



# Sorghum CCoAOMT and CCoAOMT-like gene evolution, structure, expression and the role of conserved amino acids in protein activity

M. Rakoczy<sup>1</sup> · I. Femiak<sup>1</sup> · M. Alekska<sup>1</sup> · M. Figlerowicz<sup>1</sup> · J. Podkowinski<sup>1</sup>

Received: 12 October 2017 / Accepted: 24 April 2018 / Published online: 2 May 2018  
© The Author(s) 2018

## Abstract

Sorghum is a crop plant that is grown for seeds, sucrose, forage and biofuel production. In all these applications, lignin is a superfluous component that decreases the efficiency of technological processes. Caffeoyl-coenzyme A *O*-methyltransferase (CCoAOMT) is an enzyme involved in monolignol synthesis that affects the efficiency of lignification and lignin composition. The sorghum genome harbors one CCoAOMT gene and six closely related CCoAOMT-like genes. The structures of four sorghum CCoAOMT-like enzymes suggest that these proteins might methylate caffeoyl coenzyme A and contribute to monolignol synthesis. In this study, two sorghum genes, CCoAOMT and one CCoAOMT-like, were found to be highly expressed in leaves, stems and immature seeds. The promoters of these genes possess clusters of transcription factor-binding sites specific for lignification, and this suggests that they are important for lignification. Phylogenetic analysis revealed that one sorghum CCoAOMT-like enzyme is closely related to ancestral cyanobacterial CCoAOMT-like proteins. The remaining CCoAOMT-like enzymes, including the one highly expressed in the leaves and stem, are closely related to CCoAOMT. Genes from these two groups possess different, evolutionarily conserved gene structures. The structure of the sorghum CCoAOMT-like protein from the ancestral clade was modeled and differences between enzymes from the two clades were analyzed. These results facilitate a better understanding of the evolution of genes involved in lignification, and provide valuable data for sorghum improvement through traditional breeding or molecular genetic techniques. The findings suggest that CCoAOMT-like genes might be recruited in lignification and raise questions of the frequency of such functional shifts.

**Keywords** Sorghum (*Sorghum bicolor* L.) · CCoAOMT · Gene expression · Phylogeny · Protein structure

## Introduction

Sorghum (*Sorghum bicolor* L.,  $2n=20$ ), which belongs to the Poaceae family, is a C4 crop plant that is cultivated for seeds, sugar and green biomass used as forage (Carpita and McCann 2008; Mohammed et al. 2015). Recently, sorghum was used for biofuel production, and it is a model bioenergy plant (Calvino and Messing 2012; Mullet et al. 2014; Vermerris 2011). In all these applications, lignin—the second

most abundant component of the plant cell wall—positively affects plant growth and biomass production, but it decreases the efficiency of technological processes. Attempts to improve the quality of plant material drew attention to caffeoyl-coenzyme A *O*-methyltransferase (CCoAOMT), one of the enzymes involved in lignin monomer synthesis (Anterola and Lewis 2002; Bottcher et al. 2013; Goujon et al. 2003). The sorghum genome harbors one CCoAOMT gene and six closely related genes coding for CCoAOMT-like enzymes (Walker et al. 2016; Xu et al. 2009). To investigate the impact of these genes on lignification, it is necessary to explore their expression pattern, association with biological processes and phylogenetic relationships.

CCoAOMTs and CCoAOMT-like enzymes form a family within *O*-methyltransferases characterized by molecular weight ranging between 26 and 29 kDa and divalent metal ions in the active site (Ferrer et al. 2005; Joshi and Chiang 1998; Lee et al. 2008; Liu et al. 2016; Walker et al. 2016; Widiez et al. 2011). Despite the similarities, CCoAOMTs

Communicated by S. Hohmann.

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s00438-018-1441-6>) contains supplementary material, which is available to authorized users.

✉ J. Podkowinski  
jpodkow@man.poznan.pl

<sup>1</sup> Institute of Bioorganic Chemistry PAS, ul. Noskowskiego 12/14, 61-704 Poznan, Poland

are substrate specific and use two methyl acceptors, caffeoyl-coenzyme A or 5-hydroxyferuloyl-coenzyme A, but CCoAOMT-like enzymes use a variety of compounds, including flavonoids, anthocyanins, phenylpropanoids or alkaloids (Ibdah et al. 2003; Inoue et al. 1998; Widiez et al. 2011). The enzymes from these two groups are usually associated with different biological processes. CCoAOMTs provide precursors for coniferyl and sinapyl alcohols, which are subsequently used in lignin synthesis (Barros et al. 2015; Dixon et al. 2001; Raes et al. 2003; Vanholme et al. 2010), and CCoAOMT-like enzymes, which process a broad spectrum of substrates, are involved in a variety of secondary metabolism pathways that are often specific for a given plant. The role of CCoAOMT in lignin synthesis was first recognized in *Z. violacea* and *N. tabacum*, where the enzyme was identified in all lignified tissues (Maury et al. 1999; Ye et al. 1994; Zhong et al. 1998). The direct evidence of CCoAOMT's role in lignin synthesis came from transgenic plants with increased or decreased CCoAOMT gene expression (Guo et al. 2001; Ye et al. 2001). Different approaches used to downregulate CCoAOMT, including antisense RNA, RNA interference, or gene knockout, resulted in decreased lignin content in plants, which was associated with collapsed vascular bundles, lower height, lower biomass and developmental retardation (Do et al. 2007). Plants with downregulated CCoAOMT genes also exhibit modified lignin subunit composition due to disturbances in syringyl and guaiacyl monolignol synthesis (Li et al. 2013; Meyermans et al. 2000; Pang et al. 2014; Pinçon et al. 2001; Toraman et al. 2016; Zhong et al. 1998). CCoAOMT overexpression produces the opposite effects, including increased lignin content, plant height and silique length (Zhang et al. 2014).

Recent results suggest that the role of substrate-specific CCoAOMTs is not limited to lignin synthesis and secondary cell wall formation. The CCoAOMT from petunia was shown to participate in the phenylpropene pathway leading to the synthesis of volatile attractants in petals (Shaipulah et al. 2016). The *Vitis vinifera* substrate-specific CCoAOMT is a multifunctional *O*-methyltransferase that may contribute to anthocyanin methylation (Giordano et al. 2016). Maize ZmCCoAOMT2, aside from its role in lignin synthesis as an enzyme, was shown to confer quantitative resistance to fungal pathogens (Wang and Balint-Kurti 2016; Yang et al. 2017). The process depends on ZmCCoAOMT2 interactions with nucleotide-binding leucine-rich-repeat proteins and suppression of a hypersensitive response.

Several lines of evidence demonstrated that CCoAOMT-like enzymes might contribute to monolignol synthesis. Most CCoAOMT-like enzymes are usually associated with species-specific processes (Ibdah et al. 2003; Kopycki et al. 2008a; Widiez et al. 2011), but the sorghum enzymes from this group were suggested to methylate the same substrate as substrate-specific CCoAOMTs (Walker et al. 2016). The crystal

structure comparison of the substrate-specific CCoAOMTs from the dicot plant *Medicago sativa* and the monocot plant *S. bicolor* with the promiscuous CCoAOMT-like enzyme from *M. crystallinum* revealed that differences in substrate specificity result from small topological changes in the N-terminal amino acids and a C-terminal variable loop (Ferrer et al. 2005; Hoffmann et al. 2001; Kopycki et al. 2008a; Walker et al. 2016).

The evolutionary relationships between CCoAOMT and CCoAOMT-like enzymes are more complex than classification based on substrate specificity. CCoAOMT-like enzymes are divided into two clades: one grouping CCoAOMT-like enzymes closely related to ancestral, cyanobacterial CCoAOMT-like and another, which comprises close relatives of substrate-specific CCoAOMTs (Ibdah et al. 2003; Ma and Luo 2015; Sahr et al. 2010; Widiez et al. 2011). All sorghum CCoAOMT-like enzymes for which the previous structural analysis suggests CCoAOMT activity are closely related to substrate-specific CCoAOMTs (Walker et al. 2016). However, the importance of these enzymes for lignification processes depends also on efficient expression in proper tissues.

In this study, the expression of sorghum CCoAOMT and CCoAOMT-like genes was examined in highly lignified organs: mature leaves, stem and immature seeds. Two genes were found to be highly expressed in these organs: substrate-specific CCoAOMT and one of the CCoAOMT-like genes. The promoters of these genes possess clusters of transcription factor-binding sites that are specific for lignification, suggesting a role in lignification processes. The phylogenetic analysis of sorghum enzymes and well-characterized homologous proteins from plants and cyanobacteria revealed early dedifferentiation of CCoAOMTs and CCoAOMT-like enzymes, from CCoAOMT-like enzymes related to ancestral cyanobacterial enzymes. These two clades differ in terms of exon–intron organization of their genes. The amino acids essential for CCoAOMT and CCoAOMT-like enzymes were identified using protein structural analysis, homology modeling and conservation analysis. These data suggest that sorghum CCoAOMT-like genes might be recruited to monolignol synthesis and lignification processes and shed new insights into the origin of substrate-specific CCoAOMTs. The results concerning CCoAOMT and CCoAOMT-like genes are associated with long-distance transport in plants, plant response to pathogens, mechanical resistance, and the quality of plant material used as food or forage, and may be applied to improve sorghum through traditional or molecular breeding.

## Materials and methods

### Plant materials

*Sorghum bicolor* cv. RONA1 plants grown in the field for 4 months (110 days) were in the growth stage III, between flowering and grain formation, when the plant material was collected (Gerik et al. 2003). The plant samples, consisting of mature leaves, stems and immature seeds from up to five plants, were pooled and flash frozen in liquid nitrogen a few minutes after collecting.

### RNA isolation

Total RNA was isolated from mature leaves, the inner section of the stem, or immature seeds, frozen in liquid nitrogen and ground into powder in a mortar in the presence of liquid nitrogen. RNA was isolated with RNeasy Plant Kit (Qiagen) or extracted with 0.1 M NaCl, 2% SDS, 50 mM Tris/HCl pH 8.0, 10 mM EDTA buffer in the presence of phenol–chloroform–isoamyl alcohol and precipitated two times with 2 M LiCl, resuspended in RNase-free water and precipitated with 2.5 volumes of ethanol. Finally, RNA was resuspended in RNase-free water. RNA concentration and quality was evaluated using UV absorbance at 260 nm followed by electrophoresis on an agarose gel under denaturing conditions (Sambrook et al. 1989) or RNA 6000 Nano Chip (Agilent). The reverse transcription (RT) reaction was carried out using Maxima First Strand cDNA Synthesis Kit for RT-qPCR (Thermo Scientific) with 100–500 ng/μL of total RNA as a template.

### Digital droplet PCR (ddPCR)

Reverse transcription (RT) was carried out using Maxima First Strand cDNA Synthesis Kit for RT-qPCR (Thermo Scientific) with 100–500 ng/μL of total RNA as a template. A single ddPCR reaction was composed of 1.0 μL of cDNA samples (RT reaction) diluted 1:5–1:50 in water, 200 nM primers (Table 1), QX200 ddPCR EvaGreen Supermix (Bio-Rad) and MilliQ water to a total reaction volume of 21 μL. The reaction was partitioned into droplets by the QX200 Droplet Generator (Bio-Rad) and amplified in a C1000 Touch Thermal Cycler (Bio-Rad). The amplification conditions were composed of initial denaturation at 95 °C for 5 min, 40 cycles of 95 °C for 30 s, 58 °C for 30 s and 72 °C for 45 s, and a final extension at 72 °C for 2 min followed by signal stabilization at 4 °C for 5 min and 90 °C for 5 min. After completing the amplification, the fluorescence of droplets was registered using a QX100 droplet reader (BioRad) and data were analyzed using QuantaSoft software. The expression of *S. bicolor* genes was normalized using ubiquitin carrier protein-Sobic.009G174900 (Johnson et al. 2014).

### Bioinformatic and phylogenetic analysis

Sequences were assembled, edited and processed with BioEdit software, BLAST and The Sequence Manipulation Suite (Stothard 2000).

ChloroP 1.1 was used to predict chloroplast transit peptide in protein sequences (Emanuelsson et al. 1999) and CELLO2GO was used for protein subcellular localization (Yu et al. 2014).

**Table 1** List of primers

Application		ID	Primer sequence, from 5' to 3'
<i>S. bicolor</i> genes RT-ddPCR expression analysis	SbCCoAOMT-2	Mz47	TTCGTGGACGCGACAAGGT
		Mz48	ATGGCGGCGTTGAACCCCT
	SbCCoAOMT-6	Mz49	AAGGTGGAGTCCGAGGGCA
		Mz51	CAGCGTCGACGAACACGAAGT
		Mz52	AACCTCGGCGCCTTCGACTT
	SbCCoAOMT-5	Mz53	GCGGCAATCACGGCGTTGAA
		Mz56	GCGATCAAGGACCTCAATGT
	SbCCoAOMT-4	Mz57	TGAGTGATGAGTTCCGACGA
		Mz62	GGTGGCAGCATCCCGAATGT
	SbCCoAOMT-3	Mz63	TGTTCTCCGCTCGTGCTTG
		Mz86	GGGCTTCCTCGGTAGATTTC
	SbCCoAOMT-7	Mz87	GGCGATGAACCTCTTGACAT
		S95	TTCTTGGGAATTGTCGCCAT
	SbCCoAOMT-1	S96	ATCCACGGTAGGAGCAGTAT
Sobic.009g174900		S99	CTGACGATTACCTTCTGCTC
		S100	CCTCACTATGGATGGCAATG

The CCoAOMT amino acid sequences were aligned using ClustalW. A phylogenetic tree was constructed using MEGA 6.0 software (Tamura et al. 2013) with the maximum likelihood method based on the JTT matrix-based model (Jones et al. 1992). Bootstrap-supported consensus trees were inferred from 1000 replicates (Felsenstein 1985) and branches with less than 40% bootstrap support were collapsed.

UCSF Chimera, a visualization system for exploratory research and analysis, was used for structure conservation analysis.

Conservation mapping of the corresponding sequence alignment onto the structure of SbCCoAOMT-1 (PBD:5KVA) was performed using Chimera with the default setting. For conservation style, AL2CO was chosen with sum of pairs as the type of equation.

The 3D model of the SbCCoAOMT-7-trimmed protein was predicted to examine structural differences between the CCoAOMT-like protein from clade 2 and SbCCoAOMT-1 (PDB: 5KVA) from clade 1a. Structure modeling was performed using the I-TASSER server with default structural template selection. I-TASSER (Iterative Threading ASSEmblly Refinement)—<http://zhanglab.ccmb.med.umich.edu/I-TASSER>.

Databases:

Phytozome (<https://phytozome.jgi.doe.gov>),

GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>),

PLACE (Higo et al. 1997) (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>).

Figures and diagrams presenting genes and promoter regions structure were prepared with Illustrator for Biological Sequences (IBS) (Liu et al. 2015) and Gene Structure Display Server (GSDS) (Hu et al. 2014).

## Primers used in this study

See Table 1.

## Accession numbers of sequences used in this article

Sequence data from this article can be found in the Phytozome and GenBank/EMBL data libraries, ID of genomic sequences used for gene structure analysis, which are not linked to protein sequences are given in parenthesis: *Arabidopsis thaliana*—AT4G34050.1, AT1G24735.1, AT1G67980, AT1G67990, AT4G26220.1, AT3G62000.1, AT3G61990.1; *Chlamydomonas reinhardtii*—XP\_001693484 (gene Cre13.g571950.t1.2); *Medicago truncatula*—Medtr4g094815, Medtr4g085590, Medtr4g094368, Medtr2g070410, *Medicago sativa*—AAC28973; *Mesembryanthemum crystallinum*—AAN61072; *Nicotiana tabacum*—AAC49913, CAA91228, CAB05369; *Oryza sativa*—BAA78733 (gene LOC\_Os06g06980.1), AAT68023

(gene LOC\_Os08g38900.1), BAD08718 (gene LOC\_Os08g38910.1), BAG98414 (gene LOC\_Os09g30360.1), *Petroselinum crispum*—AAA33851; *Physcomitrella patens*—Pp3c23\_5530V3.1, Pp3c4\_20870V3.5; *Populus*—BAA19102; *Plagiochasma appendiculatum*—ALS88170; *Selaginella moellendorffii*—locus 101568, XP\_002963752 (gene NW\_003314264 REGION: 2755091..2756705), XP002976256 (gene NW\_003314291 REGION: 568356..569800); *Sorghum bicolor*—Sobic.010G052200.1 (SbCCoAOMT-1), Sobic.002G242300.1 (SbCCoAOMT-2), Sobic.007G218700.1 (SbCCoAOMT-3), Sobic.007G218800.1 (SbCCoAOMT-4), Sobic.007G218500.1 (SbCCoAOMT-5), Sobic.007G217200.1 (SbCCoAOMT-6); Sobic.007G043200.1 (SbCCoAOMT-7); Sobic.007G047300.1 (COMT, caffeic acid *O*-methyltransferase), Sobic.009G174900 (ubiquitin carrier protein); *Synechocystis* sp.—WP\_010873795.1; *Vanilla planifolia*—ADZ76153, ADZ76154; *Zea mays*—CAB45149 (gene GRMZM2G127948), AFW65160 (gene GRMZM2G332522\_T02), DAA40360 (gene GRMZM2G004138\_T01), XP\_008662866 (gene GRMZM2G077486\_T01).

## Results

### *S. bicolor* genome harbors seven gene coding for CCoAOMT and CCoAOMT-like proteins

The *Sorghum bicolor* genome was queried with sequences of well-characterized caffeoyl-coenzyme A *O*-methyltransferases (CCoAOMT) and CCoAOMT-like proteins. This yielded seven genes: Sobic.010G052200, Sobic.002G242300.1, Sobic.007G218700, Sobic.007G218800.1, Sobic.007G218500, Sobic.007G217200 and Sobic.007G043200, Table 2. In this study, the genes and their peptides were renamed SbCCoAOMT-1, SbCCoAOMT-2, SbCCoAOMT-3, SbCCoAOMT-4, SbCCoAOMT-5, SbCCoAOMT-6 and SbCCoAOMT-7, Table 2. Screening the sorghum genome and sequence databases with these sequences identified genes or cDNAs coding for the same seven proteins or more distantly related sequences, such as Sb07g00386, of higher similarity to COMT than CCoAOMT or CCoAOMT-like proteins.

The evolutionary relationships among the sorghum CCoAOMT and CCoAOMT-like proteins and their homologs from other plants were ascertained using Mega 6, Fig. 1. The sorghum CCoAOMT and CCoAOMT-like genes were categorized into three clades that were named according to Widiez et al. (2011): clade 1a—true CCoAOMTs; clade 1c—grass CCoAOMT-like proteins; and clade 2—CCoAOMT-like proteins closely related



**Table 2** *S. bicolor* CCoAOMT and CCoAOMT-like protein names used in this study and physicochemical parameters: MW—molecular weight, pI—isoelectric point, GRAVY—the grand average of hydrophathy

Gene/protein name in this study	Locus id in phytozome	Protein length	MW (kDa)	pI (pH)	GRAVY [-2 .. +2]	Predicted subcellular localization	Clade
SbCCoAOMT-1	Sobic.010G052200.1	261 aa	29.07	5.05	-0.261	Cytoplasm	1a
SbCCoAOMT-2	Sobic.002G242300.1	241 aa	25.73	5.08	0.139	Chloroplast	1c
SbCCoAOMT-3	Sobic.007G218700.1	250 aa	27.36	5.35	-0.155	Cytoplasm, Chloroplast	1c
SbCCoAOMT-4	Sobic.007G218800.1	246 aa	27.23	4.87	-0.107	Cytoplasm	1c
SbCCoAOMT-5	Sobic.007G218500.1	267 aa	28.76	5.34	0.005	Cytoplasm, chloroplast	1c
SbCCoAOMT-6	Sobic.007G217200.1	144 aa	15.91	4.38	0.041	Cytoplasm	1c
SbCCoAOMT-7	Sobic.007G043200.1	306 aa	32.99	9.06	0.103	Chloroplast	2
SbCCoAOMT-7 trimmed at Ser-67		239 aa	26.37	7.30	0.016		2

Protein subcellular localization was analyzed with CELLO2GO. The trimmed SbCCoAOMT-7 transit peptide at Ser-67 is based on ChloroP analysis

to cyanobacterial enzymes, Fig. 1. SbCCoAOMT-1 is the only sorghum protein in clade 1a, the grouping of CCoAOMT enzymes with high substrate specificity. This clade harbors CCoAOMTs associated with lignin synthesis, including *A. thaliana* ATG434050, *N. tabacum* CAA91228, *Petroselinum crispum* AAA33851, and *M. sativa* AAC28973 (Do et al. 2007; Ferrer et al. 2005; Martz et al. 1998; Pakusch et al. 1989). Five closely related sorghum CCoAOMT-like genes: SbCCoAOMT-2, SbCCoAOMT-3, SbCCoAOMT-4, SbCCoAOMT-5, and SbCCoAOMT-6 are grouped in clade 1c specific for grass CCoAOMT-like enzymes promiscuous with respect to the methyl acceptor. The SbCCoAOMT-7 gene is the only sorghum gene in clade 2, which comprises CCoAOMT-like proteins closely related to cyanobacterial enzymes. Similar to other plant proteins from this clade, the SbCCoAOMT-7 protein has an N-terminal extension of 67 amino acids.

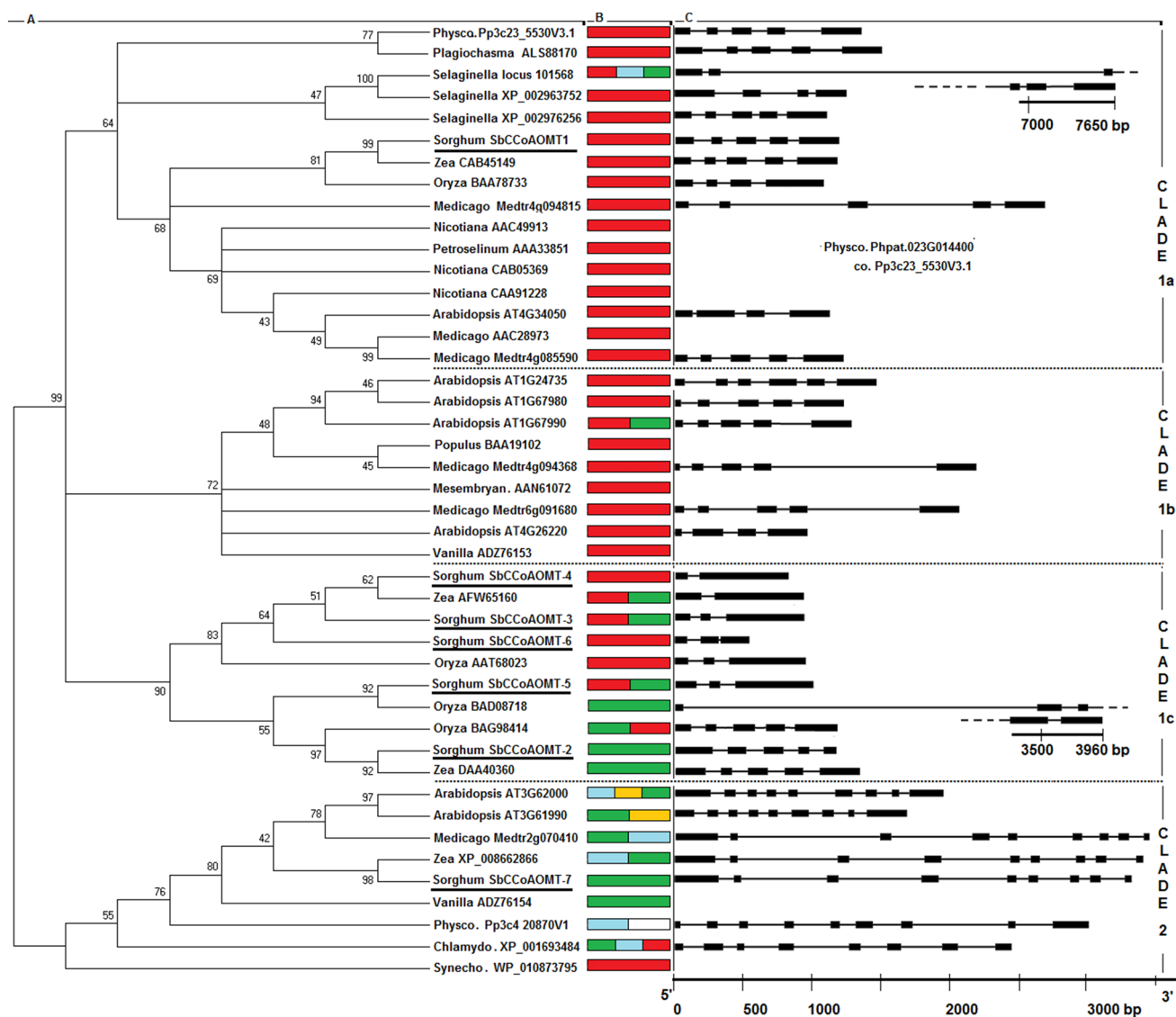
The analysis of CCoAOMT and CCoAOMT-like gene structures supports the phylogenetic relationships inferred from the protein phylogenies. The genes exhibit two different exon–intron patterns within the coding region. Genes composed of five exons are characteristic for superclade 1, but clade 2 genes are composed of nine exons, Fig. 1. Exon boundaries are conserved within clades and differ between superclade 1 and clade 2 even when the exon–intron junctions are localized in conserved regions, Fig. S-1. The clade-specific gene structure is masked in some genes, such as SbCCoAOMT-3, SbCCoAOMT-4 and SbCCoAOMT-5, by 3' end-terminal intron elimination.

SbCCoAOMT-6, the shortest sorghum CCoAOMT-like gene, possesses a unique exon–intron structure, with exon boundaries different from other genes, Fig. S-1. The location of the SbCCoAOMT-6 exon boundaries is associated with two indels specific for this gene: a duplication of the

motif EHL DALLA DEG in the central part of the protein and a 30 amino acid deletion in the variable loop, Fig. S-1.

### ***S. bicolor* CCoAOMT and CCoAOMT-like proteins: amino acid conservation, structure of superclade 1 and clade 2 proteins**

The sorghum CCoAOMT and CCoAOMT-like genes encode proteins 144–306 amino acid residues in length with calculated molecular weights of 15.91–32.99 kDa, Table 2. The shortest protein is SbCCoAOMT-6, which possesses an N-terminal deletion of 90 amino acids, a 34 amino acid deletion in the variable loop and a 16 amino acid insertion in the central region, Fig. S-1. These large deletions suggest that SbCCoAOMT-6 is not a full-length protein, and it is shorter than the cyanobacterial protein WP 010873795. The deletions in the insertion loop of SbCCoAOMT-2 and SbCCoAOMT-7 proteins are shorter, 8–9 amino acids in length, and the length of their insertion loops is similar to *Z. mays* (AFW 65160, XP 008662866), *O. sativa* (AAT 68023) or *A. thaliana* (AT3G62000), Fig. S-1. The longest sorghum CCoAOMT-like protein is SbCCoAOMT-7 due to its highly hydrophobic N-terminal extension. The ChloroP program assigned the first 67 N-terminal amino acids of SbCCoAOMT-7 as a chloroplast transit peptide. A similar transit peptide targeting the enzyme to plastids is present in Vp-OMT5 (ADZ76154) from *Vanilla planifolia*, and AT3G61990 from *A. thaliana* possesses a transit peptide that targets the enzyme to the endoplasmic reticulum (Sahr et al. 2010; Widiez et al. 2011). The apparent N-terminal extension is also present in other clade 1c proteins, and analysis of their subcellular localization with CELLO2GO and ChloroP suggests that SbCCoAOMT-2 is targeted to plastids, while SbCCoAOMT-3 and SbCCoAOMT-5 are



**Fig. 1** Phylogeny, subcellular localization and gene structure of plant CCoAOMT and CCoAOMT-like proteins. The analyzed proteins include the sorghum proteins (underlined) and their well-characterized homologs. A—the unrooted, maximum likelihood tree was constructed with 1000 bootstrap replications; the original tree was condensed with a 40% cut-off value. The clades are named according to Wiedez et al. (2011): clade 1a—true CCoAOMTs; clade 1b—non-grass CCoAOMT-like enzymes; clade 1c—grass CCoAOMT-like enzymes; and clade 2—CCoAOMT-like proteins closely related to

cyanobacteria. The peptide IDs or accession numbers are preceded by genus name or shortcut: Physco.—*Physcomitrella*, Chlamydo.—*Chlamydomonas*, Synecho.—*Synechocystis*; B—subcellular localization identified with CELLO2GO web server: red—cytoplasmic, green—chloroplast, blue—mitochondrial, yellow—nuclear, white—extracellular; C—the gene structure within the coding region with exons marked by black boxes and introns by lines. (Color figure online)

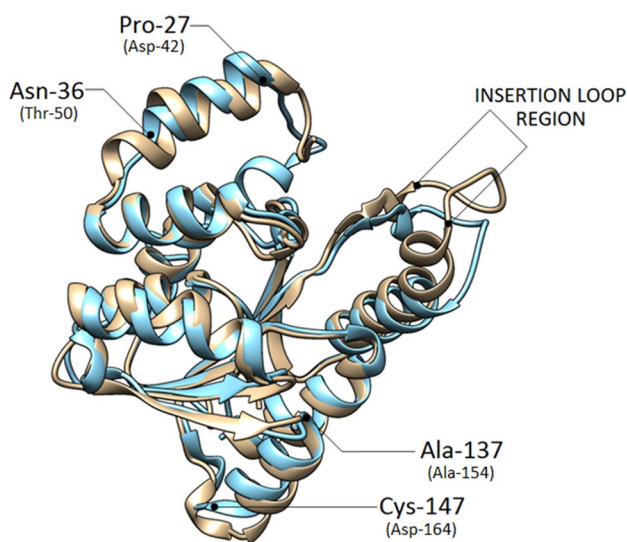
targeted to two compartments: cytosol and plastids, Fig. 1 (Yu et al. 2014). Similar to sorghum, *O. sativa* and *Z. mays* clade 1c proteins also possess transit peptides targeting them to plastids, Fig. 1b.

Forty-five highly conserved amino acids preserved across all clades in more than 95% of the analyzed sequences were identified in the set of 43 CCoAOMT and CCoAOMT-like proteins used for the phylogenetic study (SbCCoAOMT-6 was excluded from the conservancy analysis), Fig. S-1 and

CCoAOMT-CCoAOMT-like alignment in supplementary files. These amino acids are conserved in the sorghum CCoAOMT and CCoAOMT-like enzymes except SbCCoAOMT-6, which is missing six conserved amino acids due to two large deletions.

To investigate structural differences between the substrate-specific enzyme SbCCoAOMT-1 from superclade 1 and CCoAOMT-like protein from clade 2, we compared the crystallographic structure of SbCCoAOMT-1 (PDB

ID: 5KVA) with the model of the SbCCoAOMT-7 protein. For structure modeling, the SbCCoAOMT-7 protein was trimmed at Ser-67 to cut off the chloroplast transit peptide absent in the mature enzyme. The structure of the trimmed peptide was predicted using the I-TASSER server, which employs PDB datasets in the structure-threading method. The template structures applied for structure modeling involved two bacterial *O*-methyltransferases structures, PDB ID: 3R3H and 2HNKA, and human catechol-*O*-methyltransferase structure PDB ID: 2AVD, all with normalized Z score of above 1 (2.93, 3.97 and 2.81, respectively). Subsequently, the resulting model of the SbCCoAOMT-7 protein with the highest confidence score ( $C$  score =  $-0.08$ ) and the SbCCoAOMT-1 structure (PDB ID: 5KVA) was aligned using MatchMaker in Chimera (Fig. 2). The most divergent region is Asp-195 to Asp-200 of the insertion loop in trimmed SbCCoAOMT-7 protein. The variability of this region is associated with discrimination between different methyl group acceptors. The differences between the SbCCoAOMT-1 and SbCCoAOMT-7 structures also involve the shortening of two  $\alpha$ -helices: region Pro-27 to Asn-36 of trimmed SbCCoAOMT-7 (Asp-42 to Thr-50 of SbCCoAOMT-1) and region Ala-137 to Cys-147 of trimmed SbCCoAOMT-7 (Ala-154 to Asp-164 of SbCCoAOMT-1). The first of the  $\alpha$ -helices corresponds to the region of reference structure 5KVA involved in dimerization. Furthermore, the other discrepancy between these two structures concerns



**Fig. 2** Superimposition of the model of SbCCoAOMT-7 trimmed protein (blue) and the structure of SbCCoAOMT-1, PDB ID: 5KVA (tan). The regions of the highest discrepancy between SbCCoAOMT-1 and SbCCoAOMT-7 are shown:  $\alpha$ -helix Pro-27 to Asn-36,  $\alpha$ -helix Ala-137 to Cys-147 and variable loop; amino acid coordinates are according to the trimmed SbCCoAOMT-7 protein. The corresponding amino acids of the SbCCoAOMT-1 protein are marked in brackets. (Color figure online)

the N-terminal  $\alpha$ -helix localized between Ala-1 and Arg-18 of the trimmed SbCCoAOMT-7, which spatial orientation is ambiguous as this region does not possess any templates in the PDB database.

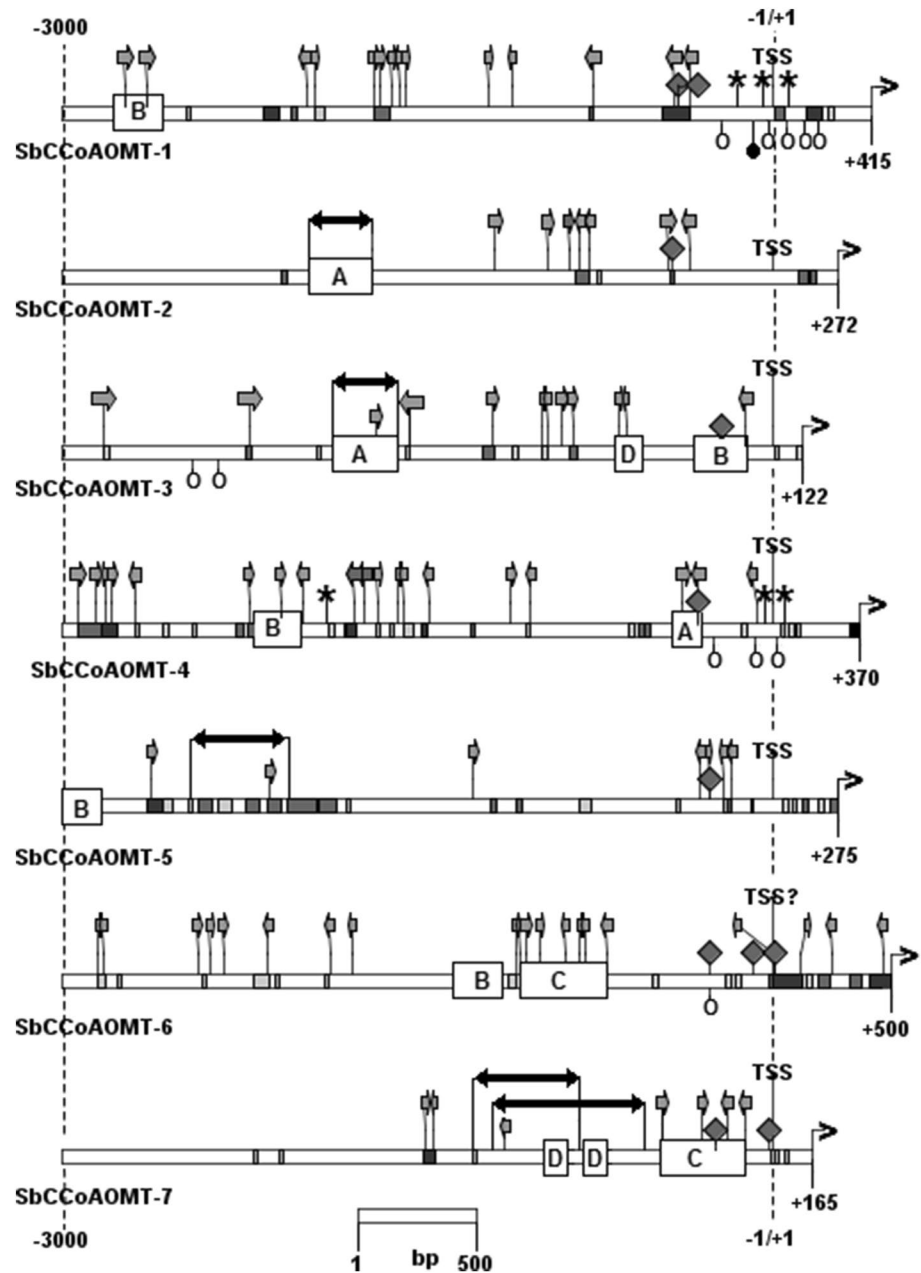
### Promoter regions of two sorghum CCoAOMT-like genes are enriched in binding sites for transcription factors involved in lignification

The promoter region of sorghum CCoAOMT and CCoAOMT-like genes spanning the 5' UTR and 3 kb above the transcription start site were scrutinized for inverted and direct repeats, conserved regions longer than 15 bp, and binding sites for transcription factors involved in lignification, Fig. 3. The distribution of these elements revealed a high variance in the architecture of the upstream regions of sorghum CCoAOMT and CCoAOMT-like genes. The sequence conservation is weak even within the vicinity of the transcription start sites and 5' UTRs. Four classes of motifs conserved in at least two of the promoters with 97–74% identity and of 38–361 bp in length were identified and named motifs A, B, C and D, Table S-1. The highest conservation was observed for the 361 bp long motif C present in the SbCCoAOMT-6 and SbCCoAOMT-7 promoter regions. The motifs are not limited to CCoAOMT and CCoAOMT-like promoters, and they are scattered throughout the sorghum genome.

The clusters of lignification-associated transcription factor binding sites were identified in the promoters of two genes: SbCCoAOMT-1 (six binding sites) and SbCCoAOMT-4 (four binding sites), Fig. 3. The clusters are composed of two motifs: (1) MybPlant motif (Place s000167), related to the P-box motif in promoters of phenylpropanoid biosynthesis genes that also regulate lignin biosynthesis (Tamagnone et al. 1998), and (2) ACIIPVPAL2 motif required for vascular-specific expression of phenylalanine ammonia lyase 2 in bean and involved in xylem-localized regulation of genes for lignin biosynthesis enzymes (Patzlaff et al. 2003). The clusters of these motifs are present exclusively in SbCCoAOMT-1 and SbCCoAOMT-4 genes where they occupy similar localization:  $-274$  to  $+248$  bp in CCoAOMT-1 and  $-247$  to  $+14$  bp in CCoAOMT-4.

Searching sorghum CCoAOMT and CCoAOMT-like promoter regions for AC regulatory elements associated with lignification (Raes et al. 2003; Weng et al. 2008) identified six highly similar or identical motifs. Five of these motifs are near the transcription start sites of the SbCCoAOMT-1 and SbCCoAOMT-4 genes within the clusters of the MybPlant and ACIIPVPAL2 transcription factor-binding sites, Fig. 3.

**Fig. 3** The structure of *S. bicolor* CCoAOMT and CCoAOMT-like promoter regions. The regions comprising 5' UTR and 3 kb above the transcription start site were screened for inverted and direct repeats—marked with black and gray arrows; short conserved regions—marked as small, gray rectangles; conserved regions longer than 30 bp—marked with big rectangles and labeled A to D; binding sites for transcription factor involved in lignification processes—open circles (Myb-Plant motif) and black circles (ACIIPVPAL2); motifs similar to AC elements—asterisks; TATA boxes—squares. Transcription start site is marked as TSS and start codon is marked with a black arrow



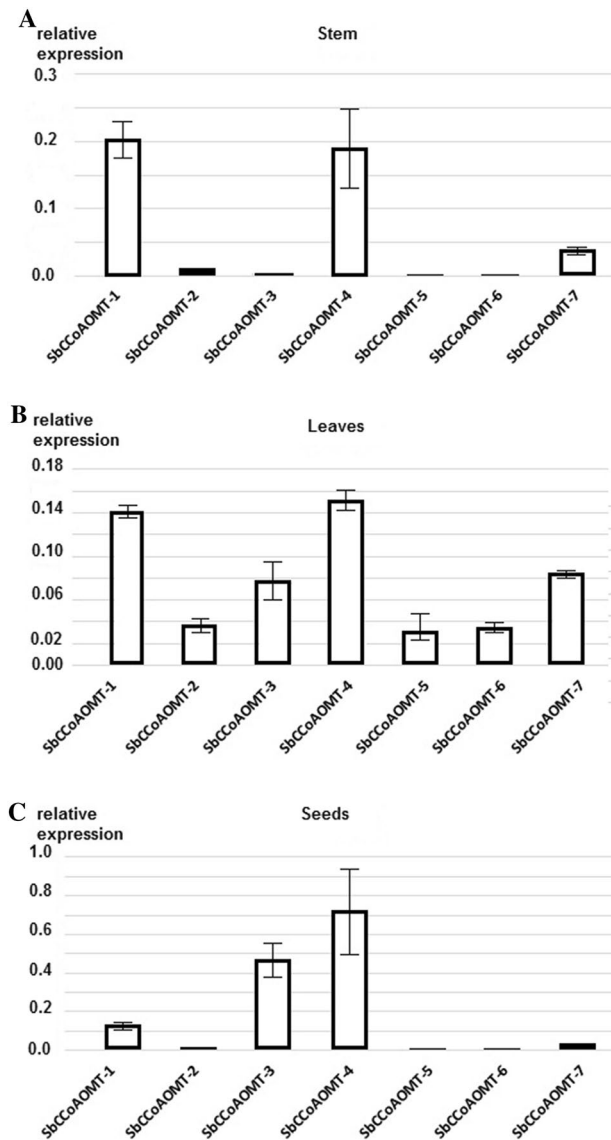
### Expression of *S. bicolor* CCoAOMT and CCoAOMT-like genes in field-grown plants

The expression of the sorghum CCoAOMT and CCoAOMT-like genes in the inner section of the stem, mature leaves and immature seeds of field-grown plants was analyzed by RT and digital PCR, Fig. 4. The Sb09g023560 gene encoding a ubiquitin-conjugating enzyme expressed at a constant and high level was used for data normalization between different samples (Johnson et al. 2014).

Two genes, SbCCoAOMT-1 and SbCCoAOMT-4, were expressed at similar, high levels in the stems of mature

sorghum plants. The transcripts of these genes were approximately five times less abundant than transcripts of the reference gene Sb09g023560. The expression of the other sorghum CCoAOMT-like genes in the stem was 5–100 times lower than that of SbCCoAOMT-1 and SbCCoAOMT-4. In leaves, the expression of SbCCoAOMT-1 and SbCCoAOMT-4 was still dominant but lower than in the stem, and the expression of the other genes was 3–5 times lower than that of SbCCoAOMT-1 and SbCCoAOMT-4. In immature seeds, SbCCoAOMT-1, SbCCoAOMT-3 and SbCCoAOMT-4 were highly expressed. The highest expression, three times higher than in the stem, was SbCCoAOMT-4. The SbCCoAOMT-1 gene





**Fig. 4** Relative expression of *S. bicolor* CCoAOMT and CCoAOMT-like genes in **a** the inner section of the stem, **b** mature leaves, and **c** immature seeds. The transcript levels were quantified using RT-ddPCR and normalized relative to the expression of the Sb09g023560 gene, which encodes a ubiquitin-conjugating enzyme; values are the mean of two independent RT-ddPCR experiments  $\pm$  SD

was expressed in seeds at a similar level as in the stem, and the expression of SbCCoAOMT-3 was between SbCCoAOMT-4 and SbCCoAOMT-1.

## Discussion

The sorghum genome contains seven CCoAOMT and CCoAOMT-like genes. One of them, SbCCoAOMT-1, belongs to clade 1a comprising substrate-specific CCoAOMTs (Paterson et al. 2009; Walker et al. 2016).

The high expression of SbCCoAOMT-1 gene in stems and leaves, organs of high lignification, is characteristic for this clade, Fig. 4. In this study, SbCCoAOMT-4, one of clade 1c genes, showed high expression similar to SbCCoAOMT-1 in leaves and stems. The promoter regions of both these genes possess clusters of transcription factor-binding sites and AC elements associated with lignification, Fig. 3. These findings were supported by localization of these two proteins in the cytoplasm, and their ability to catalyze caffeoyl-coenzyme A methylation to feruloyl-coenzyme A (Walker et al. 2016), which strongly suggest that SbCCoAOMT-1 and SbCCoAOMT-4 mainly function in monolignol synthesis associated with lignification, Fig. 1 and Table 2.

Apart from the highly expressed SbCCoAOMT-1 and SbCCoAOMT-4 genes, the SbCCoAOMT-2, SbCCoAOMT-3, SbCCoAOMT-5, SbCCoAOMT-6 and SbCCoAOMT-7 genes had activity in mature leaves, but it was 2–7 times lower. High diversity of CCoAOMT-like gene expression in leaves might be associated with the exposure of these organs to a variety of stresses, such as mechanical stress, damage, biological pathogens, high temperature, or intensive light irradiation. In seeds, the expression of the SbCCoAOMT-4 and SbCCoAOMT-3 genes exceeded that of SbCCoAOMT-1, which was expressed at a similar level as in leaves and stem, Fig. 4. The role of these genes in seeds is probably associated with forming protective structures of lignin. The SbCCoAOMT-3 protein, which possesses N-terminal extension and is probably targeted to cytosol and plastids, might be involved in synthesizing some protectants.

The SbCCoAOMT-6 gene encodes the shortest of 44 analyzed proteins, including the cyanobacterial protein. A previous report suggested that the SbCCoAOMT-6 protein is incomplete and lost enzymatic activity due to large deletions and insertion (Walker et al. 2016), Fig. S-1. The unusual structure of the SbCCoAOMT-6 gene and exon–intron boundaries different from the genes from superclade 1 and clade 2, are suggesting that the gene might have arisen as a result of pseudogene activation.

Forty-five highly conserved amino acids were identified in the set of 43 proteins used for the phylogenetic analysis and more than half of them are preserved in all analyzed proteins, Fig. S-1 (100% conserved amino acids are labeled with black dots). The high conservation of amino acids results from interactions preventing their substitution and may be used for protein structural analysis. The role of one-third of the highly conserved amino acids was assigned in previous studies on various plants (Brandt et al. 2015; Ferrer et al. 2005; Hoffmann et al. 2001; Kopycki et al. 2008a, b; Walker et al. 2016). The highly conserved amino acids correspond to SbCCoAOMT-1 Gly-101, Thr-104, Ser-107, Asp-177, Asp-203, and Asn-204, which are involved in SAM or divalent metal ion binding, or Ala-178, Lys-180, and Asp-252 involved in interactions with methyl group

acceptors or transmethylation products. Five other highly conserved amino acids correspond to *M. sativa* ACC28973 Glu-85 and Asp-165 involved in SAM and metal binding, *Synechocystis* sp. CCoAOMT-like Lys-3 and Trp-173, and *N. tabacum* ACC49913 Arg-220, participating in interactions with the methyl donor or acceptor (SbCCoAOMT-1 corresponding amino acids, respectively: Glu-99, Asp-179, Lys-35, Trp-207, Arg-242).

Nonetheless, a majority of the highly conserved amino acids, including seventeen 100% conserved amino acids, were not previously identified as essential for CCoAOMT and CCoAOMT-like enzymes. Analysis of the SbC-CoAOMT-1 model suggests that Tyr-46, which has 100% conservation, together with a less-conserved Leu-43, located in the close vicinity of N-terminal dimerization domains, might be involved in dimer formation, Fig. S-2. The high conservation of Val-102 and Gly-105, along with Tyr-106 (with 95% conservation), which are located between amino acids directly interacting with SAM, suggests their importance for SAM positioning and orientation. The other group of 100% conserved amino acids includes Phe-175, Leu-193 and Gly-198 together with Gly-197 (with 95% conservation) and seems to be essential for the positioning and orientation of amino acids interacting with SAM or the methyl group acceptor, Fig. S-2.

The phylogenetic analysis suggests that the dedifferentiation of clade 1a from clade 2 is associated with early stages of land plant evolution. As a result, the sorghum proteins from these clades have been separated for approximately 430 MYA (Labeeuw et al. 2015; Renault et al. 2017; Weng and Chapple 2010). Despite such long independent evolution, the structures of SbCCoAOMT-1 and SbCCoAOMT-7 proteins appear to be highly similar, Fig. 2. The most significant differences in the structures of these proteins depend on three regions: (1) Asp-42–Thr-50 from the dimerization region, (2) Ala-154–Asp-164, and (3) Leu-213–Lys-221 in the insertion loop (all coordinates according to SbCCoAOMT-1). The dimerization region Asp-42–Thr-50 contains two highly conserved amino acids preserved probably due to their importance for dimer formation, Figs. S-1 and S-2. The region Ala-154–D-164 contains one highly conserved amino acid that is involved in interactions with the methyl donor. The region of the highest discrepancy between SbC-CoAOMT-1 and SbCCoAOMT-7 is located between Leu-213–Lys-221, in the SbCCoAOMT-1 insertion loop, which does not contain any highly conserved amino acids. The strong structural differences in this region associated with sequence remodeling and deletions in the SbCCoAOMT-7 insertion loop, suggest that SbCCoAOMT-7 uses a different methyl acceptor than SbCCoAOMT-1, Fig. S-1. The insertion loop, which is highly variable between clades, shows high conservation within clade 1a and 2, Fig. S-2.

The CCoAOMT and CCoAOMT-like proteins phylogeny together with the specificity of gene structures within the clades, suggest that cyanobacterial genes from clade 2 are ancestors of all CCoAOMT and CCoAOMT-like genes. The presence of the genes from clades 1a and 2 in the *Physcomitrella patens* genome and the identification of clade 1a gene in liverwort *P. appendiculatum* suggest that superclade 1 arose at the very early stage of land plant evolution (Xu et al. 2015). During manuscript preparation, the *Marchantia polymorpha* genome sequence was released (Bowman et al. 2017). This allowed us to identify two liverwort CCoAOMT genes from clade 1a and one CCoAOMT-like gene from clade 2 (not shown). The *M. polymorpha* CCoAOMT-like genes are the most closely related to *P. appendiculatum*, *Ph. patents*, or *S. moellendorffii* homologs and possess an exon–intron structure characteristic for their clades that reinforce the localization of the origin of superclade 1 on the earliest stage of land plant evolution (Emiliani et al. 2009; Weng and Chapple 2010). The analysis of the *S. moellendorffii* genome identified exclusively clade 1a genes. The lack of *S. moellendorffii* clade 2 homologs might result from genes lost during the evolution of Selaginellaceae, after their dedifferentiation from the other early embryophytes.

In summary, two of seven sorghum CCoAOMT and CCoAOMT-like genes, SbCCoAOMT-1 and SbC-CoAOMT-4, revealed high expression in leaves, stem and immature seeds—organs of high lignification activity. Moreover, the clusters of lignin associated transcription factor binding sites are present exclusively in promoters of these two genes; together with a previously reported ability of these enzymes to *O*-methylate caffeoyl-coenzyme A, these results point to the important role of SbCCoAOMT-1 and SbCCoAOMT-4 for lignin synthesis. One of these enzymes, SbCCoAOMT-1, belongs to a clade grouping substrate-specific enzymes known to be involved in lignification, whereas the other, SbCCoAOMT-4, belongs to a clade of substrate-promiscuous CCoAOMT-like enzymes usually associated with species-specific processes, distinct from lignification. The results of this study suggest that enzymatic activity of SbCCoAOMT-4 is associated with lignification, rather than with other pathways; however, more research is required to determine the frequency of such functional reversion in this clade.

The amino acid conservation analysis together with homology modeling allowed us to identify amino acids essential for CCoAOMT and CCoAOMT-like enzymes. The structure of SbCCoAOMT-7, a sorghum protein from the ancestral CCoAOMT-like clade was predicted and compared with SbCCoAOMT-1 from the substrate-specific CCoAOMT clade. The structure comparison allowed us to identify the most structurally divergent regions between these proteins such as the N-terminal  $\alpha$ -helix involved in

dimerization and the insertion loop involved in methyl acceptor discrimination.

The phylogenetic analysis of CCoAOMT and CCoAOMT-like enzymes categorized these proteins into four clades. The ancestral position of the clade comprising CCoAOMT-like proteins closely related to cyanobacterial enzymes is supported by the gene exon structure specific to this clade.

In conclusion, these findings suggest that genes from both groups, CCoAOMT and CCoAOMT-like, contribute to lignin formation in sorghum and may be the key to improving the quality of plant material for bioenergy applications. Furthermore, the results of the conservation analysis suggest potential targets for site-directed mutagenesis, that may affect CCoAOMT enzyme activity. In addition, the differences in gene structures between the clades provide insight into early stages of the land plant evolution, when the demand for monolignols fixed duplicated CCoAOMT genes. Taken together, the results extend our knowledge of CCoAOMT and CCoAOMT-like genes as an important plant gene family and suggest that substrate-promiscuous CCoAOMT-like enzymes in other plants might also be involved in lignification processes.

**Acknowledgements** This work was carried out with the support of National Centre for Research and Development (NCBiR) under Grant no. PBS1/A8/9/2012 (Project Sormisol). The authors would like to thank Dr. Katarzyna Kluzek and Professor Joanna Wesoly from Laboratory of High Throughput Technologies in Institute of Molecular Biology and Biotechnology (Adam Mickiewicz University) for their help in ddPCR experiments. The authors thank Dr. Joanna Sarzynska (IBCH PAS) for her advices and help in protein structure modeling.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflicts of interest with the contents of this article.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

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

Barros J, Serk H, Granlund I, Pesquet E (2015) The cell biology of lignification in higher plants. *Ann Bot (Lond)* 115:1053–1074

Botcher A, Cesarino I, Brombini dos Santos A, Vicentini R, Mayer JLS, Vanholme R, Morreel K, Goeminne G, Moura JCMS., Nobile PM, Carmello-Guerreiro SM, Anjos IA, Creste S, Boerjan W, de Andrade Landell MG, Mazzafera P (2013) Lignification

in sugarcane: biochemical characterization, gene discovery, and expression analysis in two genotypes contrasting for lignin content. *Plant Physiol* 163:1539–1557

Bowman JL, Kohchi T, Yamato KT, Jenkins J, Shu S, Ishizaki K, Yamaoka S, Nishihama R, Nakamura Y, Berger F, Adam C, Aki SS, Althoff F, Araki T, Arteaga-Vazquez MA, Balasubramanian S, Barry K, Bauer D, Boehm CR, Briginshaw L, Caballero-Perez J, Catarino B, Chen F, Chiyoda S, Chovatia M, Davies KM, Delmans M, Demura T, Dierschke T, Dolan L, Dorantes-Acosta AE, Eklund DM, Florent SN, Flores-Sandoval E, Fujiyama A, Fukuzawa H, Galik B, Grimanelli D, Grimwood J, Grossniklaus U, Hamada T, Haseloff J, Hetherington AJ, Higo A, Hirakawa Y, Hundley HN, Ikeda Y, Inoue K, Inoue SI, Ishida S, Jia Q, Kakita M, Kanazawa T, Kawai Y, Kawashima T, Kennedy M, Kinose K, Kinoshita T, Kohara Y, Koide E, Komatsu K, Kopischke S, Kubo M, Kyozuka J, Lagercrantz U, Lin SS, Lindquist E, Lipzen AM, Lu CW, De Luna E, Martienssen RA, Minamino N, Mizutani M, Mizutani M, Mochizuki N, Monte I, Mosher R, Nagasaki H, Nakagami H, Naramoto S, Nishitani K, Ohtani M, Okamoto T, Okumura M, Phillips J, Pollak B, Reinders A, Rövekamp M, Sano R, Sawa S, Schmid MW, Shirakawa M, Solano R, Spunde A, Suetsugu N, Sugano S, Sugiyama A, Sun R, Suzuki Y, Takenaka M, Takezawa D, Tomogane H, Tsuzuki M, Ueda T, Umeda M, Ward JM, Watanabe Y, Yazaki K, Yokoyama R, Yoshitake Y, Yotsui I, Zachgo S, Schmutz J (2017) Insights into land plant evolution garnered from the *Marchantia polymorpha* genome. *Cell* 171(2):287.e15–304.e15

Brandt W, Manke K, Vogt T (2015) A catalytic triad—Lys-Asn-Asp—is essential for the catalysis of the methyl transfer in plant cation-dependent *O*-methyltransferases. *Phytochemistry* 113:130–139

Calvino M, Messing J (2012) Sweet sorghum as a model system for bioenergy crops. *Curr Opin Biotechnol* 2(23):323–329

Carpita NC, McCann MC (2008) Maize and sorghum: genetic resources for bioenergy grasses. *Trends Plant Sci* 13(8):415–420

Dixon RA, Chen F, Guo D, Parvathi K (2001) The biosynthesis of monolignols: a “metabolic grid”, or independent pathways to guaiacyl and syringyl units? *Phytochemistry* 57:1069–1084

Do C, Pollet C, Thévenin J, Sibout R, Denoue D, Barrière Y, Lapiere C, Jouanin L (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

Emanuelsson O, Nielsen H, von Heijne G (1999) ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci* 8(5):978–984

Emiliani G, Fondi M, Fani R, Gribaldo S (2009) A horizontal gene transfer at the origin of phenylpropanoid metabolism: a key adaptation of plants to land. *Biol Direct* 4:7

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39:783–791

Ferrer J, Zubieta C, Dixon R, Noel J (2005) Crystal structures of alfalfa caffeoyl coenzyme A 3-*O*-methyltransferase. *Plant Physiol* 137:1009–1017

Gerik T, Bean B, Vanderlip R (2003) Sorghum growth and development. Texas Cooperative Extension. <http://publications.tamu.edu/>. Accessed 18 Jan 2017

Giordano D, Provenzano S, Ferrandino A, Vitali M, Pagliarani C, Roman F, Cardinale F, Castellarin S, Schubert A (2016) Characterization of a multifunctional caffeoyl-CoA *O*-methyltransferase activated in grape berries upon drought stress. *Plant Physiol Biochem* 101:23–32

Goujon T, Sibout R, Eudes A, MacKay J, Jouanin L (2003) Genes involved in the biosynthesis of lignin precursors in *Arabidopsis thaliana*. *Plant Physiol Biochem* 41:677–687

- 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
- Higo K, Ugawa Y, Iwamoto M, Higo H (1997) PLACE: a database of plant cis-acting regulatory DNA elements. *Nucleic Acids Res* 26:358–359
- Hoffmann L, Maury S, Bergdoll M, Thion L, Erard M, Legrand M (2001) Identification of the enzymatic active site of tobacco caffeoyl-coenzyme A *O*-methyltransferase by site-directed mutagenesis. *J Biol Chem* 276:36831–36838
- Hu B, Jin J, Guo A, Zhang H, Luo J, Gao G (2014) GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* 31:1296–1297
- Ibdah M, Zhang X, Schmidt J, Vogt T (2003) A novel Mg<sup>2+</sup>-dependent *O*-methyltransferase in the phenylpropanoid metabolism of *Mesembryanthemum crystallinum*. *J Biol Chem* 278:43961–43972
- Inoue K, Sewalt V, Balance G, Ni W, Stürzer K, Dixon R (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
- Johnson SM, Lim F-L, Finkler A, Fromm H, Slabas AR, Knight MR (2014) Transcriptomic analysis of *Sorghum bicolor* responding to combined heat and drought stress. *S. bicolor* gene for normalization of expression. *BMC Genom* 15:456
- Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* 8(3):275–282
- Joshi CP, Chiang VL (1998) Conserved sequence motifs in plant *S*-adenosyl-L-methionine-dependent methyltransferases. *Plant Mol Biol* 37(4):663–674
- Kopycki JG, Rauh D, Chumanevich AA, Neumann P, Vogt T, Stubbs MY (2008a) Biochemical and structural analysis of substrate promiscuity in plant Mg<sup>2+</sup>-dependent *O*-methyltransferases. *J Mol Biol* 378:154–164
- Kopycki JG, Stubbs MT, Brandt W, Hagemann M, Porzel A, Schmidt J, Schliemann W, Zenk MH, Vogt T (2008b) Functional and structural characterization of a cation-dependent *O*-methyltransferase from the cyanobacterium *Synechocystis* sp. strain PCC 6803. *J Biol Chem* 283:20888–20896
- Labeuw L, Martone PT, Boucher Y, Case R (2015) Ancient origin of the biosynthesis of lignin precursors. *Biol Direct* 10:23
- Lee YJ, Kim BG, Chong Y, Lim Y, Ahn JH (2008) Cation dependent *O*-methyltransferases from rice. *Planta* 227:641–647
- Li X, Chen W, Zhao Y, Xiang Y, Jiang H, Zhu S, Cheng B (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
- Liu W, Xie Y, Ma J, Luo X, Nie P, Zuo Z, Lahrmann U, Zhao Q, Zheng Y, Zhao Y, Xue Y, Ren J (2015) IBS: an illustrator for the presentation and visualization of biological sequences. *Bioinformatics* 31:3359–3361
- Liu X, Luo Y, Wu H, Xi W, Yu J, Zhang O, Zhou Z (2016) Systematic analysis of *O*-methyltransferase gene family and identification of potential members involved in the formation of *O*-methylated flavonoids in *Citrus*. *Gene* 575:458–472
- Ma QH, Luo HR (2015) Biochemical characterization of caffeoyl coenzyme A 3-*O*-methyltransferase from wheat. *Planta* 242:113–122
- Martz F, Maury S, Pinçon G, Legrand M (1998) cDNA cloning, substrate specificity and expression study of tobacco caffeoyl-CoA 3-*O*-methyltransferase, a lignin biosynthetic enzyme. *Plant Mol Biol* 36:427–437
- 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* 121:215–223
- Meyermans H, Morreel K, Lapierre C, Pollet B, De Bruyn A, Busson R, Herdewijn P, Devreese B, Van Beeumen J, Marita JM, Ralph J, Chen C, Burggraeve B, Van Montagu M, Messens E, Boerjan W (2000) Modifications in lignin and accumulation of phenolic glucosides in poplar xylem upon down-regulation of caffeoyl-coenzyme A *O*-methyltransferase, an enzyme involved in lignin biosynthesis. *J Biol Chem* 275:36899–36909
- Mohammed R, Are AK, Bhavanasi R, Munghate RS, Kavi Kishor PB, Sharma HC (2015) Quantitative genetic analysis of agronomic and morphological traits in sorghum, *Sorghum bicolor*. *Front Plant Sci* 6:945
- Mullet J, Morishige D, McCormick R, Truong S, Hilley J, McKinley B, Anderson R, Olson SN, Rooney W (2014) Energy sorghum—a genetic model for the design of C4 grass bioenergy crops. *J Exp Bot* 65(13):3479–3489
- Pakusch AE, Kneusel RE, Matern U (1989) *S*-adenosyl-L-methionine:trans-caffeoyl-coenzyme A 3-*O*-methyltransferase from elicitor-treated parsley cell suspension cultures. *Arch Biochem Biophys* 271(2):488–494
- Pang SL, Ong SS, Lee HH, Zamri Z, Kandasamy KI, Choong CY, Wickneswari R (2014) Isolation and characterization of CCoAOMT in interspecific hybrid of *Acacia auriculiformis* × *Acacia mangium*—a key gene in lignin biosynthesis. *Genet Mol Res* 13:7217–7238
- Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otiillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Mehboob-ur-Rahman, Ware D, Westhoff P, Mayer KFX, Messing J, Rokhsar DS (2009) The *Sorghum bicolor* genome and the diversification of grasses. *Nature* 457:551–556
- Patzlaff A, McInnis S, Courtenay A, Surman C, Newman LJ, Smith C, Bevan MW, Mansfield S, Whetten RW, Sederoff RR, Campbell MM (2003) Characterisation of a pine MYB that regulates lignification. *Plant J* 36(6):743–754
- Pinçon G, Maury S, Hoffmann L, Geoffroy P, Lapierre C, Pollet B, Legrand M (2001) Repression of *O*-methyltransferase genes in transgenic tobacco affects lignin synthesis and plant growth. *Phytochemistry* 57:1167–1176
- 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
- Renault H, Alber A, Horst NA, Basilio Lopes A, Fich EA, Kriegshausler L, Wiedemann G, Ullmann P, Herrgott L, Erhardt M, Pineau E, Ehrling J, Schmitt M, Rose JKC, Reski R, Werck-Reichhart D (2017) A phenol-enriched cuticle is ancestral to lignin evolution in land plants. *Nat Commun* 8:14713
- Sahr T, Adam T, Fizames C, Maurel C, Santoni V (2010) *O*-carboxyl- and *N*-methyltransferases active on plant aquaporins. *Plant Cell Physiol* 51:2092–2104
- Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor
- Shaipulah NFM, Muhlemann JK, Woodworth BD, Van Moerkercke A, Verdonk JC, Ramirez AA, Haring MA, Dudareva N, Schuurink RC (2016) CCoAOMT down-regulation activates anthocyanin biosynthesis in Petunia. *Plant Physiol* 170:717–731



- Stothard P (2000) The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques* 28:1102–1104
- Tamagnone L, Merida A, Parr A, Mackay S, Culianez-Macia FA, Roberts K, Martina C (1998) The AmMYB308 and AmMYB330 transcription factors from antirrhinum regulate phenylpropanoid and lignin biosynthesis in transgenic tobacco. *Plant Cell* 10:135–154
- Tamura K, Stecher G, Peterson D, Filipinski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 30(12):2725–2729
- Toraman HE, Vanholme R, Borén E, Vanwonterghem Y, Djokic MR, Yildiz G, Ronsse F, Prins W, Boerjan W, Van Geem KM, Marin GB (2016) Potential of genetically engineered hybrid poplar for pyrolytic production of bio-based phenolic compounds. *Bioresour Technol* 207:229–236
- Vanholme R, Demedts B, Morreel K, Ralph J, Boerjan W (2010) Lignin biosynthesis and structure. *Plant Physiol* 153(3):895–905
- Vermerris W (2011) Survey of genomics approaches to improve bioenergy traits in maize, sorghum and sugarcane. *J Integr Plant Biol* 53(2):105–119
- Walker AM, Sattler SA, Regner M, Jones JP, Ralph J, Vermerris W, Sattler SE, Kang CH (2016) The structure and catalytic mechanism of *Sorghum bicolor* caffeoyl-CoA *O*-methyltransferase. *Plant Physiol* 172:78–92
- Wang GF, Balint-Kurti PJ (2016) Maize homologs of CCoAOMT and HCT, two key enzymes in lignin biosynthesis, form complexes with the NLR Rp1 protein to modulate the defense response. *Plant Physiol* 171:2166–2177
- Weng JK, Chapple C (2010) The origin and evolution of lignin biosynthesis. *New Phytol* 187(2):273–285
- Weng JK, Li X, Bonawitz ND, Chapple C (2008) Emerging strategies of lignin engineering and degradation for cellulosic biofuel production. *Curr Opin Biotechnol* 19:166–172
- Widiez T, Hartman TG, Dudai N, Yan Q, Lawton M, Havkin-Frenkel D, Belanger FC (2011) Functional characterization of two new members of the caffeoyl CoA *O*-methyltransferase-like gene family from *Vanilla planifolia* reveals a new class of plastid-localized *O*-methyltransferases. *Plant Mol Biol* 76:475–488
- Xu Z, Zhang D, Hu J, Zhou X, Ye X, Reichel KL, Stewart NR, Syrenne RD, Yang X, Gao P, Shi W, Doepcke C, Sykes RW, Burris JN, Bozell JJ, Cheng MZ, Hayes DG, Labbe N, Davis M, Stewart CN Jr, Yuan JS (2009) Comparative genome analysis of lignin biosynthesis gene families across the plant kingdom. *BMC Bioinform* 10(Suppl. 11):S3
- Xu RX, Zhao Y, Gao S, Zhang YY, Li DD, Lou HX, Cheng AX (2015) Functional characterization of a plastidial cation-dependent *O*-methyltransferase from the liverwort *Plagiochasma appendiculatum*. *Phytochemistry* 118:33–41
- Yang Q, He Y, Kabahuma M, Chaya T, Kelly A, Borrego E, Bian Y, El Kasmi F, Yang L, Teixeira P, Kolkman J, Nelson R, Kolomiets M, L Dangel J, Wisser R, Caplan J, Li X, Lauter N, Balint-Kurti P (2017) A gene encoding maize caffeoyl-CoA *O*-methyltransferase confers quantitative resistance to multiple pathogens. *Nat Genet* 49(9):1364–1372
- Ye Z-H, Kneusel RE, Matern U, Varner JE (1994) An alternative methylation pathway in lignin biosynthesis in *Zinnia*. *Plant Cell* 6:1427–1439
- Ye Z-H, Zhong R, Morrison WH III, Himmelsbach DS (2001) Caffeoyl coenzyme A *O*-methyltransferase and lignin biosynthesis. *Phytochemistry* 57:1177–1185
- Yu CS, Cheng CW, Su WC, Chang KC, Huang SW, Hwang JK, Lu CH (2014) CELLO2GO: a web server for protein subcellular localization prediction with functional gene ontology annotation. *PLoS ONE* 9(6):e99368
- Zhang G, Zhang Y, Xu J, Niu X, Qi J, Tao A, Zhang L, Fang P, Lin LH, Su J (2014) The CCoAOMT1 gene from jute (*Corchorus capsularis* L.) is involved in lignin biosynthesis in *Arabidopsis thaliana*. *Gene* 546:398–402
- Zhong R, Morrison WH, Negrel J, Yea ZH (1998) Dual methylation pathways in lignin biosynthesis. *Plant Cell* 10:2033–2045