



# Bioinformatic and expression analysis of the *OMT* gene family in *Pyrus bretschneideri* cv. Dangshan Su

X. Cheng, Y. Xiong, D.H. Li, J. Cheng, Y.P. Cao, C.C. Yan, Q. Jin, N. Sun, Y.P. Cai and Y. Lin

School of Life Science, Anhui Agricultural University, Hefei, Anhui, China

Corresponding author: Y.P. Cai

E-mail: ypcaiah@163.com

Genet. Mol. Res. 15 (3): gmr.15038664

Received March 29, 2016

Accepted May 18, 2016

Published September 2, 2016

DOI <http://dx.doi.org/10.4238/gmr.15038664>

Copyright © 2016 The Authors. This is an open-access article distributed under the terms of the Creative Commons Attribution ShareAlike (CC BY-SA) 4.0 License

**ABSTRACT.** With high nutritional value in its fruits, Dangshan Su pear has been widely cultivated in China. The stone cell content in fruits is a key factor affecting fruit quality in pear, and the formation of stone cells has been associated with lignin biosynthesis. O-Methyltransferase (OMT) is a key enzyme involved in lignin metabolism within the phenylpropanoid pathway. Here, we screened 26 *OMT* genes from the *Pyrus bretschneideri* cv. Dangshan Su genome using the DNATools software. To characterize the *OMT* gene family in pear, gene structure, chromosomal localization, and conserved motifs of *PbOMTs* were analyzed. *PbOMTs* were divided into two categories, type I (designated *PbCCOMTs*) and type II (designated *PbCOMTs*), indicating the differentiation of function during evolution. Based on the analysis of multiple sequence alignment, cis-element prediction, and phylogenetic relationships, two candidate genes, *PbCCOMT1* and *PbCCOMT3*, were selected for the analysis of temporal and spatial gene expression

in pear. The promoter regions of both *PbCCOMT1* and *PbCCOMT3* contain regulatory motifs for lignin synthesis. Moreover, the two genes show high similarity and close phylogenetic relationships with *CCOMTs* in other species. Expression analysis showed that transcript levels of two *PbCCOMTs* were positively associated with the contents of both stone cells and lignin during the development of pear fruit. These results suggest that *PbCCOMT1* and *PbCCOMT3* are closely associated with lignin biosynthesis. These findings will help clarify the function of *PbOMTs* in lignin metabolism and to elucidate the mechanisms underlying stone cell formation in pear.

**Key words:** *OMT* gene family; *Pyrus*; Lignin synthesis; *PbCCOMTs*; Phylogenetic analysis

## INTRODUCTION

*Pyrus bretschneideri* cv. Dangshan Su originated in Dangshan County, Anhui Province and is widely planted in China. As its fruit is very nutritious and has important medicinal value, this variety is in great demand both at home and abroad. In recent years, the number of stone cells in the fruits of Dangshan Su pears has increased as a result of the deterioration of field environments and field management (Cai et al., 2010; Jin et al., 2013). Stone cells are formed by parenchyma cells following the secondary thickening and lignification of primary walls. This acts as a type of sclerenchymatous cells in the pear pulp, which determines the internal quality of pear fruits (Cai et al., 2010; Jin et al., 2013). An increased concentration of stone cells in pear fruit leads to a muricated taste, poor edible quality, and loss of economic value (Choi and Lee, 2013; Zhao et al., 2013; Yan et al., 2014). During the formation of stone cells, lignin is deposited in the corner of the primary wall of parenchyma cells, from where it extends to the intercellular and secondary wall layers along the primary wall corner until the whole cell lumen is covered. Ultimately, a solid stone cell is formed (Lu et al., 2011; Jin et al., 2013). Therefore, lignin biosynthesis is thought to be involved in the generation of enlarged stone cells in pear fruit and in their presence in increased numbers. As a key enzyme in the phenylpropane pathway, O-methyltransferases (OMTs) are involved in the synthesis of a series of secondary metabolites such as lignins, flavonoids, and phytoalexins. OMTs can be divided into two types: type I, a caffeoyl-CoA-3-O-methyltransferase (CCOMT), which is a key enzyme involved in lignin biosynthesis; and type II, a caffeic acid O-methyltransferase (COMT), which is involved in lignin metabolism and flavonoid methylation (Joshi and Chiang, 1998). During lignin biosynthesis and using caffeoyl-CoA as a substrate, CCOMT transfers the methyl group from S-adenosylmethionine (SAM) to the lignin precursor benzene at the third carbon position, generating feruloyl-CoA. Conversely, COMT is responsible for methylation of lignin precursors such as coniferaldehyde and 5-hydroxyconiferaldehyde (Joshi and Chiang, 1998).

Both types of *OMTs* exist in plants, and individual members of this family have different functions. In mutant screens, the *Atomt1* mutant of *Arabidopsis thaliana* can be identified by its reduced levels of syringyl (S) units and flavonoid, and its increased levels of 5-hydroxyguaiacyl, which is a precursor of S-units. These results suggest that *AtOMT1*

is associated with the metabolism of both lignin and flavonoids (Goujon et al., 2003). An *Arabidopsis* mutant (*ccomt1*) was found to have decreased lignin content in the stem and an increased ratio of S/G lignin units. An *Arabidopsis* double mutant (*comt1 ccomt1*) showed lower levels of G units than were found in wild-type *Arabidopsis* (Do et al., 2007).

*CCOMT* genes have been targeted in order to lower the total lignin content while increasing the ratio of S/G lignin units via RNA interference technology. For example, the suppression of *PrCCOMT* expression by interference RNA led to a 20% reduction in the total lignin content (Wagner et al., 2011). Suppression of *CCOMT* expression in maize showed that the lignin content was decreased and the ratio of S/G lignin units was increased (Li et al., 2013). Research showed that heterologously expressed anti-*AcCCOMT* increased the levels of S units, G units, and total lignin in tobacco (Pang et al., 2014). The results of previous studies have shown that *PbCCOMT* genes are expressed at high levels during the early stages of fruit development, which might promote the accumulation of pear stone cells (Wu et al., 2012). Furthermore, a higher ratio of G units can lead large trouble in degradation of polymeride lignin, which subsequently increases the degree of sclereid polymerization, and seriously affects the quality and taste of pear fruit (Yan et al., 2014). This indicates that *CCOMT* is a valuable regulatory gene.

van Parijs et al. (2015) analyzed the phylogenetic evolution of *COMTs* in *Lolium perenne*. Wang et al. (2013) screened nine *OMT* genes involved in lignin synthesis from the whole genome of *Eucalyptus* using methods such as RNA-seq, qRT-PCR, and a composite evolutionary tree. However, no recent studies on the *OMT* gene family in pears have been performed using genome-wide methods. In the present study, based on the published pear genome sequence, the *OMT* gene family in pear was investigated to analyze its chromosome distribution, phylogenetic relationships, conserved motifs, and the expression pattern of candidate *PbCCOMTs* relevant to lignin biosynthesis at the genome-wide level. These results should provide some useful information for further studies on the functions of *PbOMTs*, developmental mechanisms of stone cells, and regulation of pear lignin metabolism.

## MATERIAL AND METHODS

### Identification and sequence analysis of *OMT* family genes in pear

The pear genome sequence was downloaded from GigaDB (<http://gigadb.org/dataset/100083>), and the DNATOOLS software was used to construct a local database from the protein sequences of the latest complete pear genome. Two *Arabidopsis OMTs* (At5g54160 and At4g34050), which are relevant to lignin biosynthesis, were used as initial query sequences to search the pear protein database using the BLASTp program ( $P = 0.001$ ). Redundant sequences were then removed manually. Hidden Markov Models of the Pfam (<http://pfam.sanger.ac.uk/search>) and SMART (<http://smart.embl-heidelberg.de/>) databases were used to confirm each candidate sequence containing the signature methyltransf 2 (PF00891) or methyltransf 3 (PF01596) domains of *OMT* genes. The molecular weight (MW) and isoelectric point (pI) of proteins were calculated with the Compute pI/Mw tool ([http://web.expasy.org/compute\\_pi/](http://web.expasy.org/compute_pi/)) (Bjellqvist et al., 1993). Information on *PbOMTs*, including amino acid sequences, chromosome location, clone number, and open reading frame length was obtained from the pear genome sequence database.

## Phylogenetic analysis, conserved protein motifs, and prediction of gene structure

Phylogenetic relationships were analyzed using MEGA 5.1 (Tamura et al., 2011) with the neighbor-joining (NJ) method (bootstrap = 1000). Information on proteins used for phylogenetic analysis is listed in [Table S1](#).

The conserved motifs encoded by each *OMT* gene were identified using the Multiple Em for Motif Elicitation (MEME) program (<http://meme-suite.org/tools/meme>) (Bailey et al., 2006). Parameters were set as follows: optimum motif width  $\geq 6$  and  $\leq 200$ ; maximum number of motifs 20. To predict the exon-intron structure of the *PbOMT* genes, the genomic sequences and their predicted coding sequences were compared using GSDS (<http://gsds.cbi.pku.edu.cn/>).

## Estimating Ka/Ks ratios for duplicated gene pairs

Duplicated *OMT* gene pairs in pear were aligned by ClustalW. Next, the Ka (non-synonymous substitution rate) and Ks (synonymous substitution rate) values were calculated using the DnaSP v5.0 software (Rozas and Rozas, 1995), and the Ka/Ks ratio was analyzed to assess the selection pressure for each gene pair.

## Chromosomal locations and gene duplications

Individual *OMT* genes were mapped on pear chromosomes according to their starting positions provided in the genome annotation document. The image showing chromosome location was generated by the MapInspect software (<http://mapinspect.software.informer.com>).

We defined gene duplication according to the following criteria: 1) the aligned sequence covered  $>80\%$  of the longer gene, 2) the identity of the aligned regions was  $>80\%$ . Based on the chromosomal locations of the *OMT* genes, two types of gene duplications were recognized: tandem duplication and segmental duplication (Kong et al., 2013; Wei et al., 2013).

## Alignment of amino acid sequences encoded by *OMTs* and analysis of their secondary components

Protein sequences of pear and *Arabidopsis* were aligned using Iden and Sim in Sequence Manipulation Suite (Stothard, 2000). Multiple comparisons of each amino acid sequence were carried out using the BioEdit software and secondary elements were analyzed using the online software Esript3.0 (Robert and Gouet, 2014).

## Cis-element analysis and interaction networks

For promoter analysis, 2000-bp genomic DNA sequences from upstream of the initiation codon (ATG) were retrieved from the pear genome sequence. Analysis of cis-acting regulatory DNA elements (cis-element) was carried out using the PLANT CARE program (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) (Magali et al., 2002). Functional protein association networks were integrated in the STRING software (<http://string-db.org/>).

## RNA extraction and reverse transcription

To examine the expression of *PbOMT* genes, buds, young stems, mature leaves, and fruits were obtained from 50-year-old pear trees grown on a farm in Dangshan, Anhui, China. All samples were stored in liquid nitrogen until use. Fruits were collected at different days after flowering (DAF) (15, 39, 47, 55, 63, and 145 DAF).

Total RNA was isolated from all samples using a total RNA isolation kit (Tiangen, China). Reverse transcription was performed using a M-MLV Rtase cDNA Synthesis Kit (TaKaRa, China).

## Primer design and semi-quantitative RT-PCR

Primers were designed using Primer premier 5 and are listed in [Table S2](#). Semi-quantitative RT-PCR as performed as follows: predegeneration for 5 min, followed by 30 cycles of degeneration for 30 s at 94°C, annealing for 30 s at 55°C, extension for 1 min at 72°C, and a final 72°C extension for 10 min. Each analysis was set up in triplicate.

## RESULTS

### Identification of the *OMT* genes in pear

A total of 39 candidate genes were identified, and Pfam and SMART databases were subsequently used to confirm the presence of the methyltransf 2 or methyltransf 3 domain in all candidate genes. As a result, 26 members of the *OMT* gene family were identified in the pear genome, including of 19 *COMT* genes and seven *CCOMT* genes, designated as *PbCOMT1* to *PbCOMT19*, and *PbCCOMT1* to *PbCCOMT7*, respectively (Table 1).

**Table 1.** Sequence characteristics of 26 *OMT* genes identified in pear.

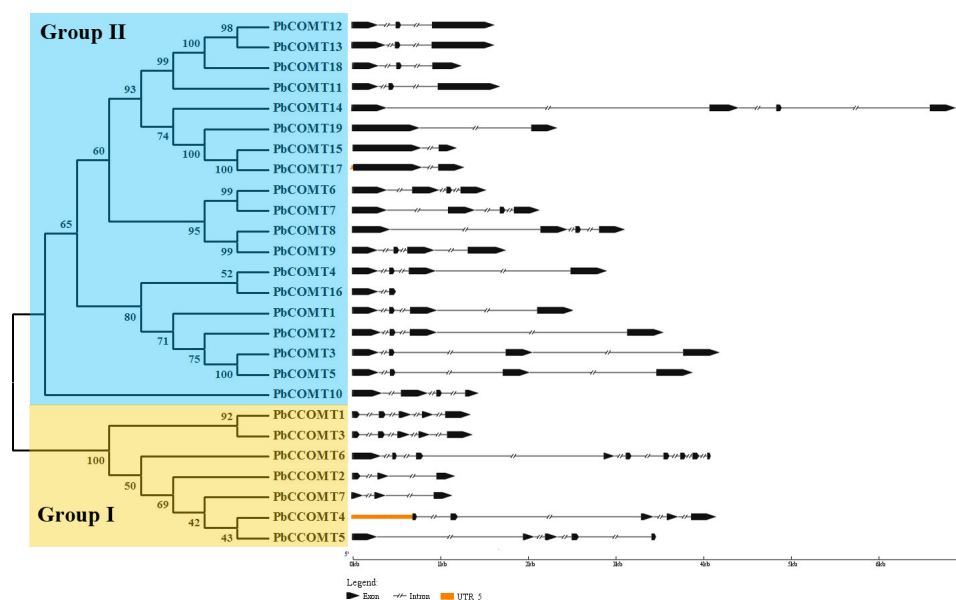
Name	Sequence ID	Chromosome	Length (aa)	MW (kDa)	pI	ORF (bp)
<i>PbCOMT1</i>	Pbr013510.1	Chr1	365	39.74	5.63	1096
<i>PbCOMT2</i>	Pbr013512.1	Chr1	377	41.00	5.67	1132
<i>PbCOMT3</i>	Pbr040044.1	Chr7	367	40.05	5.63	1102
<i>PbCOMT4</i>	Pbr040039.1	Chr7	365	39.59	5.75	1096
<i>PbCOMT5</i>	Pbr040045.1	Chr7	367	39.87	5.82	1102
<i>PbCOMT6</i>	Pbr035407.1	/	358	39.48	5.28	1075
<i>PbCOMT7</i>	Pbr025887.1	Chr10	358	39.00	5.70	1075
<i>PbCOMT8</i>	Pbr007791.1	Chr1	370	41.03	5.78	1111
<i>PbCOMT9</i>	Pbr025883.1	Chr10	370	40.97	5.96	1111
<i>PbCOMT10</i>	Pbr003001.1	/	291	32.35	6.16	874
<i>PbCOMT11</i>	Pbr021100.1	Chr15	361	40.02	5.59	1084
<i>PbCOMT12</i>	Pbr021099.1	Chr15	363	40.19	6.08	1090
<i>PbCOMT13</i>	Pbr021098.1	Chr15	392	43.38	6.14	1177
<i>PbCOMT14</i>	Pbr032564.1	Chr12	369	40.53	5.26	1108
<i>PbCOMT15</i>	Pbr030677.1	Chr9	332	36.50	5.78	997
<i>PbCOMT16</i>	Pbr040038.1	Chr7	126	13.76	6.27	379
<i>PbCOMT17</i>	Pbr030676.1	Chr9	362	39.94	6.10	1087
<i>PbCOMT18</i>	Pbr039628.1	Chr10	236	26.51	6.25	709
<i>PbCOMT19</i>	Pbr036056.1	Chr17	358	39.39	5.38	1075
<i>PbCCOMT1</i>	Pbr034039.1	Chr17	247	27.86	5.43	742
<i>PbCCOMT2</i>	Pbr028157.1	/	149	16.34	7.66	448
<i>PbCCOMT3</i>	Pbr038709.1	/	247	27.82	5.32	742
<i>PbCCOMT4</i>	Pbr019305.1	Chr9	235	26.70	5.43	706
<i>PbCCOMT5</i>	Pbr028166.1	/	235	26.31	5.14	706
<i>PbCCOMT6</i>	Pbr025246.1	Chr5	307	34.12	8.91	922
<i>PbCCOMT7</i>	Pbr015171.1	Chr16	161	17.59	7.66	484

“/” indicates that the specific chromosome is not known. MW: molecular weight; ORF: open reading frame.

To characterize PbOMT proteins, their physicochemical properties (pI and MW) were analyzed. As shown in Table 1, most *PbCOMT* sequences were longer and had a higher MW than those of *PbCCOMT* genes. *PbCOMT* proteins were between 126- and 392-aa long, with maximum and minimum molecular weights of 43.38 and 13.76 kDa, respectively, and isoelectric points between 5.26 and 6.27. *PbCCOMT* proteins are between 149- and 307-aa long, with maximum and minimum molecular weights of 34.12 and 16.34 kDa, and isoelectric points from 5.14 to 8.91.

### Intron-exon structure and phylogenetic analysis of pear *OMT* genes

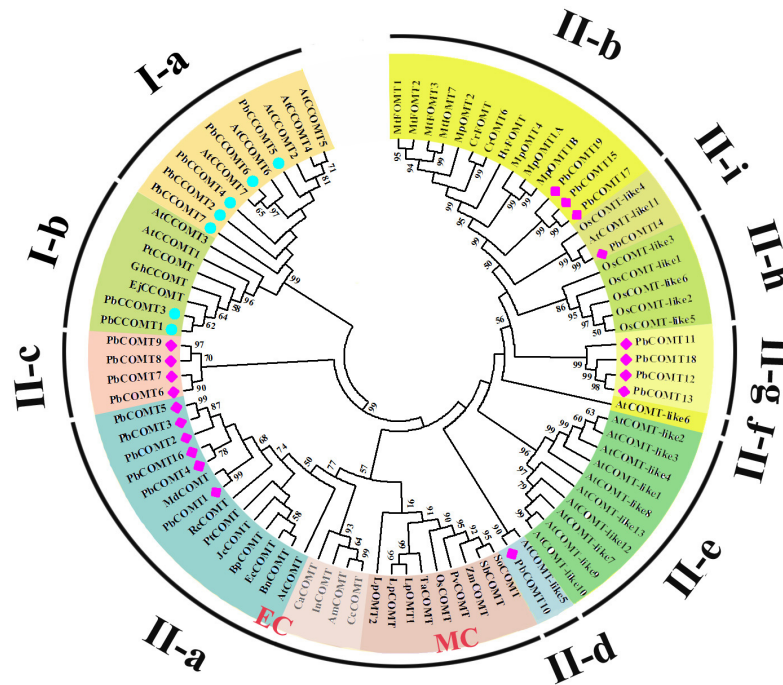
Based on the phylogenetic analysis (Figure 1), 26 *OMT* genes were divided into two major groups, with *PbCOMTs* in group I (seven genes) and *PbCCOMTs* in group II (19 genes). This implies that there were evolutionary differences between the two groups. Twenty-six *OMT* genes were clustered into eight gene pairs with high bootstrapping values, with the exception of *PbCCOMT4/5* and *PbCOMT4/16*. We further analyzed the structures of 26 *PbOMT* genes in order to support the utility of the phylogenetic tree in the clarification of evolution relationships between *PbOMTs*. The results showed that these gene structures were very similar within the same subtypes, which revealed their close evolutionary relationship. Based on the number of introns in pear *OMT* genes, the *OMTs* can be divided into five types. Among 26 *OMT* genes, except eight introns that were located in *PbCCOMT6*, in the rest of the family numbers, there were 4 genes with 4 introns, 11 genes with 3 introns, 6 genes with 2 introns, and 4 genes with 1 intron.



**Figure 1.** Phylogenetic relationships and gene structure of pear *OMT* proteins.

To further clarify the genetic relationships and to predict the biological function of *OMT*-gene products, 87 *OMT* proteins from different species were used to construct a NJ tree

(Figure 2). These proteins can be divided into two classes (classes I and II). Class I consists of all *CCOMT* genes, which are further divided into two subclasses (subclass I-a and I-b).



**Figure 2.** Neighbor-joining phylogenetic tree based on 87 OMT protein sequences from various plants. Diamonds represent PbCOMTs; blue dots represent PbCCOMTs. COMTs from subclass II-a were divided into two groups: a monocot cluster (MC) and a eudicot cluster (EC). Navy blue and pale pink shading represents the EC, pink shading represents the MC.

Five *Arabidopsis* *CCOMTs* and five pear *CCOMTs* were included in subclass I-a. Subclass I-b included two *PbCCOMTs* and two *AtCCOMTs*. Among them, *AtCCOMT3* (At4g26220) is associated with the methylation of flavonone and flavanone (Wils et al., 2013). *AtCCOMT1* and *PtCCOMT* are associated with lignin synthesis (Do et al., 2007; Du et al., 2013), *PbCCOMT1* and *PbCCOMT3* gathered for a subclass with these two genes that may imply they are involved in lignin metabolism process.

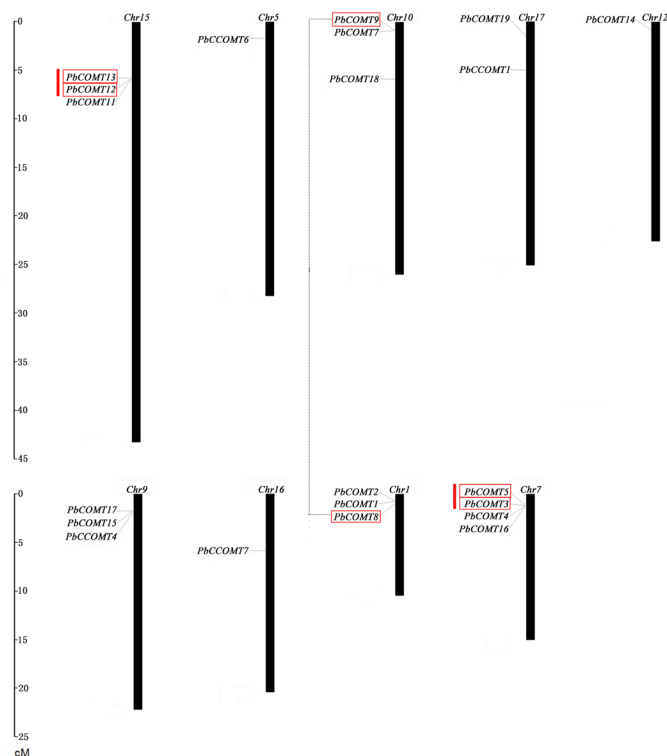
Class II included all *COMTs*, which were further divided into nine subclasses (II-a-i). Pear *COMTs* were distributed in subclasses II-a, -b, -c, -d, -g, and -i. Within subclass II-c and -g, eight pear *COMTs* were included. Only *AtCOMT*- or *OsCOMT*-like genes were included in subclasses II-e, -f, and -h, which is likely to be associated with species differentiation. *PbCOMT15*, -17, and -19 were included in subclass II-b, which belongs to the flavonoid OMTs, suggesting that these three *PbCOMTs* may be associated with flavonoid methylation (Kim et al., 2010; Sung et al., 2011). Most of the genes in subclass II-a are involved in lignin biosynthesis and include sequences from both monocots and eudicots, which could be further subdivided into two groups. The eudicot cluster could clearly be divided into two clades due to species variation. *PbCCOMTs* were grouped in subclass II-a of the eudicot cluster. Hirano et al. (2012) classified lignin metabolism related *ZmCOMT*, *SbCOMT*, *OsCOMT*, *LpCOMT*,

*LpCOMT1*, *LpCOMT3*, *AtOMT* as true *COMTs*, including *PbCOMT1*~*5* and *PbCOMT16*. There is preliminary evidence that one pear *COMT* (accession No. KC905086), *PbCOMT1*, is associated with lignin metabolism. The functions of the other five *PbCOMTs* in this group require further investigation.

The NJ tree also showed that *PbCOMT14* and *OsCOMT-like4* are closely related to *AtCOMT-like11*. Although some *OMT* genes in rice and pear could be grouped with *AtOMTs*, the bootstrap values in the tree nodes were low. For example, the bootstrap value of the *PbCCOMT6* containing clade was only 65%.

### Chromosomal localization and duplication of *OMT* genes in pear

To define the distribution of *PbOMTs* on chromosomes, and to analyze gene duplication events, a chromosome location analysis was performed. A total of 21 of 26 *OMT* genes were distributed on chromosomes 1, 5, 7, 9, 10, 12, 15, 16, and 17 (Figure 3), and most were located at the end of the chromosomes. As shown in Figure 3, chromosome 7 contains the maximum number of genes (four), with one each on chromosomes 5, 12, and 16. The majority of these *COMT* genes are present as clusters in the pear genome. There is one gene cluster on chromosomes 1, 7, 9, 10, and 15, respectively. *PbCCOMT4* forms a gene cluster with *PbCOMT15* and *PbCOMT17*, suggesting their close relationship or co-regulation.



**Figure 3.** Chromosomal locations of pear *OMT* genes on nine chromosomes. Genes involved in segmental duplication are joined by dashed lines, and the red rectangle indicates the gene cluster on each chromosome.



Based on the evolutionary relationships and chromosomal distribution of the *PbOMT* genes, one gene pair involved in segmental duplication and two pairs involved in tandem duplication were identified (Table 2). Generally, a Ka/Ks ratio of >1 indicates accelerated evolution with a positive selection, a ratio of 1 indicates pseudogenes with neutral selection, while a ratio <1 indicates functional constraints with negative or purifying gene selection. In this study, the Ka/Ks ratios of three gene pairs were all less than 1, indicating that their evolution was strongly influenced by purifying selection pressure although there was no strong differentiation for function.

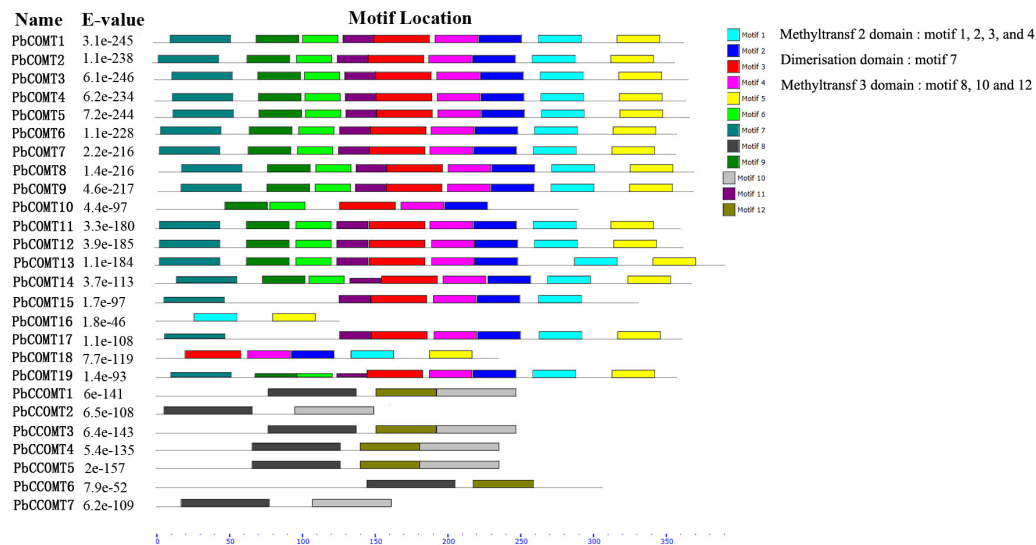
**Table 2.** Ka/Ks analysis for duplicated *OMT* genes in pear.

Duplicated gene 1	Duplicated gene 2	Ka	Ks	Ka/Ks	Purifying selection	Duplicated type
<i>PbCOMT8</i>	<i>PbCOMT9</i>	0.043	0.150	0.287	Yes	Segmental
<i>PbCOMT3</i>	<i>PbCOMT5</i>	0.027	0.047	0.574	Yes	Tandem
<i>PbCOMT12</i>	<i>PbCOMT13</i>	0.030	0.065	0.461	Yes	Tandem

### Sequence conservation analysis

To further unrevealing the relationship among the members of *OMT* gene family, the distribution of the conserved motifs in individual member was analyzed using MEME software online. A total of 12 conservative motifs were identified (Figure 4 and [Figure S1](#)).

In addition, the *OMT* family could be divided into *COMT* and *CCOMT* types, based on clear differences in their conserved motifs. Using the Pfam software, these conserved motifs were further annotated. As shown in Figure 4 and Table 3, motifs 1-4 encode a methyltransf 2 domain, which only exists in *PbCOMT*s, and is mostly found in the middle of the sequence. All the *PbCOMT* proteins include these four motifs, except for *PbCOMT16*.



**Figure 4.** Distribution of 12 putative conserved motifs in pear *OMT* proteins.

**Table 3.** Motif composition of OMT proteins from pear.

Motif No.	Width	Conserved amino acid sequences
1	29	WICHDWSDDHCIKLLKNCWNAIPDNGKVI
2	29	IKGINFDLPHVIKHAPQYPGEVHVGGDMF
3	38	PFNKAHGMMTFEYHSTDPFRNQVFNKAMACHSTICMCK
4	29	YKGFEGVTQLVDVGGGTGNAIWMIVSKYP
5	29	MMAQNPGGKERTEEEFMALAKGAGFQGFR
6	24	RLYGLTPVCKCFVTNEDGVSLAPL
7	41	QRREEENYCFAMQLANSSVLQVMQAAIELDFDIMAKAGP
8	60	MNAKKTMEELGVYTGYSLLTTALALPDDGKITAIDVNKEAYEVGLPFIQKAGVEHKINFCQ
9	29	DAPVMLDRMLRLLASYSILTCSLRTPDD
10	54	WFGTVAEPEENVQEYARYRKHVLELNSFLAADPRIELCMVSVGDGLTLCRRLC
11	21	QDKVFVESWYHLKDAVLEGGI
12	41	EEYGSFDFAFVDADKDNMNYHKRLIDLKVKVGLIAYDNTL

Motif 7 encodes a dimerization domain, which is also only present in PbCOMTs. The dimerization domain is annotated using Pfam as follows: it was found at the N-terminal of OMT in many plants, indicating its relevance to the formation of protein dimers. Accordingly, it is likely that COMT proteins are only present in monomeric forms because of the absence of a dimerization domain. Motifs 8, 10, and 12 encode a methyltransferase domain, which is only found in PbCCOMTs. Therefore, motifs 8, 10, and 12 could be used as feature motifs to distinguish COMTs and CCOTs. Other motifs, such as 5, 6, 9, and 11 only exist in PbCOMTs, and their functions are unknown. Next, we used the OMT orthologs to construct the interaction network using the STRING software in *Arabidopsis* (Figure S2). OMT1 (PbCOMT1) was related to AT1G15950, 4CL2 (Pb4CL) and CCoAOMT1 (PbCCOMT1; PbCCOMT3).

Based on the analysis of conserved motifs, the sequence features of the two major types of genes in the OMT family, the COMTs and CCOTs, were analyzed at the protein level by multiple comparison (Figure 5). The results of previous studies have found that the conserved domain in the COMT gene possesses the following motifs: (V, I, L) (V, L) (D, K) (V, I) GGXX (G, A), GINFDLPHV, EHVGGDMF, NGKVI, and GGKERT (Joshi and Chiang, 1998). As shown in Figure 5, the five elements found in the PbCOMT proteins are relatively well conserved, with the exception of PbCOMT16, in which the first two motifs have been lost. Thus, these specific motifs could be used as features to identify COMT genes. Among the different PbCOMT proteins, the amino acid composition of these five motifs differed because of their specific substrates.

In plants, CCOT has been found with three SAM-binding motifs, namely A, B, C, and other characteristic motifs, namely E, H, G, D, F (Joshi and Chiang, 1998). In the present study, these eight motifs were also identified in PbCCOMTs, with the loss of motif D in PbCCOMT7 and motif G in PbCCOMT2 and PbCCOMT7. It is likely that the differential distribution of motifs among the PbCCOMTs will be related to their functional differentiation. All screened PbOMTs possess the characteristic structure of full-length OMT proteins, which indicates that the results obtained by screening and identification are reliable.

### Identification and spatio-temporal expression analysis of candidate PbCCOMTs in *P. bretschneideri* cv. Dangshan Su

To date, evidence has clearly shown that the COMT and CCOT genes from *Arabidopsis* and maize are associated with lignin biosynthesis (Do et al., 2007; Li et al., 2013).



**Figure 5.** Multiple sequence alignment of OMT proteins from pear, maize, and *Arabidopsis*. A: CCOMTs; B: COMTs. Conserved sequence motifs in COMT are boxed and bolded; three SAM-binding motifs, A, B, and C, are indicated below; CCOMT signature motifs are indicated as D, E, F, G, and H.

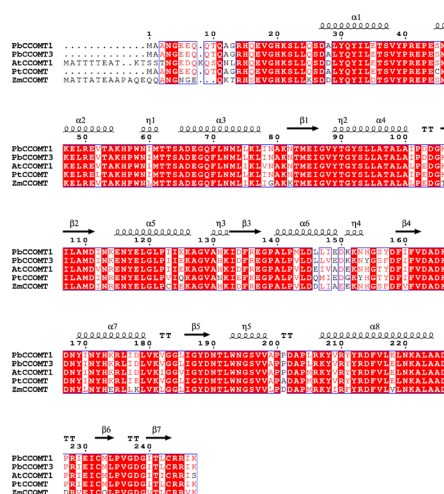
In order to screen pear *CCOMT* genes related to lignin biosynthesis, Iden and Sim in Sequence Manipulation Suite methods were used to analyze the identity and similarity of seven pear *CCOMT* genes and to compare with those from *Arabidopsis* and maize (Table 4). Comparisons based on amino acid sequence further demonstrate that PbCCOMT1 and PbCCOMT3 are more similar to AtCCOMT1 and ZmCCOMT than to the other PbCCOMTs, indicating that PbCCOMT1 and PbCCOMT3 may act as CCOMT enzymes. This was consistent with the results of previous evolutionary tree clustering in this study. Therefore, *PbCCOMT1* and *PbCCOMT3* were used as candidate genes for further study.

**Table 4.** Sequence identity and similarity among *Arabidopsis*, maize, and pear *CCOMT* protein homologs.

Iden/Sim (%%)	AtCCOMT1	ZmCCOMT
PbCCOMT1	85/86	77/81
PbCCOMT2	31/33	33/35
PbCCOMT3	85/86	76/80
PbCCOMT4	49/54	51/56
PbCCOMT5	46/50	48/53
PbCCOMT6	27/34	29/35
PbCCOMT7	33/36	33/37

Iden: identity; Sim: similarity.

The amino acid sequences of *PbCCOMT1* and *PbCCOMT3* were aligned with those of three other *CCOMTs* related to lignin biosynthesis, and using the ESPrpt tool, the distribution of secondary structure elements was screened (Figure 6). Higher conservation was observed among the five *CCOMTs*. The secondary structures of PbCCOMT1 and PbCCOMT3 are mainly composed of  $\alpha$  helices and random curls, with some stretching sequences and  $\beta$  angles. The secondary structure of the proteins encoded by *PbCCOMT1* and *PbCCOMT3* contain 36.44/34.01%  $\alpha$  helix, 21.05/23.89% extension chain, 12.96/11.34%  $\beta$  angle, and 29.55/30.77% random curl, respectively.



**Figure 6.** Alignment of the protein sequences encoded by *PbCCOMT1* and *PbCCOMT3* and other lignin CCOMTs. Secondary structure elements extracted from the protein structure of MsCCOMT (PDB id code: 1SU1) are indicated on top of the alignment using Esprpt. The arrows indicate  $\beta$ -sheet structure, the helices indicate helical structure motifs, and TT indicates turns.

The promoter sequences in *PbCCOMT1* and *PbCCOMT3* affects their expression. Both the *PbCCOMT1* and *PbCCOMT3* promoters possess light-responsive elements (Table 5), indicating their expression may be affected by light.

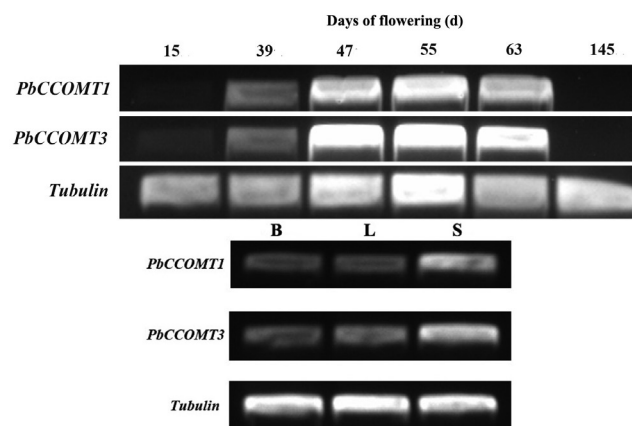
**Table 5.** Nucleotide motifs identified in promoter regions of putative lignin *CCOMT* genes in pear.

Motif name	Motif location		Motif function
	<i>PbCCOMT1</i>	<i>PbCCOMT3</i>	
AC-I	1298	178,1290	Most of the MYBs in this network can bind to the AC elements (I, II, and III) in the promoters of lignin biosynthesis genes and regulate their expression
AC-II	1299,1300	1291	
WUN-motif	-	273	Wound-responsive element
TGACG-motif	796	525,782	Cis-acting regulatory element involved in MeJA-responsiveness
MBS	-	1146	MYB Binding Site involved in drought-inducibility
Sp1	547	134	Light responsive element
TC-rich repeats	1444	405,1432	Cis-acting element involved in defense and stress responsiveness
GARE-motif	440,1153	261	Gibberellin-responsive element
HSE	462,521	965	Cis-acting element involved in heat stress responsiveness
LTR	-	749	Cis-acting element involved in low-temperature responsiveness
TCA-element	45,154,237,810, 1143	226,1203, 1204,1431,1471	Cis-acting element involved in salicylic acid responsiveness

MBS: MYB binding site; HSE: heat shock element; LTR: low temperature responsive.

In addition, the two *PbCCOMT* promoters also contain a large number of cis-elements for biological and abiotic stress responses, these include a WUN-motif, MBS, TC-rich repeats, HSE, LTR, TCA-element, and hormone-response related elements, such as those sensitive to methyl jasmonic acid (MeJA), salicylic acid, and gibberellin acid. Moreover, *PbCCOMT1* and *PbCCOMT3* have AC components, which are generally associated with the regulation of lignin metabolism by combining MYB transcription factors (Raes et al., 2003; Liu et al., 2015). This indicates that the two genes play an important role in lignin biosynthesis.

Semi-quantitative RT-PCR was used to analyze the expression of candidate pear *CCOMTs* in different plant parts and at different development periods. As shown in Figure 7, expression of *PbCCOMT1* and *PbCCOMT3* was detected in the mature leaves, young stems, buds, and fruits. This may be because of the high lignin content in stems, therefore the expression of these genes are higher in stems than buds and mature leaves.



**Figure 7.** Expression patterns of candidate *CCOMT* genes in pear. *PbCCOMT1* and *PbCCOMT3* expression patterns in pulp during the development of pear fruits. *PbCCOMT1* and *PbCCOMT3* expression patterns in pear buds (B), mature leaves (L), and young stems (S).

The expression of *PbCCOMT1* and *PbCCOMT3* followed similar trends in pear fruits. Low levels of these genes were present in fruit at 15 and 145 DAF. The expression levels in fruit from 47 to 63 DAF were significantly higher than those in mature leaves, young stems, and buds. *PbCCOMT1* was maximally expressed 55 DAF. The levels of *PbCCOMT3* expression between 47 and 63 DAF were higher than those of *PbCCOMT1*. These results are consistent with the trend observed in pear for lignin and stone cell contents (Cai et al., 2010; Lu et al., 2011), indicating that the two candidate *CCOMT* genes are probably involved in stone cell development and lignin biosynthesis in DangShan Su pear fruit.

## DISCUSSION

Various members of the *OMT* gene family are involved in multiple branched pathways in phenylpropanoid metabolism and are mainly responsible for modification of plant secondary metabolite methylation. Two categories of *OMTs* exist: *COMTs* and *CCOMTs*. During lignin metabolism, *COMT* catalyzes 5-hydroxy-coniferyl aldehyde, caffeic acid, and 5-hydroxy-coniferyl alcohol methylation to generate ferulic acid, sinapic alcohol, and sinapic aldehyde in the synthesis of S-lignin, respectively (Osakabe et al., 1999), *CCOMT* is responsible for transferring caffeoyl CoA to feruloyl CoA, and the latter subsequently participates in G-lignin synthesis.

The recent completion of whole genome sequencing in *Arabidopsis* (Raes et al., 2003), *Populus trichocarpa* (Shi et al., 2010), and *Eucalyptus grandis* (Carocha et al., 2015) has shown that these model organisms contain *COMTs* and *CCOMTs*. A total of 13 *COMT-like* genes and seven *CCOMT* genes exist in *Arabidopsis*. In the present study, *AtCOMT-like5*, *-6*, and *-11* were grouped along one evolutionary branch, and the remaining 10 genes were grouped in another class, similar to previously reports (Raes et al., 2003). Among seven *CCOMTs*, only *AtCCOMT1* belongs to the true *CCOMT* clade involved in lignin biosynthesis (Sibout et al., 2005; Do et al., 2007). *PbCCOMT1* and *PbCCOMT3* were grouped with *AtCCOMT1*, implying that they have a role in lignin metabolism. Although the functions of other members of this gene family remain unclear, *AtCCOMT2*, and *AtCCOMT4-7*, were expressed at different development periods in inflorescence stems (Raes et al., 2003). Six *CCOMTs* and 24 *COMTs* were isolated from *P. trichocarpa* and seven *COMTs* and two *CCOMTs* were isolated from *E. grandis*, respectively. In most plants, including pear, there are more members of *COMTs* than there are of *CCOMTs*. In this study, 19 *PbCOMTs* and seven *PbCCOMTs* were identified. *CCOMTs* should have different functions, such as involvement in lignin and flavonoid biosynthesis in pears. These results are consistent with previous reports that some *CCOMTs* can catalyze the synthesis of flavonoids in *Arabidopsis* and *Camellia sinensis* (Wils et al., 2013; Zhang et al., 2015).

Analysis of chromosomal localization indicated that members of the *PbOMT* gene family are distributed in the pear genome in the form of gene clusters. Together with the phylogenetic relationship, these results showed that gene duplication events within the *OMT* gene family played a significant role in genome expansion in pear. In the present study, characteristics of *PbOMTs* were analyzed based on multiple comparisons of amino-acid sequences. The results from the sequence comparison and analysis of the conservative motifs, showed that both *PbCOMTs* and *PbCCOMTs* possessed the conserved domains shared by these proteins in other species. However, different members of the *PbOMT* gene family diverged and obtained various functions during the course of evolution. Generally, it is thought

that products of genes with similar sequences should carry similar functions. We found that PbCCOMT1 and PbCCOMT3 were highly similar to AtCCOMT1 and ZmCCOMT, and showed that PbCCOMT1 and PbCCOMT3 are involved in lignin metabolism in the same way as their homologous proteins AtCCOMT1 and ZmCCOMT.

The AC, MeJA-, and light-responsive elements in *PbCCOMT1* and *PbCCOMT3* were identified by cis-element analysis. Gharechahi et al. (2013) identified different *CCOMT* proteins from the hairy root of *Silybum marianum* L. treated by MeJA. Wang et al. (2013) found that the contents of stone cells and lignin decreased by 31.6 and 13.9%, respectively, and that the quantity and diameter of stone cells was reduced by 28.0 and 45.8%, respectively, in bagged pear fruits. This is likely because bagging affects the light intensity. Thus, MeJA and light might affect the expression of *CCOMTs* during stone cell development. During production, we can attempt to control light levels or spray MeJA onto the fruits to control the stone cell content. In addition, the regulatory function of AC elements upstream of genes related to lignin synthesis has been shown (Zhao and Dixon, 2011). *AtCCOMT1* in *A. thaliana* possesses this AC element and has been found to be involved in lignin synthesis (Raes et al., 2003; Do et al., 2007). Thus, *PbCCOMT1* and *PbCCOMT3* both contain AC elements and may participate in lignin synthesis in pear.

Semi-quantitative RT-PCR showed that the expression of two *CCOMT* genes (*PbCCOMT1* and *PbCCOMT3*) was associated with stone cell and lignin contents during fruit development. These results suggest that *PbCCOMT1* and *PbCCOMT3* play key roles in the development of stone cells in Dangshan Su pear. High levels of *PbCCOMT1* and *PbCCOMT3* expression in stems is consistent with the expression of *PtCCOMT* in *Populus tomentosa*. High levels of *CCOMT* expression is associated with xylem development in woody plants (Shi et al., 2010; Yang et al., 2014). At present, many studies have used genetic engineering to suppress *CCOMT* expression to downregulate lignin (Wagner et al., 2011; Li et al., 2013; Pang et al., 2014; Zhang et al., 2014). In the present study, two *CCOMT* genes (*PbCCOMT1* and *PbCCOMT3*) involved in lignin biosynthesis have been identified. Subsequent research will focus on isolating the two genes and analyzing their function in lignin biosynthesis using transgenic methods. These genes could be used as targets for antisense and interference RNA technology to reduce the lignin content and production of stone cells and therefore increasing the quality of pear fruit.

## Conflicts of interest

The authors declare no conflict of interest.

## ACKNOWLEDGMENTS

Research supported by grants from the National Natural Science Foundation of China (#31171944 and #30771483).

## REFERENCES

- Bailey TL, Williams N, Misleh C and Li WW (2006). MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res.* 34: W369-W373. <http://dx.doi.org/10.1093/nar/gkl198>
- Bjellqvist B, Hughes GJ, Pasquali C, Paquet N, et al. (1993). The focusing positions of polypeptides in immobilized

- pH gradients can be predicted from their amino acid sequences. *Electrophoresis* 14: 1023-1031. <http://dx.doi.org/10.1002/elps.11501401163>
- Cai YP, Li GQ, Nie JQ, Lin Y, et al. (2010). Study of the structure and biosynthetic pathway of lignin in stone cells of pear. *Sci. Hortic.* 125: 374-379. <http://dx.doi.org/10.1016/j.scienta.2010.04.029>
- Carocha V, Soler M, Hefer C, Cassan-Wang H, et al. (2015). Genome-wide analysis of the lignin toolbox of *Eucalyptus grandis*. *New Phytol.* 206: 1297-1313. <http://dx.doi.org/10.1111/nph.13313>
- Choi JH and Lee SH (2013). Distribution of stone cell in Asian, Chinese, and European pear fruit and its morphological changes. *J. Appl. Bot. Food Qual.* 86: 185-189.
- Do CT, Pollet B, Thévenin J, Sibout R, et al. (2007). Both caffeoyl Coenzyme A 3-O-methyltransferase 1 and caffeic acid O-methyltransferase 1 are involved in redundant functions for lignin, flavonoids and sinapoyl malate biosynthesis in *Arabidopsis*. *Planta* 226: 1117-1129. <http://dx.doi.org/10.1007/s00425-007-0558-3>
- Du Q, Gong C, Pan W and Zhang D (2013). Development and application of microsatellites in candidate genes related to wood properties in the Chinese white poplar (*Populus tomentosa* Carr.). *DNA Res.* 20: 31-44. <http://dx.doi.org/10.1093/dnares/dss031>
- Gharechahi J, Khalili M, Hasanloo T and Salekdeh GH (2013). An integrated proteomic approach to decipher the effect of methyl jasmonate elicitation on the proteome of *Silybum marianum* L. hairy roots. *Plant Physiol. Biochem.* 70: 115-122. <http://dx.doi.org/10.1016/j.plaphy.2013.05.031>
- Goujon T, Sibout R, Pollet B, Maba B, et al. (2003). A new *Arabidopsis thaliana* mutant deficient in the expression of O-methyltransferase impacts lignins and sinapoyl esters. *Plant Mol. Biol.* 51: 973-989. <http://dx.doi.org/10.1023/A:1023022825098>
- Hirano K, Aya K, Kondo M, Okuno A, et al. (2012). *OsCAD2* is the major *CAD* gene responsible for monolignol biosynthesis in rice culm. *Plant Cell Rep.* 31: 91-101. <http://dx.doi.org/10.1007/s00299-011-1142-7>
- Kim BG, Sung SH, Chong Y, Lim Y, et al. (2010). Plant flavonoid O-methyltransferases: substrate specificity and application. *J. Plant Biol.* 53: 321-329. <http://dx.doi.org/10.1007/s12374-010-9126-7>
- Kong X, Lv W, Jiang S, Zhang D, et al. (2013). Genome-wide identification and expression analysis of calcium-dependent protein kinase in maize. *BMC Genomics* 14: 433. <http://dx.doi.org/10.1186/1471-2164-14-433>
- Li X, Chen W, Zhao Y, Xiang Y, et al. (2013). Downregulation of caffeoyl-CoA O-methyltransferase (*CCoAOMT*) by RNA interference leads to reduced lignin production in maize straw. *Genet. Mol. Biol.* 36: 540-546. <http://dx.doi.org/10.1590/S1415-47572013005000039>
- Liu JY, Osbourn A and Ma PD (2015). MYB transcription factors as regulators of phenylpropanoid metabolism in plants. *Mol. Plant* 8: 689-708. <http://dx.doi.org/10.1016/j.molp.2015.03.012>
- Lu XP, Liu YZ, An JC, Hu HJ, et al. (2011). Isolation of a cinnamoyl CoA reductase gene involved in formation of stone cells in pear (*Pyrus pyrifolia*). *Acta Physiol. Plant.* 33: 585-591. <http://dx.doi.org/10.1007/s11738-010-0583-x>
- Magali L, Patrice D, Gert T, Kathleen M, et al. (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. <http://dx.doi.org/10.1093/nar/30.1.325>
- Jin Q, Yan CC, Qiu JX, Zhang N, et al. (2013). Structural characterization and deposition of stone cell lignin in Dangshan Su pear. *Sci. Hortic.* 155: 123-130. <http://dx.doi.org/10.1016/j.scienta.2013.03.020>
- Joshi CP and Chiang VL (1998). Conserved sequence motifs in plant S-adenosyl-L-methionine-dependent methyltransferases. *Plant Mol. Biol.* 37: 663-674. <http://dx.doi.org/10.1023/A:1006035210889>
- Osakabe K, Tsao CC, Li L, Popko JL, et al. (1999). Coniferyl aldehyde 5-hydroxylation and methylation direct syringyl lignin biosynthesis in angiosperms. *Proc. Natl. Acad. Sci. U. S. A.* 96: 8955-8960. <http://dx.doi.org/10.1073/pnas.96.16.8955>
- Pang SL, Ong SS, Lee HH, Zamri Z, et al. (2014). Isolation and characterization of *CCoAOMT* in interspecific hybrid of *Acacia auriculiformis* x *Acacia mangium* - a key gene in lignin biosynthesis. *Genet. Mol. Res.* 13: 7217-7238. <http://dx.doi.org/10.4238/2014.September.5.7>
- Raes J, Rohde A, Christensen JH, Van de Peer Y, et al. (2003). Genome-wide characterization of the lignification toolbox in *Arabidopsis*. *Plant Physiol.* 133: 1051-1071. <http://dx.doi.org/10.1104/pp.103.026484>
- Robert X and Gouet P (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* 42: 320-324. <http://dx.doi.org/10.1093/nar/gku316>
- Rozas J and Rozas R (1995). DnaSP, DNA sequence polymorphism: an interactive program for estimating Population Genetics parameters from DNA sequence data. *Comput. Applic. Biosci.* 11: 621-625.
- Shi R, Sun YH, Li QZ, Heber S, et al. (2010). Towards a systems approach for lignin biosynthesis in *Populus trichocarpa*: transcript abundance and specificity of the monolignol biosynthetic genes. *Plant Cell Physiol.* 51: 144-163. <http://dx.doi.org/10.1093/pcp/pcp175>



- Sibout R, Eudes A, Mouille G, Pollet B, et al. (2005). Cinnamyl alcohol dehydrogenase-C and -D are the primary genes involved in lignin biosynthesis in the floral stem of *Arabidopsis*. *Plant Cell* 17: 2059-2076. <http://dx.doi.org/10.1105/tpc.105.030767>
- Stothard P (2000). The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques* 28: 1102-1104.
- Sung SH, Kim BG, Chong Y and Ahn JH (2011). Characterization of phenylpropanoid O-methyltransferase from rice: molecular basis for the different reactivity toward different Substrates. *J. Plant Biol.* 54: 314-320. <http://dx.doi.org/10.1007/s12374-011-9169-4>
- Tamura K, Peterson D, Peterson N, Stecher G, et al. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28: 2731-2739. <http://dx.doi.org/10.1093/molbev/msr121>
- van Parijs FRD, Ruttink T, Boerjan W, Haesaert G, et al. (2015). Clade classification of monolignol biosynthesis gene family members reveals target genes to decrease lignin in *Lolium perenne*. *Plant Biol.* 17: 877-892. <http://dx.doi.org/10.1111/plb.12316>
- Wagner A, Tobimatsu Y, Phillips L, Flint H, et al. (2011). *CCoAOMT* suppression modifies lignin composition in *Pinus radiata*. *Plant J.* 67: 119-129. <http://dx.doi.org/10.1111/j.1365-3113X.2011.04580.x>
- Wang B, Zhang N, Yan CC, Jin Q, et al. (2013). Bagging for the development of stone cell and metabolism of lignin in *Pyrus bretschneideri* Dangshan Suli. *Acta Hortic. Sin.* 40: 531-539.
- Wei H, Li W, Sun X, Zhu S, et al. (2013). Systematic analysis and comparison of nucleotide-binding site disease resistance genes in a diploid cotton *Gossypium raimondii*. *PLoS One* 8: e68435. <http://dx.doi.org/10.1371/journal.pone.0068435>
- Wils CR, Brandt W, Manke K and Vogt T (2013). A single amino acid determines position specificity of an *Arabidopsis thaliana* CCoAOMT-like O-methyltransferase. *FEBS Lett.* 587: 683-689. <http://dx.doi.org/10.1016/j.febslet.2013.01.040>
- Wu J, Wang ZW, Shi ZB, Zhang S, et al. (2012). The genome of the pear (*Pyrus bretschneideri* Rehd.). *Genome Res.* 23: 396-408. <http://dx.doi.org/10.1101/gr.144311.112>
- Yan CC, Yin M, Zhang N, Jin Q, et al. (2014). Stone cell distribution and lignin structure in various pear varieties. *Sci. Hortic.* 174: 142-150. <http://dx.doi.org/10.1016/j.scienta.2014.05.018>
- Yang XH, Li XG, Li BL and Zhang DQ (2014). Genome-wide transcriptional profiling reveals molecular signatures of secondary xylem differentiation in *Populus tomentosa*. *Genet. Mol. Res.* 13: 9489-9504. <http://dx.doi.org/10.4238/2014.November.11.14>
- Zhang G, Zhang YJ, Xu JT, Niu XP, et al. (2014). The *CCoAOMT1* gene from jute (*Corchorus capsularis* L.) is involved in lignin biosynthesis in *Arabidopsis thaliana*. *Gene* 546: 398-402. <http://dx.doi.org/10.1016/j.gene.2014.05.011>
- Zhang Y, Lv HP, Ma CY, Guo L, et al. (2015). Cloning of a caffeoyl-Coenzyme A O-methyltransferase from *Camellia sinensis* and analysis of its catalytic activity. *J. Zhejiang Univ. Sci. B.* 16: 103-112. <http://dx.doi.org/10.1631/jzus.B1400193>
- Zhao Q and Dixon RA (2011). Transcriptional networks for lignin biosynthesis: more complex than we thought? *Trends Plant Sci.* 16: 227-233. <http://dx.doi.org/10.1016/j.tplants.2010.12.005>
- Zhao SG, Zhang JG, Zhao YP, Zhang YX (2013). New discoveries of stone cell differentiation in fruitlets of 'Yali' pears (*Pyrus bretschneideri* Rehd.). *J. Food Agric. Environ.* 11: 937-942.

## Supplementary material

**Table S1.** Sequences used for phylogenetic analyses of *OMT* proteins.

**Table S2.** Primers used for semi-quantitative RT-PCR of *PbCCOMT* homologs. *Tubulin* (accession No. AB239680.1) served as the internal standard.

**Figure S1.** Detailed amino acid composition of the 12 motifs in *PbOMT* proteins.

**Figure S2.** Functional interaction networks of OMTs in pear according to orthologs in *Arabidopsis*.