in the day and night, respectively. The plants were grown at 16 h/8 h day/night regime and at 60% humidity. An herbivory stress treatment using *Lymantria dispar* (*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.

2.2. 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 were assessed using an Agilent 2100 Bioanalyzer (Agilent Technologies). Complementary DNA (cDNA) synthesis was performed as described previously (Barakat et al., 2009).

2.3. OMT 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 (Supplementary Table 1) 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 (Supplementary Table 1). 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). ANOVA statistical analyses were performed using Statistica 6.0 software (StatSoft Poland Inc., Tulsa, OH, USA). Expression data of OMT genes under biotic (rust infection) and abiotic (mechanical damage, drought, atmospheric CO₂, UVB) stresses were collected from PopGenIE database (Sjodin et al., 2006; Rinaldi et al., 2007; Sjodin et al., 2009).

2.4. OMT sequences from model species

OMT sequences used in phylogenetic analyses include sequences from plants with fully sequenced genomes as well as other taxa representing key positions on the angiosperm phylogenetic tree. OMT sequences from *Arabidopsis* and *Oryza* (Supplementary Table 2) were retrieved from Joint Genome Institutes (JGI) and The Arabidopsis
Genome Resources (TAIR). OMT sequences from various species such as Arabidopsis (Ming et al., 2008), Vitis vinifera (Velasco et al., 2007), Medicago truncatula (http://www.medicago.org/genome), Sorghum bicolor (JGI), Physcomitrella patens (JGI) and Selaginella moellendorffii (JGI) were identified by querying their genomes. OMT sequences from various non-model species including gymnosperms were retrieved from TIGR Plant Genomics databases (http://www.jcvi.org/) and Genbank (http://www.ncbi.nlm.nih.gov/). Sequences were carefully inspected and corrected for annotation errors such as misplaced stop codon, frame shift, sequencing errors, presence of specific motifs, etc. before use. Only sequences that present at least one of the three previously described SAM-OMT motifs, (V/I/L)(V/I)(D/K)/(V/I)/GGXX(G/A), (V/I/F)/(A/P/E)/X(A/P/G)/DAAXXXK(W/Y/F), and (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, involved in the binding of SAM-OMT proteins to their substrates (Joshi and Chiang, 1998) were considered in this study.

2.5. Intron–exon structure, sequence alignment, and phylogenetic analyses of OMT genes

The intron (in)–exon (ex) structure of OMT genes was retrieved from the Joint Genome Institute website (JGI). The in–ex structure of genes, for which cDNA sequences were available, was checked by aligning genomic and cDNA sequences.

OMT 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 aligned to match the corresponding aa sequences. The WAG model (Whelan and Goldman, 2001), assuming among site rate heterogeneity (WAG + G), was used with the aa sequences. Phylogenetic analyses were performed using Maximum Likelihood (ML) implemented in PHYML v. 2.4.4 (Guindon and Gascuel, 2003). ML method was used because it is well suited for the large dataset including lineages with largely varying rates of evolution used in this study. The robustness of the phylogenetic inference was estimated by 100 bootstrap replicates.

3. Results

3.1. Characterization of SAM-OMT motifs and annotation of OMT sequences

Sequence similarity searches of several plant model genomes enabled the identification of 173 OMT protein sequences (Supplementary Table 2). Sequences that were short or presented multiple stop codons were removed and not considered in this study. In model species for which the genomes were completely sequenced, 157 OMT genes have been identified (Supplementary Table 2): 15 from Arabidopsis, 5 from Carica, 19 from Medicago, 20 from Oryza, 26 from Populus, 11 from Selaginella, 28 from Sorghum, 31 from Vitis, and 2 from Physcomitrella.

SAM-OMT consensus motifs reported previously (Joshi and Chiang, 1998) were used to annotate and sort OMT sequences in functional categories (Fig. 1). This analysis showed that 139 (80.3%), 60 (34.7%), and 105 (60.7%) sequences possess SAM-binding motifs A, B, and C, respectively, with no mismatch (Table 1 and Fig. 1). All class II genes presented the three motifs with 0–3 mismatches. Seven sequences (ZeOMT, ZmaOMT, VviOMT24, PpOMT19, OsaOMT4, SmoOMT8, and PpaOMT2) from class II presented an insertion in motifs B or C. Class I OMT genes all present motifs A and B with 0–3 mismatches, but few possess motif C. Indeed, most of class I sequences have one or two insertions inside motif C (Fig. 1). A previous study (Joshi and Chiang, 1998) reported that motifs A and B were separated by 52 aa with few exceptions. Our analysis showed that these motifs were separated by 40–49 aa with ~76% of the sequences presenting 43 aa separation between motifs. A similar situation was found for the distribution of motifs B and C. Indeed, these two motifs were separated by 19–29 aa with ~94% of the sequences showing a 19 aa separation, instead of 30 aa suggested previously (Joshi and Chiang, 1998).

Other motifs (I, J, K, and L) were reported being specific to group II OMT genes, which include all OMTs except COA OMTs, and were suggested to be involved in substrate specificity (Joshi and Chiang, 1998). Analysis of the distribution of these motifs showed that most of class II sequences possess motif I and about half possess motif J, K, and L (Fig. 1 and Table 2). Indeed, 62 (98.4%), 36 (61.02%), 47 (79.67%), and 37 (62.71%) sequences of class II (59 sequences) showed motifs I, J, K, and L, respectively. Multifunctional OMT genes from class I (Li et al., 1999; Anterola and Lewis, 2002) presented similar distribution of the four motifs. About two third (61.7%) of class I sequences possess motif I with 0–3 mismatches. However, only three sequences (SbiOMT10 and SbiOMT9, and OsaOMT12) from this class, excluding multifunction OMT genes, possess one or more of motifs J, K, and L (Fig. 1).

3.2. Phylogeny of OMT genes

A maximum likelihood (ML) phylogenetic tree using protein sequences (Fig. 1) showed that OMT sequences, except few angiosperm ones, were distributed in two major classes supported by high bootstrap values. Both major OMT classes are represented by monocot, eudicot, gymnosperm, and lycophyte sequences. Class I included biochemically characterized simple phenol, flavonoid, and multifunction OMT genes OMT genes (Li et al., 1997; Lam et al., 2007) and sequences that have high similarity to various OMTs such as simple phenol, alkaloid, and flavonoids. Class II included previously characterized COMT, COMT-like, a previously characterized multifunctional OMT (PsyOMT), and sequences having similarity to flavonoid OMT genes (Raes et al., 2003; Tuskan et al., 2006; Lam et al., 2007). A small group of eudicot and monocot sequences clustered separately from the two major classes at the base of the phylogenetic tree. Search in cDNA databases and expression analysis from this study showed that these genes were all expressed suggesting they are not pseudogenes. When the phylogenetic tree is rooted using class I genes (data not shown), this group of genes cluster with class II sequences. In addition, all these genes possess class II AA motifs A, B, C, I, J, K, and L and all showed similar exon–intron pattern as class II genes. This supports the hypothesis that these genes are members of class II that had evolved fast. Class I included a major clade (group 1) represented by angiosperm sequences and two small groups (Fig. 1, highlighted in pink and green) composed of eudicot, as well as eudicot and gymnosperm sequences. Group 2 (Fig. 1, highlighted in pink) included several eudicot sequences including AthOMT12 and MpiOMT2, which encode a predicted Orcinol and a flavonoid OMTs (Willits et al., 2004), respectively. Group 3 may correspond to multifunctional OMT genes as it includes two previously characterized multifunction OMT genes (PsyOMT1 and PsyOMT2) (Li et al., 1997). Similar situation was found for sequences from class II, where COMT and COMT-like genes, were distributed in two groups. Group 1, composed of monocot, eudicot, and gymnosperm sequences, included several bona fide COMT genes (Raes et al., 2003; Tuskan et al., 2006; Lam et al., 2007). Group 2 included several angiosperm sequences having similarity to flavonoid and alkaloid OMT genes. The distinct distribution of this group was reported previously (Raes et al., 2003). These groups of genes could correspond to derived sequences that have evolved fast after the split of eudicots and monocots. Distribution of OMT genes showed that the increase in gene number in the OMT gene family is associated with extensive duplication of which many are species-specific. Some of these duplication events are common to Medicago, Vitis, and Populus.
3.3. Organization of OMT gene family in Populus

Analysis of the gene distribution in the Populus genome showed that 24 out of 26 OMT genes (92%) were mapped to 11 chromosomes: I, II, IV, VI, IX, XI, XII, XIV, XV, and XIX (Fig. 2). Two remaining genes (PoptrOMT6 and PoptrOMT15) were distributed on scaffolds not anchored yet on the physical map. All OMT mapped genes were located on segmental duplication blocks. Some of these genes (PoptrOMT1, PoptrOMT2, PoptrOMT3, PoptrOMT10, PoptrOMT16, PoptrOMT7, PoptrOMT18, PoptrOMT19, PoptrOMT21, PoptrOMT22, PoptrOMT25, PoptrOMT26) are the result of segmental duplications or polyploidy. Among these genes, three duplicate gene pairs (PoptrOMT21–PoptrOMT25; PoptrOMT18–PoptrOMT3; PoptrOMT16–PoptrOMT10) were still located on conserved positions on homologous duplicated blocks. Other genes (PoptrOMT1, PoptrOMT2, PoptrOMT17, PoptrOMT19, PoptrOMT22, and PoptrOMT26) have lost their counter-parts (second copy). Six pairs of Populus OMT genes (PoptrOMT16–PoptrOMT8; PoptrOMT7–PoptrOMT8; PoptrOMT3–PoptrOMT14; PoptrOMT12–PoptrOMT9; PoptrOMT10–PoptrOMT20; PoptrOMT24–PoptrOMT11–PoptrOMT23–PoptrOMT13–PoptrOMT4–PoptrOMT5) were tandem duplicated. Two pairs (PoptrOMT7–PoptrOMT18; PoptrOMT3–PoptrOMT14) seem to be the result of a segmental duplication followed by a tandem duplication event. A similar situation was found for Arabidopsis where all OMT genes (100%) were either duplicated in tandem or located on segmental duplication blocks (results not shown). In Oryza, only three (15%) out of the 20 identified OMT genes were located on segmental duplication blocks or distributed in tandem (results not shown).

3.4. Intron–exon structure of Populus OMT genes

Gene structure analysis of Populus OMT genes (Fig. 3) showed the existence of five patterns of intron–exon structure: Pattern 1 (2ex/1in), pattern 2 (4ex/3in), pattern 3 (3ex/2in), pattern 4 (2ex/1in), and pattern 5 (3ex/2in). 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 1 and pattern 2 were found in Populus, Arabidopsis, and Oryza. However, patterns 3, 4, and 5 exist only in Populus and seem to be derived from pattern 1 and pattern 2. Pattern 2 was also found in a Physcomitrella sequence (PpaOMT1) (result not shown), indicating that this pattern is ancestral. Homeologous duplicate pairs (PoptrOMT21–PoptrOMT25) showed similar structure (Fig. 3). Tandem duplicated gene pairs (PoptrOMT16–PoptrOMT8; PoptrOMT7–PoptrOMT18; PoptrOMT12–PoptrOMT9; PoptrOMT10–PoptrOMT20; PoptrOMT24–PoptrOMT11–PoptrOMT23–PoptrOMT13–PoptrOMT4) also showed similar intron–exon structures.

3.5. Expression analysis of Populus OMT genes

The expression of 20 OMT Populus genes out of 26 was analyzed using quantitative real-time RT-PCR. These 20 genes, of which eight from class II including COMT or COMT-like genes, were chosen from different classes and include duplicated genes to address various evolutionary questions. The results showed that all the studied OMT genes were expressed in leaves, bark, and xylem (Fig. 4). The expression patterns of the OMT genes from each class were classified into different groups. The significance of expression differences between tissues presented below was supported by ANOVA statistical analysis.

Population class II genes were divided in three groups (Fig. 4A). Group 1 gene (PoptrOMT25) was preferably expressed in xylem. PoptrOMT25 is eight times more highly expressed in the xylem than other tissues. Genes from group 2 (PoptrOMT1, PoptrOMT3, and PoptrOMT14) showed preferential expression in bark tissue. Group 3 genes (PoptrOMT11, PoptrOMT19, and PoptrOMT18) did not show any differential expression in any tissue. Class I that include multifunctional, flavonoid, simple phenol, and catechol OMTs showed four different groups (Fig. 4B). Group 1 genes (PoptrOMT8, PoptrOMT23, and PoptrOMT4) were preferentially expressed in leaves. Group 2 genes (PoptrOMT20, PoptrOMT9, PoptrOMT12, PoptrOMT16, and PoptrOMT17) included genes that showed preferential expression in bark tissue. Group 3 (PoptrOMT13 and PoptrOMT11) was represented by genes with high expression in bark and leaves. Group 4 genes (PoptrOMT24 and PoptrOMT26) did not show any differential expression in any tissue.

To check for functional divergence of OMT genes and get insight into their role in plant defense, we analyzed their expression under various biotic and abiotic stresses (Fig. 5 and Supplementary Table 3). Expression analysis in herbivory stressed plants versus non-stressed plants showed that the two real COMT genes were also induced under biotic stresses (rust infection and herbivory) and abiotic stresses (UVB and ozone) (Supplementary Table 3). These two genes showed different expression profiles following herbivory stress (Fig. 5). While PoptrOMT21 expression was induced in bark and leaf tissues, PoptrOMT25 showed an expression increase in xylem. PoptrOMT11 and PoptrOMT13 from class I, which present high similarity to flavonoid OMTs, showed an expression increase in leaves of herbivory stressed plants compared to non-stressed plants. PoptrOMT23 (class I) showed high expression under drought, mechanical damage, and biosphere CO2 stress. PoptrOMT26 was induced under drought stress.

Duplicated gene pairs generated using segmental duplications (PoptrOMT1–PoptrOMT25, PoptrOMT3–PoptrOMT14) showed similar expression profiles. Indeed, PoptrOMT21 and PoptrOMT25 were both induced under biotic and abiotic stresses studied (Figs. 4 and 5 and Supplementary Table 3). Similarly PoptrOMT3–PoptrOMT14 were both...
induced in bark versus leaves and xylem. Tandem duplicated genes showed different expression patterns. Some tandem duplicate pairs (\textit{PoptrOMT7}–\textit{PoptrOMT18}; \textit{PoptrOMT3}–\textit{PoptrOMT14}; \textit{PoptrOMT12}–\textit{PoptrOMT9}) have conserved similar expression profiles. Others (\textit{PoptrOMT11}–\textit{PoptrOMT23}–\textit{PoptrOMT13}–\textit{PoptrOMT4}) present divergent profiles. For instance, \textit{PoptrOMT23} and \textit{PoptrOMT4} were highly expressed in leaves, while \textit{PoptrOMT11} and \textit{PoptrOMT13} were highly expressed in leaves and bark (Fig. 4 and Supplementary Table 3). \textit{PoptrOMT23} was induced following abiotic stress treatments such as mechanical damage, elevated CO\textsubscript{2} concentration, and drought. \textit{PoptrOMT24} did not show a difference in the expression between the tissues analyzed.

4. Discussion

Sequence similarity searches of several plant model genomes enabled the identification of 173 OMT protein sequences. The annotation of these sequences was yet limited to a bioinformatics prediction based on sequence similarity with previously annotated sequences. In this study, the sequences and the distribution of SAM-OMT binding motifs as well as other motifs previously reported (Joshi and Chiang, 1998) were characterized and updated. All OMT sequences analyzed in this study fit in group II OMTs or lineage B previously reported (Joshi and Chiang, 1998; Lam et al., 2007). Genes from this group encode proteins that catalyze the methylation of flavonoids, flavonols, phenylpropanoids, and phenolics. This study showed that all OMT sequences, except four sequences that showed an insertion in motif B, possess the SAM-OMT binding motifs A and B with one to three mismatches (Table 1). Motif A, which was suggested to play a major role in the binding of SAM-OMT proteins to their substrates (Vidgren et al., 1994; Joshi and Chiang, 1998), was present with zero mismatches in 80% of the sequences. Motif B was present in over 98% of sequences with a number of mismatches less than two. Motif C, which was also suggested to function in substrate binding, was found in almost all class II OMT sequences when one to three mismatches were tolerated. Three sequences (\textit{ZeOMT}, \textit{SmoOMT8}, and \textit{VviOMT23}) showed an insertion in motif C. On the opposite, almost all the OMT genes from class I except \textit{Physcomitrella}, \textit{Selaginella}, and two angiosperm sequences did not show motif C even by tolerating one to three mismatches. They all have insertions within motif C. This indicates that motif C is specific to hydroxycinnamic acid (HCA) OMTs or the so-called COMT and COMT-like genes (class II genes). Motif C was reported in all SAM-OMT genes investigated in a previous study (Joshi and Chiang, 1998). The discrepancy between this result and the ones previously reported could be due to the bias of the dataset used in the previous study towards COMT genes. Based on this result, the role of motif C in the binding of SAM-OMT proteins to their substrates is unclear. One hypothesis is that motif C is important for the binding for only class II but not class I genes. The other hypothesis is that motif C is not involved in the binding, but in substrate utilization specificity. A third of class I sequences also lack motif I. Similar situation was found for motifs J, K, and L, which were distributed mainly in class II and a small fraction of class I. Furthermore, class II sequences present different patterns of gain/loss of motifs, J, K, and L. For

<table>
<thead>
<tr>
<th>Classes</th>
<th>Motif I</th>
<th>Motif J</th>
<th>Motif K</th>
<th>Motif L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class I (103 sequences)</td>
<td>61 (62.7%)</td>
<td>18 (17.48%)</td>
<td>20 (19.42%)</td>
<td>8 (7.77%)</td>
</tr>
<tr>
<td>Class II (59 sequences)</td>
<td>62 (98.4%)</td>
<td>36 (61.02%)</td>
<td>47 (79.67%)</td>
<td>37 (62.71%)</td>
</tr>
<tr>
<td>Class III (14 sequences)</td>
<td>8 (100%)</td>
<td>8 (57.14%)</td>
<td>9 (64.29%)</td>
<td>0 (0%)</td>
</tr>
</tbody>
</table>

Table 2: The occurrence (expressed in % in parenthesis) of COMT-specific motifs previously reported (Joshi and Chiang, 1998) in plant OMTs with a maximum of three mismatches.

Fig. 2. Distribution of OMT genes on \textit{Populus} chromosomes. The position of genes is indicated with an arrowhead. The names of the chromosomes and their sizes (Mb) are indicated below each chromosome, and are based on the new assembly of the \textit{Populus} genome (version 2). Segmental duplication homologous blocks, indicated with the same color, were mapped on the chromosomes based on duplication coordinates from the \textit{Populus} genome assembly version 2.
example, all *Arabidopsis* class II OMT genes, except the real COMT gene, lack motif J. It is unknown what the biological roles of these motifs are and whether they correlate with functional divergence of OMT genes.

Phylogenetic distribution showed that OMT genes are distributed in two major classes. Class I included simple phenol OMTs, flavonoid OMTs, and multifunctional COMTs (Li et al., 1997; Lam et al., 2007). Class II included all previously described real COMT and COMT-like genes from *Arabidopsis* and *Populus* (Raes et al., 2003; Tuskan et al., 2006), as well as a multifunctional OMT (PsyOMT) (Raes et al., 2003; Tuskan et al., 2006). It is noteworthy that gymnosperm sequences cluster much more closely to real COMT than other angiosperm COMT-like genes. This suggests that COMT-like genes evolved divergent protein sequences and maybe different functions. The phylogenetic tree of OMT genes presented here is supported by the distribution of SAM-OMT motifs where motif C was exclusively present in class II OMTs. Similarly, motifs K, L, and J were also mainly distributed in class II and multifunctional OMT group from class I. It is also supported by the intron–exon structure of genes showing that all class II *Populus* genes possess the same pattern 2. The phylogenetic distribution of OMT genes generated in this study is also in accordance with a previous study showing that all OMTs, except the ones that use acetyl CoA derivatives, were distributed in two major groups, B1 and B2 (Lam et al., 2007).

The phylogenetic tree showed evidence of an early duplication that had happen in the ancestor of land plants that generated the two major classes of SAM-OMT genes. The tree also showed evidence of several duplication events of which some may have happen in the ancestor of the rosids or the poaceae as it was suggested by a previous study (Tuskan et al., 2006). Segmental duplications seem to be a predominant process for the duplication of plant OMT genes. This was supported by the genome organization results showing that most OMT genes from *Populus* (this study) and *Arabidopsis* (results not shown) were located on duplicated blocks. Tandem duplication process seems to have also played a major role in the expansion of OMT gene family in plants. For instance, ~70% of OMT genes from *Populus* and a large fraction from *Arabidopsis* (data not shown) were tandem duplicated. Similar results were reported in a previous study in Apple (*Malus × domestica*) showing five OMT genes duplicated in tandem (Han et al., 2007). These finding showed that the amplification of OMT gene family is the result of various duplication–retention events that have happened at different times during evolution of land plants (Masterson, 1994; Pichersky and Gang, 2000; Han et al., 2007).

The deep rooting of OMT classes I and II in the phylogenetic tree of land plants indicates that evolution of COMT/COMT-like
(class II) genes and the other OMT genes from class I may have happen in the ancestor of seed plants. It also suggests that evolution of lignin in land plants correlates with the evolution of COMT genes. This result is in agreement with a previous study reporting a correlation between the evolution of lignin and another monolignol biosynthesis related gene family (cinnamyl alcohol dehydrogenase) (Barakat et al., 2010). It is noteworthy that three multifunctional OMT genes were distributed in classes I and II. The evolution of multifunctional OMT in both classes could be explained by the fact that the ancestral gene encoded a multifunctional protein. Another hypothesis is that genes from both classes evolved multi-functions independently.

The lingering question is why diverse copies of OMT genes from each class have been maintained within plant genomes. The
Retention of duplicate genes through sub-functionalization and neofunctionalization was suggested as a major process in the expansion of OMT gene family (Pichersky and Gang, 2000). Our analysis showed that while some duplicate genes have conserved similar expression profiles, others evolved modified or new ones. For instance, several genes from class I were differentially expressed in leaves or barks and two of them were induced following abiotic stresses. Blast analysis showed that these genes presented a high similarity to simple phenol, flavonoid OMT genes that function in plant defense against biotic and abiotic stresses. Among Populus class II genes only PoptrCOMT25, which corresponds to a real COMT gene (Tuskan et al., 2006), was preferentially expressed in xylem under normal growth conditions. The other Populus class II OMT genes were differentially highly expressed in leaves, bark, or both. Similar results were reported for 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. These variations in expression could be associated to variations of cis regulation elements previously reported (Toquin et al., 2003). It is unknown whether the COMT-like genes still function in lignin biosynthesis. Previous studies on another gene family (cinnamyl alcohol dehydrogenase) from the lignin biosynthesis pathway (Barakat et al., 2009; Barakat et al., 2010) showed that several CAD-like genes were induced under herbivory and other biotic and abiotic stresses, suggesting that some CAD-like genes still function in monolignol biosynthesis under stress conditions. These monolignols could end up in lignin polymers or lignans involved in plant defense. However, these genes could also have evolved, under the action of divergent or convergent evolution, different functions as it suggested by the high similarity of some COMT-like sequences to flavonoid OMT genes. Furthermore, AthOMT15 that cluster with real COMT genes was previously described as a flavonoid OMT gene (Lam et al., 2007).

Previous studies (Pichersky and Gang, 2000) suggested that the evolution of new cis regulatory elements added to aa substitutions have played a major role in the functional diversification of OMT genes. While results from this study suggest a functional divergence and specialization of Populus COMT-like genes from class II in various tissues, it is not clear whether these genes act differentially on various substrates or methylate the same substrate or both. It is also unknown whether they still function in lignin biosynthesis. It is likely that some duplicates evolved new functions and methylate different substrates as it was suggested by previous studies (Gang et al., 2002), while other still function in lignin biosynthesis under various stresses. Previous studies (Gang et al., 2002) showed that sequence similarity is not sufficient to predict the biochemical function of genes as the substrate specificity could be altered by the mutation of a single or few amino acids. However, this comprehensive study lay the basis for much needed functional and biochemical analysis of OMT genes to determine their biological roles.

In conclusion, this study reports on the annotation of a large set of OMT genes and the distribution of SAM motifs specific to different functional classes. Phylogenetic analyses showed that OMT genes were distributed in two classes that have evolved in the ancestor of land plants. The phylogenetic distribution, the protein and nucleotide sequence divergence, the SAM motif distribution, and the divergence of expression in different plant tissues and under various stresses of genes from class II are suggestive of functional divergence of genes from this class. The phylogenetic, comparative genomics, and expression analyses from this study will be useful for functional genomics studies addressing the biological role and the functional diversification of SAM-OMT genes.

Fig. 5. Expression of Populus OMT genes. Relative quantification of expression was analyzed in different tissues from non-treated (control) vs. herbivore-treated leaves. The name of each gene is indicated at the top of each histogram. Tissues studied are shown at the bottom of the diagrams. ± means SE of three biological replicate samples; *P<0.05, **P<0.01, ***P<0.001 according to Student's t-test. Y axis indicates the relative expression level of each gene compared to the control tissue (leaves).
Supplementary materials related to this article can be found online at doi:10.1016/j.gene.2011.02.008

Acknowledgments

We thank Dr. Dawn Luthe for access to the real-time RT-PCR machine. We thank our colleagues Chris Frost for performing the herbivore treatments. This work was supported by The Schatz Center for Tree Molecular Genetics at Penn State.

References


Hoffmann, L., et al., 2000. Impacts of herbivore treatments. This work was supported by The Schatz Center for Tree Molecular Genetics at Penn State.


Velasco, R., et al., 2007. A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. PLOS ONE 2, e1236.


Further reading


TIGR, J. Craig Venter Institute [http://www.jvci.org/].
学霸图书馆
www.xuebalib.com

本文献由“学霸图书馆-文献云下载”收集自网络，仅供学习交流使用。

学霸图书馆（www.xuebalib.com）是一个“整合众多图书馆数据库资源，提供一站式文献检索和下载服务”的24小时在线不限IP图书馆。

图书馆致力于便利、促进学习与科研，提供最强文献下载服务。

图书馆导航：

图书馆首页 文献云下载 图书馆入口 外文数据库大全 疑难文献辅助工具