3 Biotech (2022) 12:149
https://doi.org/10.1007/s13205-022-03208-0

ORIGINAL ARTICLE

Genome-wide analysis of the *CAD* gene family reveals two *bona fide CAD* genes in oil palm

Chong Yu Lok Yusuf[1] · Nuraini Sabri Nabilah[1] · Nur Atiqah Amiza Mohd Taufik[1] · Idris Abu Seman[2] · Mohd Puad Abdullah[3]

Received: 3 January 2022 / Accepted: 21 May 2022 / Published online: 20 June 2022

Abstract

Cinnamyl alcohol dehydrogenase (CAD) is the key enzyme for lignin biosynthesis in plants. In this study, genome-wide analysis was performed to identify *CAD* genes in oil palm (*Elaeis guineensis*). Phylogenetic analysis was then conducted to select the *bona fide EgCAD*s. The *bona fide EgCAD* genes and their respective 5′ flanking regions were cloned and analysed. Their expression profiles were evaluated in various organs using RT-PCR. Seven *EgCAD* genes (*EgCAD1*-7) were identified and divided into four phylogenetic groups. EgCAD1 and EgCAD2 display high sequence similarities with other *bona fide* CADs and possess all the signature motifs of the *bona fide* CAD. They also display similar 3D protein structures. Gene expression analysis showed that *EgCAD1* was expressed most abundantly in the root tissues, while *EgCAD2* was expressed constitutively in all the tissues studied. *EgCAD1* possesses only one transcription start site, while *EgCAD2* has five. Interestingly, a TC microsatellite was found in the 5′ flanking region of *EgCAD2*. The 5′ flanking regions of *EgCAD1* and *EgCAD2* contain lignin-associated regulatory elements i.e. AC-elements, and other defence-related motifs, including W-box, GT-1 motif and CGTCA-motif. Altogether, these results imply that *EgCAD1* and *EgCAD2* are *bona fide CAD* involved in lignin biosynthesis during the normal development of oil palm and in response to stresses. Our findings shed some light on the roles of the *bona fide CAD* genes in oil palm and pave the way for manipulating lignin content in oil palm through a genetic approach.

Keywords Cinnamyl alcohol dehydrogenase · CAD · Monolignols · Lignin · Oil palm

Introduction

Lignin is a natural polymer produced by plants to support their growth and development (Yoon et al. 2015). Accumulation of lignin in the secondary cell wall strengthens the cell wall and confers it rigidity, allowing the plant to stand upright. Studies showed that the accumulation of lignin in plant tissues enhances drought tolerance (Tu et al. 2020; Wen et al. 2021). Moreover, lignin also serves an important role in plant defence by acting as a physical barrier to restrict pathogen invasion (Ma et al. 2018b; Lee et al. 2019). Various important roles of lignin in plants have been reviewed by Liu et al. (2018b). Despite its crucial biological roles in plants, lignin negatively affects the quality of animal feedstock and products of certain industries such as paper pulp and biofuels (Hodgson-Kratky et al. 2019; Ladeira Ázar et al. 2020; Chen et al. 2001; Jung and Allen 1995). Hence, manipulation of lignin content in plants through a transgenic approach has gained much attention in recent years. Reduction of lignin content in plant has been attempted by suppressing the expression of lignin biosynthetic genes in order to improve the saccharification efficiency of plant biomass (Bryant et al. 2020; Bewg et al. 2016; Poovaiah et al. 2014).

✉ Chong Yu Lok Yusuf
yusufchong@uitm.edu.my

✉ Mohd Puad Abdullah
puad@upm.edu.my

[1] Laboratory of Plant Genetic and Cell Biology, Faculty of Plantation and Agrotechnology, Universiti Teknologi MARA, Jasin Campus, 77300 Merlimau, Melaka, Malaysia

[2] Malaysian Palm Oil Board (MPOB), No. 6, Persiaran Institusi, Bandar Baru Bangi, 43000 Kajang, Selangor, Malaysia

[3] Department of Cell and Molecular Biology, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia

In most plants, lignin is composed of three main monolignols, namely the coniferyl alcohol, sinapyl alcohol and *p*-coumaryl alcohol which produce the G, S and H lignin units, respectively, upon being integrated into lignin (Wang et al. 2013). An uncommon caffeyl alcohol that forms C unit was also reported in the seed coats of vanilla orchid and cacti under the subfamily Cactoidae (Chen et al. 2013, 2012a). Recent studies also showed that other phenolic compounds behave as authentic monolignols and are integrated in the lignin of some plants (del Río et al. 2020). The lignin composition in plants varies from species to species, with the dicots having a large portion of the S and G units but trace amount of H unit, and the monocots demonstrating a higher amount of H unit, in addition to the S and G units (Chanoca et al. 2019; Vanholme et al. 2010). Regardless of the types of canonical monolignols, lignin monomers are synthesised in plant through the lignin biosynthetic pathway, whose final step is regulated by cinnamyl alcohol dehydrogenase (CAD; EC 1.1.1.195) (Xie et al. 2018).

CAD catalyses the reduction of cinnamyl aldehydes into their corresponding alcohols in the presence of NADPH as a cofactor. It is encoded by a multigene family with varying numbers of family members in higher plants. Genome-wide analysis had discovered 5, 7, 12 and 14 CAD isoforms in *Cucumis melo*, *Brachypodium distachyon*, *Oryza sativa* and *Sorghum bicolor*, respectively (Saballos et al. 2009; Jin et al. 2014; Tobias and Chow 2005; Bukh et al. 2012). Phylogenetic analysis shows that the CAD from higher plants could be divided into three to seven groups, depending on the sequences used in the analysis (Bukh et al. 2012; Jin et al. 2014; Ma 2010; Deng et al. 2013). Regardless of the number of groups that appeared in the phylogenetic tree, members of group I CAD are regarded as the *bona fide* CADs associated with lignin biosynthesis in plants. Examples of the *bona fide CAD* from higher plants are *SbCAD2* from *Sorghum bicolor*, *OsCAD2* from *Oryza sativa*, *PtoCAD1* from *Populus tomentosa*, *ZmCAD2* from *Zea mays*, etc. (Park et al. 2018; Liu et al. 2021a; Saballos et al. 2009; Chao et al. 2014).

Functional studies of *CAD* genes in various plants demonstrated the crucial roles of CAD in lignin biosynthesis. Generally, these studies were conducted through gain- or loss-of-function analyses (Martin et al. 2019; Preisner et al. 2014). Overexpression of *CAD* gene always leads to ectopic deposition of lignin in plant tissues (Li et al. 2019; Ma et al. 2018a). Additionally, overexpression of *LtuCAD1* from *Liriodendron tulipifera* rendered the *Arabidopsis cad4 cad5* double mutant to regain lignin content similar to that of wild-type plants (Xu et al. 2013). In contrast, suppression of *CAD* resulted in a significant reduction of lignin content in plants as observed in transgenic switchgrass, medicago and poplar (Saathoff et al. 2011a; Van Acker et al. 2017; Zhao et al. 2013). It is worth noting that the brown midrib phenotype reported in maize and sorghum mutants is due to a natural mutation in the *CAD* gene or reduced *CAD* expression (Tsuruta et al. 2010; Chen et al. 2012b; Li and Huang 2017). More recently, Yamamoto et al. (2020) also reported that the formation of atypical red-coloured wood in the Sekizaisou variety of mulberry plant is associated with a mutation in the *CAD1* gene.

Apart from serving important functions in plant development, *CAD* genes also play important roles in plant defence and adaptation to stresses (Tronchet et al. 2010; Kim and Huh 2019). Many studies showed that the expressions of *CAD* genes were upregulated in response to various stimuli, including low temperature, wounding, phytohormones, pathogens and UV light (Jin et al. 2014; Park et al. 2018; Kim et al. 2010; Cheng et al. 2013). The critical role played by *TaCAD12* in resistance to sharp eyespot disease has been demonstrated through overexpression and silencing of the gene in wheat plants (Rong et al. 2016). Furthermore, improved heavy metal tolerance was observed in *Arabidopsis* plant overexpressing a *SaCAD* gene from *Sedum alfredii* (Qiu et al. 2018). Liu et al. (2020) also have revealed that the *CmCAD2* and *CmCAD3* from *Cucumis melo* played a major role in drought tolerance in transgenic *Arabidopsis* plants. Meanwhile, the host plant displayed enhanced drought sensitivity when the genes were silenced individually and combined. Together, these studies underscored the crucial roles served by *CAD* genes in plant adaptation to biotic and abiotic stresses.

Despite being well studied in various staple crops, CAD is not well studied in oil crops. To our knowledge, there is no report on the study of *CAD* genes in oil palm, an oil crop that largely contributes to the edible oil in the world market. Hence, we performed a genome-wide analysis to identify the members of the *CAD* gene family in oil palm. The *bona fide CAD* genes and their respective promoters were cloned and analysed in this study.

Materials and methods

Identification of putative *EgCAD* genes in oil palm genome

Sequences of AtCAD4 and AtCAD5, the *bona fide* CAD from *Arabidopsis thaliana*, were used as query to identify potential *EgCAD* genes from the oil palm genome through tBLASTn in National Center for Biotechnology Information (NCBI). The identified potential *EgCAD* genes were examined for the presence of cinnamyl-alcohol dehydrogenase (CAD) domain or medium-chain dehydrogenases/reductases (MDR) superfamily domain using the Conserved Domain Database in NCBI.

Identification of *bona fide EgCAD* genes through phylogenetic analysis

The amino acid sequences of CAD family members from *Arabidopsis thaliana*, *Oryza sativa* and *Brachypodium distachyon* were retrieved from NCBI and Phytozome database (https://phytozome-next.jgi.doe.gov/). A phylogenetic tree was constructed using the Maximum Likelihood method based on the JTT matrix-based model in the MEGA7 software. The amino acid sequences of CADs used in the phylogenetic analysis are presented in Online Resource 1.

Plant materials

The oil palm material (variety *pisifera*, 367 P) used in the cloning of *bona fide EgCAD* genes and their 5′ flanking region sequences was sampled from the Malaysian Palm Oil Board (MPOB) Kluang Research Station (GPS coordinates: 1.956501, 103.372014) situated in Johor, Malaysia. The oil palm samples, including the coleoptile and primary root of germinated seed, young leaf and young root of 1-year-old seedling, immature fruitlet and mesocarp tissue of developing fruit collected from the MPOB/UKM Research Station (GPS coordinates: 2.906570, 101.783928) located in Selangor, Malaysia were used to study the expression profile of *bona fide EgCAD* genes.

Extraction of nucleic acids and cDNA synthesis

Total RNA samples were extracted from various organs of oil palm as described in the study by Yusuf et al. (2018b). For the samples used in gene expression analysis, cDNA was synthesized from 3 µg of total RNA using RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific, USA). Genomic DNA was isolated from the oil palm root tissues using Genomic DNA Mini Kit (Geneaid, Taiwan) according to the suggested protocol.

Cloning of the ORFs of *EgCAD1* and *EgCAD2* genes

The ORFs of *EgCAD1* and *EgCAD2* genes were amplified from the cDNA derived from root tissue by PCR. The reaction mixture contained 1 × Phusion HF buffer, 0.2 mM of dNTPs, forward and reverse primers 0.5 µM each, 2 µL of cDNA template, 1 unit of Phusion DNA Polymerase (Thermo Scientific, USA) and dH_2O to a final volume of 50 µL. The primers used in PCR are listed in Online Resource 2. The primers were designed to anneal to the 5′ and 3′ UTRs of the target genes in order to obtain the sequence of the entire ORF. The PCR was performed using the following thermal cycling profile: 98 °C (30 s); [98 °C (10 s), 62 °C (15 s), 72 °C (50 s)] × 35 cycles; 72 °C for 5 min. The gel purified PCR products were cloned using CloneJET PCR Cloning Kit (Thermo Scientific, USA) and sequenced by First BASE Laboratories Sdn Bhd (Selangor, Malaysia).

In silico analysis

The molecular weights and the theoretical isoelectric points (pI) of EgCADs were predicted using the ProtParam tool at the ExPASy website (web.expasy.org/protparam). The multiple sequence alignment was performed using the ClustalW method in BioEdit version 7.0. The amino acid sequences of *bona fide* CADs used in the multiple sequence alignment were retrieved from the NCBI and Phytozome. The GenBank/Phytozome accession numbers of the sequences used are: ZmCAD2: CAA06687, SbCAD2: XP_021314961, OsCAD2: Q6ZHS4, BdCAD5: Bradi3g06480 (Phytozome), AtCAD4: At3g19450 (Phytozome), AtCAD5: At4g34230 (Phytozome), EgCAD1: XP_010932224 and EgCAD2: XP_010943210. The 3D structures of *bona fide* EgCADs were constructed based on the crystal structure of AtCAD5 (SMTL ID: 2cf5.1.B) using SWISS-MODEL (https://swissmodel.expasy.org/).

Expression profile analysis

The expression profiles of *bona fide EgCAD* genes were studied by semi-quantitative PCR in various plant organs collected from oil palm at different developmental stages. The oil palm *GAPDH* gene (accession number: DQ267444) was used as the reference gene. The PCR was performed in a total volume of 20 µL reaction mixture consisting of 1 × *Taq* Buffer, 2 mM $MgCl_2$, 0.2 mM dNTPs, forward and reverse primers 0.2 µM each (Online Resource 2), 50 ng cDNA, 0.5 unit *Taq* DNA Polymerase (Thermo Scientific, USA) and dH_2O. The PCR profile used is as follows: 95 °C (3 min); 95 °C (20 s), 58 °C (25 s), 72 °C (25 s) for 28 cycles; 72 °C (5 min). The PCR products (5 µL) were resolved on 1.5% TAE agarose gel and documented.

Cloning and sequence analysis of the 5′ flanking regions of *EgCAD1* and *EgCAD2* genes

To obtain the regulatory sequences of *EgCAD1* and *EgCAD2*, approximately 1.5 kb of the 5′ flanking regions of *EgCAD1* and *EgCAD2* were amplified by PCR. The PCR components in the reaction mixture are identical to that used in ORF amplification, as stated in the section describing the cloning of ORFs, except that 100 ng of gDNA was used as the template and the primers used were shown in Online Resource 2. Forward primer was designed to bind at the

position approximately 1.5 kb upstream of the ORF, while the reverse primers bind to the ORF of the target gene. The thermal cycling profile used is also the same as the previous one, except that the annealing temperature was set to 60 °C (*EgCAD1*) or 62 (for *EgCAD2*) and the extension step was prolonged to 45 s. The PCR products were cloned and sequenced in the same way as stated in the section describing the cloning of ORFs. The *cis*-acting elements present in the 5′ regulatory regions of *EgCAD1* and *EgCAD2* were searched from PlantCARE online database (Lescot et al. 2002) and identified manually based on information from published studies.

Determination of the transcription start sites of *EgCAD1* and *EgCAD2* genes

The transcription start sites (TSS) of *EgCAD1* and *EgCAD2* genes were determined using 5′ Rapid Amplification of cDNA Ends (RACE) method. The cDNA template for 5′ RACE was synthesized from 1 µg of total RNA isolated from young leaf sample using the Template Switching RT Enzyme Mix (NEB, USA) and the primers listed in Online Resource 2 based on the protocol recommended by New England Biolabs. The 5′ RACE-PCR was performed in a 50-µL PCR mixture containing 1 × Phusion HF buffer, 0.2 mM of dNTPs, 0.5 µM TSO-specific primer, 0.5 µM GSP1 or GSP2 primer (Online Resource 2), 2 µL of RACE template, 1 unit of Phusion DNA Polymerase (Thermo Scientific, USA) and dH_2O. The PCR thermal cycling profile used was 98 °C for 30 s, followed by 35 cycles of denaturation at 98 °C for 10 s, annealing at 63 °C for 30 s, extension at 72 °C for 40 s and finally, a 5-min final extension step at 72 °C. The PCR products were cloned and sequenced as described in previous section.

Results

There are seven *EgCAD* genes in the oil palm genome

Initially, eight loci annotated as *CAD* were identified in the oil palm genome through tBLASTn. However, a close examination of the protein sequences encoded by these loci revealed that the CAD conserve domain was absent in one of the protein sequences examined (Online Resource 3). Therefore, only seven loci (EgCAD1-7) were considered as putative CAD, while the gene that lacks the CAD conserved domains was regarded as a pseudogene. The *EgCAD* genes possess open reading frames (ORF) of varying sizes, ranging from 1068 to 1128 bp. The ORFs of *EgCAD1-7* can be translated into polypeptides consisting of 355 to 375 amino acids, with an average molecular weight (Mw) of 39.3 kDa and a theoretical isoelectric point (pI) between 6.08 and 7.55 (Table 1). Analysis of the gene location showed the *EgCAD1-7* genes were distributed unevenly across the 16 chromosomes of oil palm. In short, three *EgCAD* genes are located on chromosome 4, one each in chromosomes 10 and 11, and two are unknown (Table 1).

Table 1 Details of *EgCAD* genes in oil palm genome

Gene	Accession	Locus	Chromosome	Domain	ORF (bp)	Amino acid	MW (kDa)	pI
EgCAD1	XM_010933922 OK539815[a]	LOC105052938	10	PLN02514	1071	356	38.5	6.15
EgCAD2	XM_010944908 OK539816[a]	LOC105061001	Unknown	PLN02514	1074	357	38.8	6.08
EgCAD3	XM_010943394	LOC105059896	Unknown	cl31545 probable cinnamyl alcohol dehydrogenase	1110	369	39.8	6.79
EgCAD4	XM_010921792	LOC105044021	4	cd05283 Cinnamyl alcohol dehydrogenases (CAD)	1077	358	39.1	6.18
EgCAD5	XM_010920701	LOC105043232	4	cd05283 Cinnamyl alcohol dehydrogenases (CAD)	1128	375	41.8	6.75
EgCAD6	XM_010920760	LOC105043276	4	cd05283 Cinnamyl alcohol dehydrogenases (CAD)	1068	355	38.6	6.75
EgCAD7	XM_029267211	LOC105054383	11	cd05283 Cinnamyl alcohol dehydrogenases (CAD)	1068	355	38.8	7.55

[a]Sequence obtained through cloning in present study

EgCAD1 and *EgCAD2* encoded for *bona fide* CAD in oil palm

To identify the *bona fide EgCAD* gene(s) responsible for lignin biosynthesis in oil palm, we performed a phylogenetic analysis on the seven EgCADs and the CAD family members from *Arabidopsis thaliana*, *Oryza sativa* and *Brachypodium distachyon*. The CAD family members from higher plants were classified into five major groups based on phylogenetic analysis. The phylogenetic tree shows that the CADs of oil palm, indicated with the filled circle (●) in Fig. 1, were distributed in four different groups. Group I comprises EgCAD1, EgCAD2, and the well-studied *bona fide* CADs from *Arabidopsis* (AtCAD4 and AtCAD5), *Brachypodium* (BdCAD5) and rice (OsCAD2). Group II is a dicot-specific group that contains only the CAD members from *Arabidopsis*, without any CAD members from monocots. Hence, none of the oil palm EgCADs was found in this group. Next, group III is the largest group that contains the CAD members from both monocot and dicot species. While multiple CAD members from each of the species included in this analysis

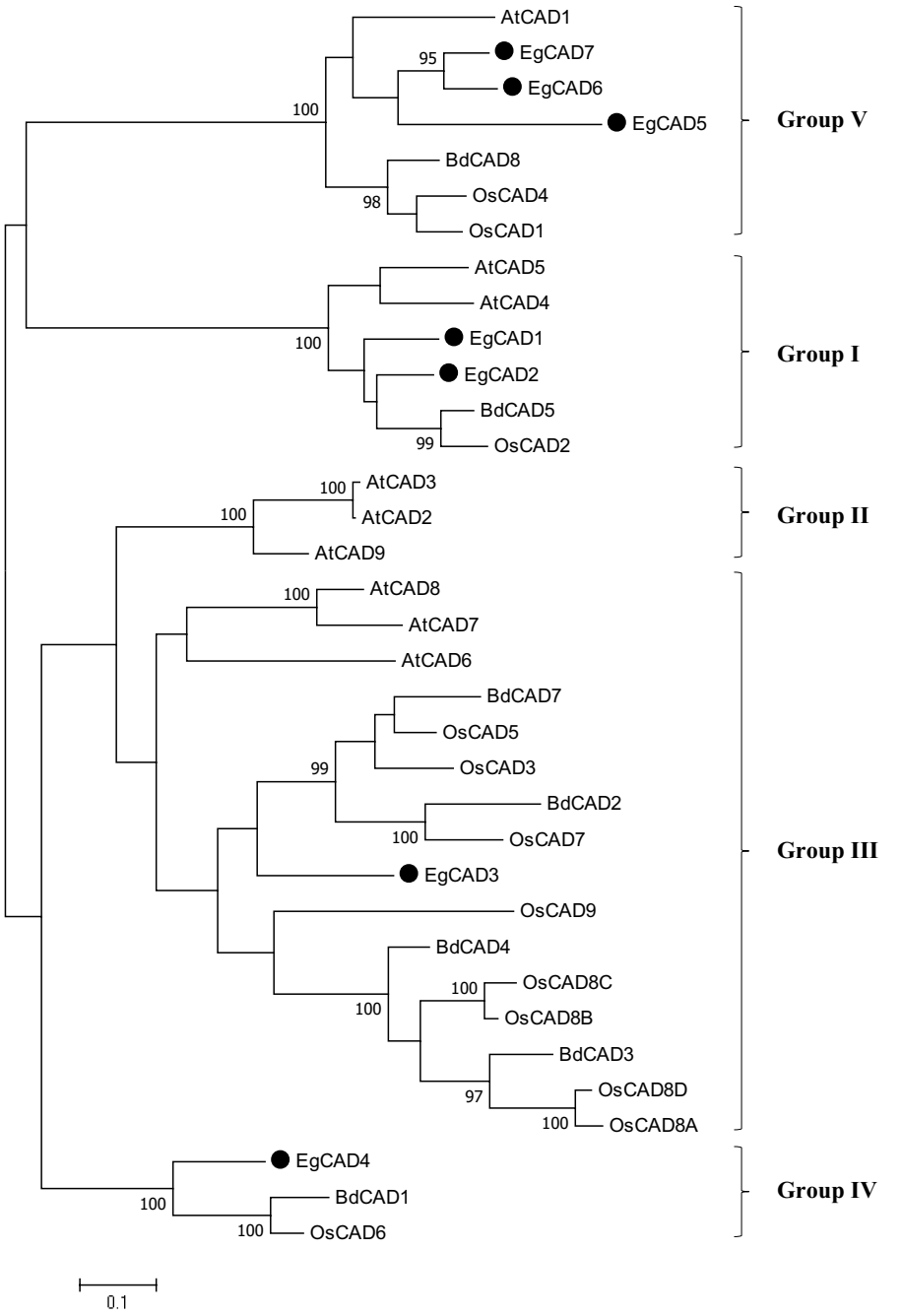

Fig. 1 Phylogenetic analysis of CADs from oil palm (*Elaeis guineensis*) and selected higher plants. The analysis involved 35 amino acid sequences of CADs from *Elaeis guineensis* (Eg), *Arabidopsis thaliana* (At), *Oryza sativa* (Os) and *Brachypodium distachyon* (Bd). The nomenclature of CADs used in present study is according to Kim et al. (2004) for *Arabidopsis thaliana*, Tobias and Chow (2005) for *Oryza sativa*, and Bukh et al. (2012) for *Brachypodium distachyon*. The phylogenetic tree was constructed using the Maximum Likelihood method in MEGA7 with 1000 bootstrap replicates. Numbers shown at the nodes represent the bootstrap values. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site

were clustered in group III, only EgCAD3 from oil palm fell into this group. Contrary to group II, group IV only consists of the CAD members from monocot species, i.e. BdCAD1, OsCAD6 and EgCAD4. Last, group V also encompasses the CAD members from monocot and dicot species. Three oil palm CAD members, namely the EgCAD5, EgCAD6 and EgCAD7, belonged to this group. From the result of the phylogenetic analysis, *EgCAD1* and *EgCAD2* are suggested to be the *bona fide CAD* genes in oil palm.

In silico analysis of *bona fide EgCAD* genes

Prior to the characterization of the *bona fide EgCAD* genes, we cloned the ORFs of *EgCAD1* and *EgCAD2* to validate their sequences. The sequences were deposited in GenBank under the accession numbers OK539815 for *EgCAD1* and OK539816 for *EgCAD2*. Our sequencing results revealed minor differences in the cDNA sequence of *EgCAD1* obtained through cloning compared to the respective sequence of the gene (LOC105052938) retrieved from the oil palm genome sequencing project. The variances were observed at positions 216 and 843 (relative to the first nucleotide of ORF), which the C^{216} and G^{843} in *EgCAD1* sequence retrieved from the genome database appear to be T^{216} and C^{843} in the sequence we cloned. However, these differences do not change the encoded amino acid sequence. On the other hand, the sequence of the cloned *EgCAD2* was identical to the respective sequence (LOC105061001) in the oil palm genome database. Next, characterization of the *bona fide EgCAD* genes were performed based on the sequences we obtained through cloning.

EgCAD1 and EgCAD2 share 81.64% sequence identity at the protein level. BLASTp analysis showed that EgCAD1 shares 81.07% identity with the CAD2 of *Phalaenopsis equestris* (accession no.: XP_020592027), while EgCAD2 and the CAD1 of *Ananas comosus* (accession no.: OAY67853) exhibit a higher identity at 86.8%. To better compare the *bona fide* CADs from oil palm and other higher plant species, we aligned the protein sequences of EgCAD1 and EgCAD2 together with the selected *bona fide* CADs that had been well studied. The multiple sequence alignment analysis revealed that EgCAD1 and EgCAD2 share the same signature motifs of CAD with other *bona fide* CADs (Fig. 2). Notably, the presence of Zn1-binding motif and Zn2-binding motif with the consensus sequences $GHE(X)_2G(X)_5G(X)_2V$ and $GD(X)_{10}C(X)_2C(X)_2C(X)_7C$, respectively, in EgCAD1 and EgCAD2, indicating these EgCADs are Zn-dependent dehydrogenases which belong to the medium-chain dehydrogenase/reductase superfamily. The analysis also revealed that a glycine-rich motif $GXG(X)_2G$ known to be the binding site of a cofactor, i.e., NADPH was well conserved in EgCAD1 and EgCAD2. Apart from that, EgCAD1 and EgCAD2 also contained the amino acid residues corresponding to the Ser211, Ser212, Ser213, Lys216 and Gly275 in AtCAD5, which were found to have a direct interaction with the cofactor (Youn et al. 2006).

We also examined the 12 amino acid residues constituting the substrate-binding pocket of CAD as identified by Youn et al. (2006) in AtCAD5. The amino acid residues forming the substrate-binding pocket of EgCAD1 and EgCAD2 were highly similar to that of AtCAD5 (Fig. 2). Only two amino acids (Val96 and Val301) differ in EgCAD1, and one (Val95) for the case of EgCAD2 when compared to AtCAD5. Interestingly, the Ile300 in EgCAD2 (equivalent to Ile300 in AtCAD5) was replaced by Val301 in EgCAD1. Although the amino acid substitution was conservative, it resulted in a greater substrate-binding pocket in EgCAD1, suggesting that EgCAD1 has a greater substrate versatility than EgCAD2. Next, we constructed the three-dimensional (3D) models of EgCAD1 and EgCAD2 proteins through homology modelling to examine their structures (Fig. 3). As anticipated, EgCAD1 and EgCAD2 share a very similar protein structure with AtCAD5. Both 3D models were homodimers consisting of the nucleotide-binding and catalytic domains (Fig. 3a, b). Furthermore, the homodimers also contained the zinc ions required for its catalytic activity. A close examination of the 3D models also revealed that the β strand (βF) of the two subunits interacted with each other to form a dimmer in both EgCADs studied (Fig. 3a, b). The difference in the structures of the substrate-binding pockets of EgCAD1 and EgCAD2 is clearly displayed in Fig. 3c, d. Judging by the features displayed by the EgCAD1 and EgCAD2 protein models, it is suggested that the two EgCADs studied are *bona fide* CAD.

The *bona fide EgCAD* genes exhibited different expression behaviours in oil palm

To infer the role of *bona fide EgCAD* genes in oil palm, we analysed the expression profiles of *EgCAD1* and *EgCAD2* genes in various organs, including the vegetative and reproductive parts of oil palm (Online Resource 4), by semi-quantitative PCR. Surprisingly, the two *EgCAD* genes displayed two distinct expression patterns (Fig. 4). *EgCAD1* manifested varying expression levels in different oil palm organs, and the highest gene expression level was observed in the primary root of germinated seeds. The gene was also abundantly expressed in the coleoptile of germinated seeds and the young root tissues of 1-year-old plants. In addition, the transcript of *EgCAD1* was also detectable in immature fruitlets and, to a lesser extent, in young leaf tissues collected from 1-year-old seedlings. However, no expression of *EgCAD1* was observed in the mesocarp tissue of developing fruits. Contrary to *EgCAD1*, *EgCAD2* was expressed at high levels in all the organs studied regardless of the plant's developmental stage, reflecting its importance in lignin biosynthesis. Together, the gene expression profiles suggest that *EgCAD2* play a predominant role in monolignol production in oil palm, and it is aided by *EgCAD1* in certain tissues, particularly in root tissues.

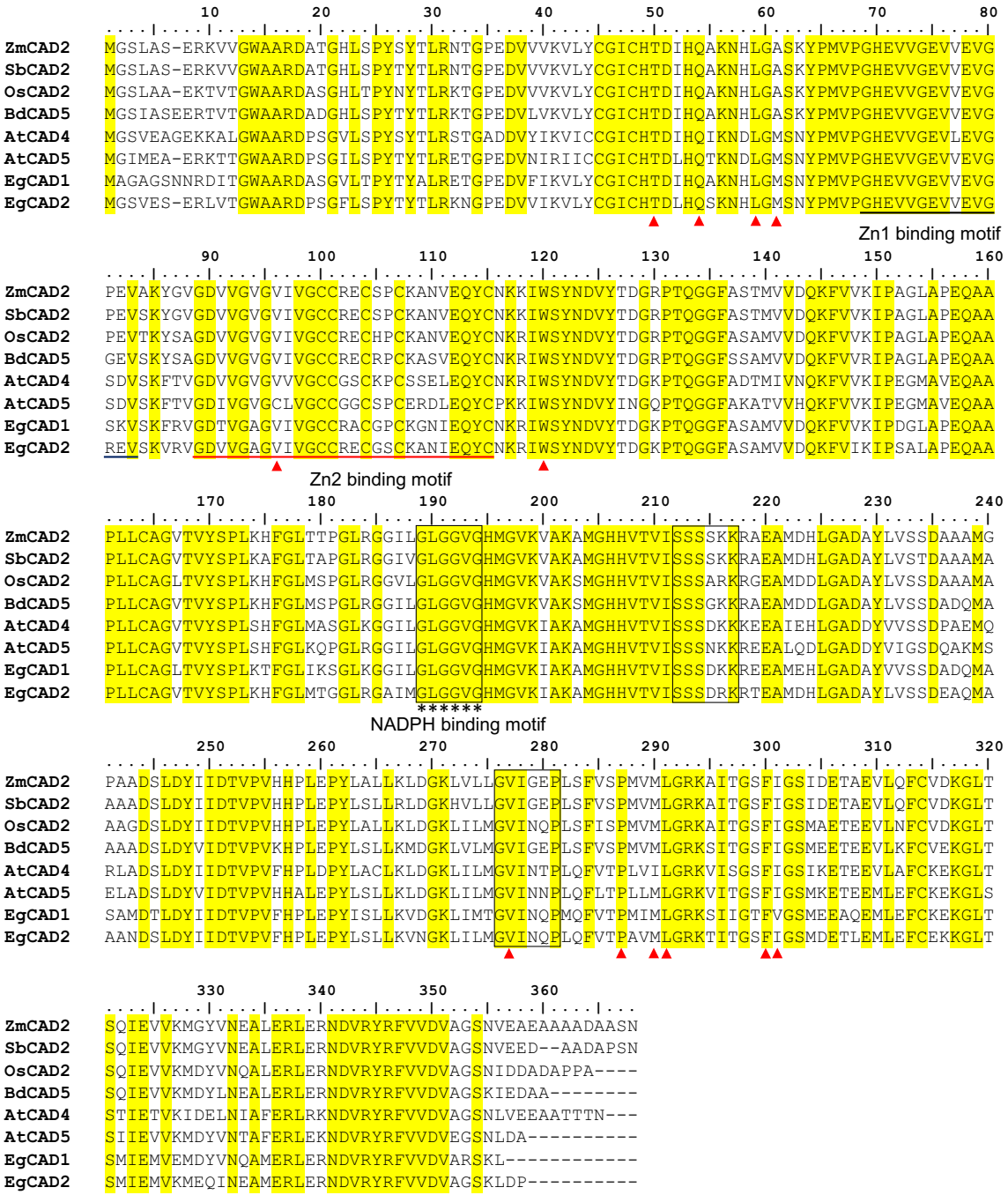

Fig. 2 Amino acid sequence alignment of EgCAD1, EgCAD2 and the *bona fide* CADs of *Zea mays* (ZmCAD2), *Sorghum bicolor* (SbCAD2), *Oryza sativa* (OsCAD2), *Brachypodium distachyon* (BdCAD5) and *Arabidopsis thaliana* (AtCAD4 and AtCAD5). The identical amino acid residues are shaded. The conserved Zn^{2+}-binding motifs are underlined. The NADP-binding motif is indicated with asterisk (*). The residues constituting a substrate-binding pocket are marked with triangles (▲) and the residues forming the loops that interact with NADP as proposed by Youn et al. (2006) are boxed

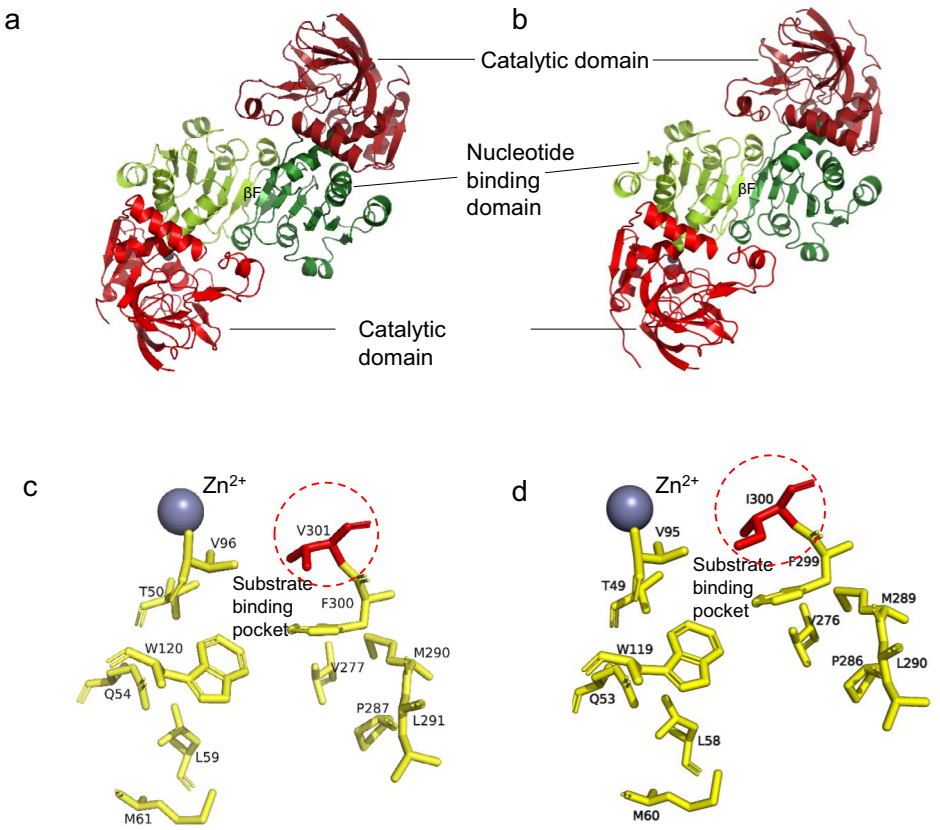

Fig. 3 Three-dimensional protein structure models of *bona fide* EgCADs. The homodimer structures of EgCAD1 (**a**) and EgCAD2 (**b**) were constructed using SWISS-MODEL workspace based on the crystal structure of *Arabidopsis* AtCAD5 (SMTL ID: 2cf5.1.B). The nucleotide-binding and catalytic domains of the bottom subunit are colored in green and red, respectively; while their counterparts in the upper subunit are shown in dark green and dark red, respectively. Comparison of the substrate binding pockets of EgCAD1 (**c**) and EgCAD2 (**d**). Difference in the side chain of amino acid residues that forming the substrate binding pocket is marked with a circle

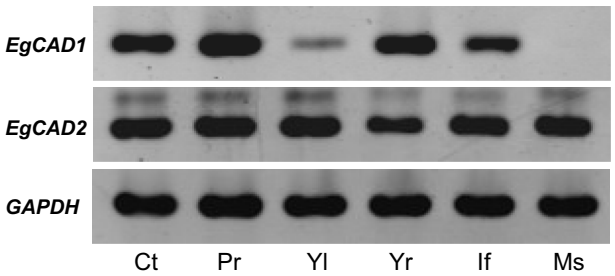

Fig. 4 Expression profiles of *EgCAD1* and *EgCAD2* genes in various organs of oil palm. Agarose gel electrophoresis of the RT-PCR products amplified from the cDNA samples of coleoptile (Ct) and primary root (Pr) of germinating seed, young leaf (Yl) and young root (Yr) of 1-year-old palm, immature fruitlet (If) and mesocarp tissue (Ms) of developing fruit. Refer Online Resource 4 for more information on the samples studied. *GAPDH* was used as the internal control for the gene expression analysis

The regulatory regions of *EgCAD1* and *EgCAD2* contain lignin and defence-related *cis*-regulatory elements

To obtain more clues for predicting the functions of the genes, we cloned and analysed the 5′ flanking sequences of *EgCAD1* and *EgCAD2*. The 5′ flanking sequences of both genes (approximately 1.5 kb upstream of the ORF) were obtained and deposited in GenBank under the accession numbers OK539817 and OK539818 for *EgCAD1* and *EgCAD2*, respectively. The transcription start sites (TSS) of both genes were determined empirically by 5′ RACE in this study. The TSS of *EgCAD1* was an adenine nucleotide (Online Resource 5) located 41 bp upstream of the translation initiation site (ATG), revealing the *EgCAD1* gene possesses a 41 bp 5′ untranslated region (UTR). The TSS denotes the +1 position of the *EgCAD1* transcript and serves as a reference point to determine the positions of the *cis*-acting elements in the 5′ flanking region. The sequencing result showed five different TSSs in *EgCAD2* (Fig. 5 and Online Resource 5), indicating the occurrence of alternative transcription initiation (ATI) in *EgCAD2*. The TSSs identified correspond to the adenine (A), or guanine (G) nucleotides positioned 273 bp (A), 262 bp (G), 213 bp (G), 205 bp (G) and 79 bp (A) upstream of the translation initiation site (ATG), resulting in the formation of different lengths of 5′ UTR. Despite varying lengths, the 5′ UTR of *EgCAD2* is longer than that of *EgCAD1*. To streamline the characterization of the 5′ flanking region, the position of the most upstream TSS, i.e., the adenine nucleotide located 273 bp upstream of the translation initiation site of *EgCAD2*, was regarded as +1 and referred to as distal TSS hereafter (Fig. 5).

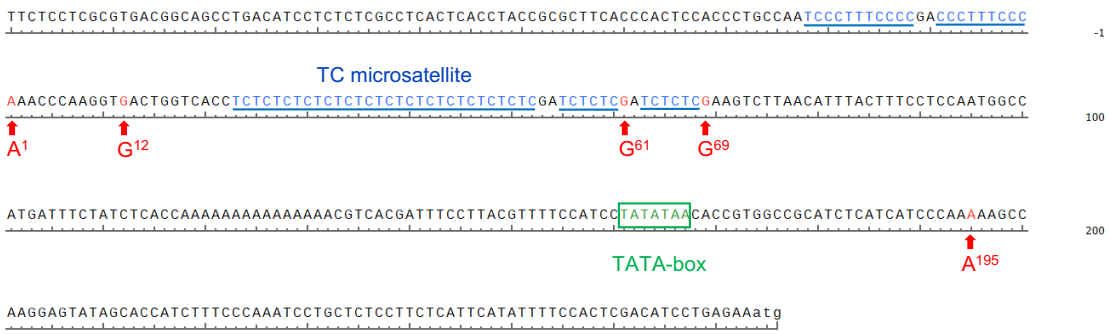

Fig. 5 Organization of the DNA elements that potentially involved in transcription initiation of *EgCAD2*. All of the TSSs are indicated with an arrow under the nucleotides. The most upstream transcription start site (TSS), i.e. the A^1 nucleotide is given number + 1. The nucleotides that constituting the TATA-box are boxed. While the TC microsatellite and TC rich sequences are underlined. The nucleotides in small letters (atg) represent the translation initiation site of *EgCAD2*

Sequence analysis further revealed that *EgCAD1* contains a TATA-box with the sequence "TATAAAT" located at position −30. In the 5′ flanking region of *EgCAD2*, we only found a TATA-box (TATATAA) at position 161 (Fig. 5), which is 34 nucleotides upstream of the most downstream TSS (A^{195}) of *EgCAD2*. This suggests that the other alternative TSSs (A^1, G^{12}, G^{61} and G^{69}) might employ other core promoter elements to initiate gene transcription like other TATA-less promoters. Similar to the sorghum *vp1* promoter (Carrari et al. 2001), which is a TATA-less promoter, a microsatellite consisting of 15 CT repeats (or TC microsatellite) was found upstream of the G^{61} and G^{69} TSSs. We also noticed two 6-base-long TC-rich motifs (TCTCTC) situated immediately upstream of the G^{61} and G^{69}. In addition, another two TC-rich regions (TCCCTTTCCCC and CCCTTTCCC) spanning positions −22 to −12 and −9 to −1 were present upstream of the distal TSS (A^1) (Fig. 5).

Comparison of the 5′ flanking sequences revealed that *EgCAD1* and *EgCAD2* shared similar *cis*-regulatory elements in their regulatory regions. Details of the *cis*-regulatory elements identified in the regulatory regions of *EgCAD1* and *EgCAD2* are listed in Table 2. These *cis*-regulatory elements can be divided into two groups based on their functions. The first group of regulatory elements encompassed the Skn-1_motif and AC elements involved in plant development. In consonance with the role of *CAD* in lignin biosynthesis, both of the *bona fide EgCAD* genes contain AC elements in their regulatory regions. It is worth noting that the AC elements serve as the binding sites for MYB transcription factors that regulate lignin biosynthesis in plant. Interestingly, *EgCAD1* possesses only one AC element (AC-II) in its regulatory region, but *EgCAD2* contains two types of AC elements (AC-I and AC-III) located adjacent to the TSS (Fig. 6).

The second group consists of the regulatory elements responsive to stresses and phytohormones. Several defence-related motifs, such as the CGTCA-motif, TGACG-motif, W-box and GT-1 motif were found in the regulatory regions of *EgCAD1* and *EgCAD2*, implying the genes play a role in plant defence. Furthermore, the two *bona fide EgCAD* genes also possess the ACGTATERD1 motif responsive to water stress. Additionally, sequence analysis revealed that the TATC-box motif, which is involved in gibberellin responsiveness, is only present in the regulatory region of *EgCAD1,* but not *EgCAD2*; this might be a factor contributing to the tissue-selective expression of *EgCAD1* in oil palm. Overall, the results indicate that *EgCAD1* and *EgCAD2* are likely involved in oil palm development and adaption to stresses.

Discussion

Oil palm is one of the important oil crops cultivated in many countries for edible oil production. Besides, the oil palm biomass is used as the feedstock to produce biofuels (Mahlia et al. 2019). Thus, oil palm plantation promotes international trade and contributes to global economic growth. Unfortunately, the growth of oil palm is greatly impaired by the basal stem rot disease and environmental stresses, leading to reduced palm oil production (Paterson 2019; Gorea et al. 2020). Lignin has been known to serve important functions in plant development and defence. Many studies showed that it is possible to improve plant growth and disease resistance by enhancing lignin accumulation in plant cell walls (Mutuku et al. 2019; Zhang et al. 2019; Tang et al. 2019). However, this strategy may be impracticable or unfeasible in oil palm as far as oil palm biomass for downstream processing activities is

Table 2 *Cis*-regulatory elements identified in the regulatory regions of *EgCAD1* and *EgCAD2*

No	*cis*-acting elements	Motif	Function	Source	*EgCAD1*	*EgCAD2*
1	-10PEHVPSBD	TATTCT	Activated by blue, white or UV-A light	Thum et al. (2001)	X	X
2	AC-I	ACCTACC	Xylem-specific expression of lignin biosynthetic genes	Zhong and Ye (2009), Hatton et al. (1995)		X
3	AC-II	ACCAACC	Xylem-specific expression of lignin biosynthetic genes	Hatton et al. (1995), Zhong and Ye (2009)	X	
4	AC-III	ACCTAAC	Xylem-specific expression of lignin biosynthetic genes	Zhong and Ye (2009), Hatton et al. (1995)		X
5	ACGTATERD1	ACGT	Required for etiolation-induced expression of erd1 (early responsive to dehydration) in Arabidopsis	Simpson et al. (2003)	X	X
6	Box-4	ATTAAT	Part of a module for light response	PlantCARE	X	X
7	CGTCA-motif	CGTCA	Involved in the MeJA-responsiveness	Rouster et al., (1997), Wang et al. (2010)	X	X
8	CURECORECR	GTAC	Copper-response element	Kropat et al. (2005)	X	X
9	GATABOX	GATA	Required for high level, light regulated, and tissue specific expression	Rubio-Somoza et al. (2006)	X	X
10	Skn-1_motif	GTCAT	Required for endosperm expression	PlantCARE	X	X
11	TATC-box	TATCCCA	Involved in gibberellin responsiveness	Gubler and Jacobsen (1992)	X	
12	TGACG-motif	TGACG	Binding site for bZIP transactivating factors, involved in the MeJA-responsiveness	Rouster et al. (1997)	X	X
13	W-box	TTGACC	Binding site for members of the WRKY family of transcription factors, responsible for the pathogen inducibility Recognized specifically by salicylic acid (SA)-induced WRKY DNA binding proteins	Rushton et al. (1996) Xu et al. (2006)	X	X
14	GT-1 motif	GAAAAA	Required for respond towards pathogen and NaCl-induction	Park et al. (2004)	X	X

"X" indicates the presence of regulatory element

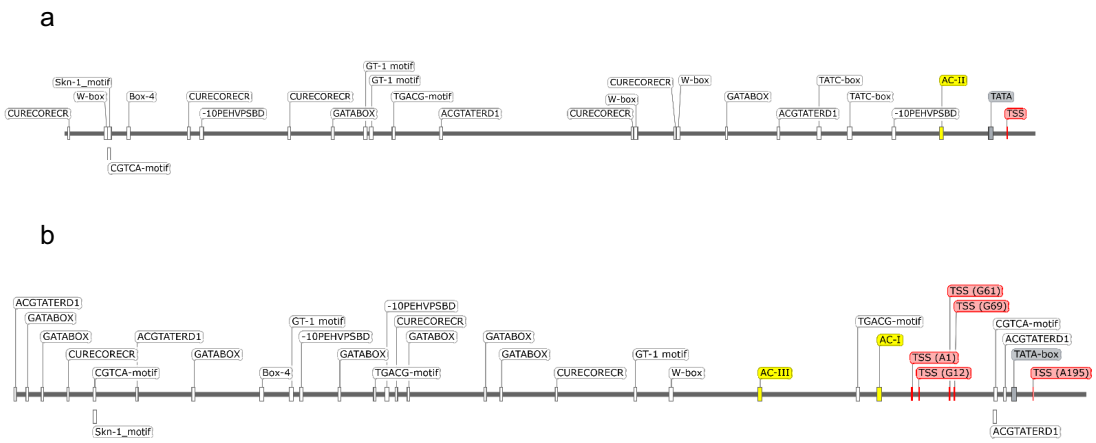

Fig. 6 Distribution of *cis*-elements in the 5′ regulatory regions of *EgCAD1* (**a**) and *EgCAD2* (**b**). Functions of the *cis*-elements were summarized in Table 2

concerned, as lignin impedes the enzymatic saccharification of oil palm biomass during biofuel production, resulting in additional cost incurred for the removal of lignin from lignocellulosic materials (Lockhart 2015; Ladeira Ázar et al. 2020). In addition, modulation of lignin biosynthesis in oil palm requires identification of lignin biosynthetic genes in oil palm and an in-depth understanding of the regulation of the target gene.

As CAD controls the final step of the biosynthesis of monolignols (monomers of lignin), it has been a target for alteration of lignin content in plants. Previous studies showed that perturbation of *CAD* in switchgrass and *Brachypodium* reduced the lignin content of transgenic plants without compromising the plant growth and resulted in improved saccharification efficiency (Saathoff et al. 2011a; Fu et al. 2011; Bouvier d'Yvoire et al. 2013). However, some studies demonstrated that suppression of *CAD* altered the lignin composition but not the lignin content of transgenic plants (Baucher et al. 1999; Fornalé et al. 2012; Trabucco et al. 2013). Since the CADs from higher plants are encoded by a multigene family, and they are divided into different groups, it has been proposed that a specific group of *CAD* genes is associated with a specific function in plant. For example, the group I *CAD* is well recognized as *bona fide CAD* responsible for lignin biosynthesis (Ma 2010; Hirano et al. 2012), while the group IV *CAD* was shown to be associated with defence (Rong et al. 2016). Therefore, selecting the right *CAD* gene to manipulate is critical to accomplish the desired effects.

There are two *bona fide CAD* genes in oil palm

Among the higher plants, the *CAD* gene family of *Arabidopsis thaliana* has been extensively studied. Out of the nine putative *AtCAD* genes, *AtCAD4* and *AtCAD5* had been identified as *bona fide CAD* genes, which played a major role in lignin biosynthesis in *Arabidopsis* (Sibout et al. 2005, 2003). Perturbation of these genes affected the lignin content and composition in the plant and led to higher disease susceptibility (Tronchet et al. 2010; Sibout et al. 2005, 2003). It is also worth noting that *AtCAD1* also takes part in lignin biosynthesis in the elongating stems of *Arabidopsis* (Eudes et al. 2006). A study of the AtCAD4 and AtCAD5 proteins revealed that the latter possesses higher catalytic capacity and used sinapyl aldehyde as its substrate more effectively compared to the former (Kim et al. 2004). To better characterise the AtCAD5, Youn et al. (2006) also studied its crystal structure and determined the amino acid residues involved in its catalytic activity. Owing to the detailed study of the members of the *CAD* gene family from *Arabidopsis*, *AtCAD*s always serve as a reference for the characterization of the *CAD* gene family in other species.

In the present study, we have identified seven putative *EgCAD* genes (*EgCAD*1-7) in the oil palm genome (Table 1). We also found another six loci annotated as *CAD* in oil palm genome (data not shown), but they were considered as pseudogenes. Previously, Singh et al. (2013) reported that abundant homologous duplicated sequences are present in the 16 chromosome pairs of oil palm due to segmental duplications. Therefore, the presence of multiple *EgCAD* genes in the oil palm genome is probably due to the segmental duplications of chromosome arms. Apart from segmental duplication, whole-genome duplication also occurred in many angiosperms during evolution, producing multiple duplicate genes in their genomes (Qiao et al. 2019; Ren et al. 2018). Among the putative *EgCAD* genes identified, *EgCAD1* and *EgCAD2* are suggested to be the *bona fide CAD* genes of oil palm judging by the features exhibited, particularly the high similarity between them and the *bona fide CAD* genes from other higher plant species. In addition, our phylogenetic analysis showed that the EgCAD1 and EgCAD2 are clustered together with the *bona fide* CADs, which were determined to play major roles in lignin biosynthesis in other plants (Fig. 1). To date, all of the studied higher plants possess one or two copies of *bona fide CAD*s. For example, a single copy of *bona fide CAD* is found in the genomes of *Brachypodium distachyon*, *Liriodendron tulipifera* and *Populus tomentosa* (Bukh et al. 2012; Xu et al. 2013; Chao et al. 2014), while there are two *bona fide CAD*s in the genomes of *Arabidopsis thaliana*, *Panicum virgatum* and *Cucumis melo* (Jin et al. 2014; Kim et al. 2004; Saathoff et al. 2011b).

EgCAD2 is accountable for the biosynthesis of lignin throughout plant development

Gene expression analysis revealed that the two *bona fide EgCAD* genes exhibited different expression patterns in the tissues studied (Fig. 4), indicating that the two genes might play different roles in oil palm. *EgCAD1* expressed preferentially in the root tissues under normal growth conditions. A previous study showed that the expression of *EgCAD1* elevated in oil palm seedlings upon pathogen infection (Goh et al. 2018). It is most likely that the induced gene expression after pathogen infection is associated with the biosynthesis of cellular components required for plant defence. Previously, Rong et al. (2016) showed that the expression of *TaCAD12* also elevated in wheat plants in response to *Rhizoctonia cerealis* infection, and the role of *TaCAD12* in disease resistance had been determined through functional analyses. Hence, we suggest that *EgCAD1* plays a role in plant defence, particularly in pathogen-induced lignin biosynthesis for plant defence.

Contrary to *EgCAD1*, *EgCAD2* was expressed constitutively in all the tissues studied, indicating that *EgCAD2* plays a predominant role in lignin biosynthesis during the growth and development of oil palm. Undoubtedly, this might be attributed to the presence of multiple TSSs and AC elements in the 5′ flanking region of *EgCAD2* (Figs. 4, 5). The AC element is required to direct the expression of lignin biosynthetic genes in lignified tissues (Zhong and Ye 2009). Previously, Bukh et al. (2012) reported

that *BdCAD5*, a *bona fide CAD*, displayed a constitutive expression pattern in *Brachypodium distachyon*. In rice, the *OsCAD2* responsible for lignin biosynthesis is also expressed constitutively throughout all stages of plant development (Park et al. 2018; Martin et al. 2019). Since lignin is indispensable during plant development, the constitutive expression of *EgCAD2* is most probably accountable for the biosynthesis of lignin throughout plant development. Therefore, we suggest that *EgCAD2* is a *bona fide CAD* that acts as the key player in lignin biosynthesis in oil palm and mainly contributes to the production of monolignols required for the growth and development of oil palm.

Lignin and defence related *cis*-regulatory elements control the expression of *EgCAD1* and *EgCAD2*

The expression behaviours of *EgCAD1* and *EgCAD2* reflect the different manners the genes are regulated in oil palm. Since gene expression is largely regulated by its promoter and *cis*-regulatory elements, the cloning and analysis of the 5′ flanking sequences of *EgCAD1* and *EgCAD2* would provide essential information for predicting their promoter activities. As anticipated, AC elements were found to be present in the 5′ flanking sequences of both genes. It is worth noting that AC elements serve as the binding sites for the MYB transcription factors, which activate the transcriptions of lignin biosynthetic genes (Zhou et al. 2009). In addition, studies also showed that AC elements were required to direct the expression of a gene in vascular tissues (Séguin et al. 1997; Hatton et al. 1995). Raes et al. (2003) reported that AC elements are present in the regulatory region of most lignin biosynthetic genes in *Arabidopsis*. Our previous works also showed the presence of AC elements in the regulatory regions of *EgPAL1* and *Eg4CL1* genes from oil palm, another two important genes in the phenylpropanoid pathway (Yusuf et al. 2018a, b). Hence, the presence of AC elements in the regulatory regions of *EgCAD1* and *EgCAD2* further implies that these genes are *bona fide CAD* genes involved in lignin biosynthesis.

Apart from the *cis*-acting elements mentioned above, we also found several defence-related motifs, including W-box, GT-1 motif and CGTCA-motif in the 5′ flanking regions of *EgCAD1* and *EgCAD2* (Table 2). It is well known that defence-related transcription factors such as WRKY and trihelix are usually produced in plants upon pathogen infection (Liu et al. 2021b; Bi et al. 2021; Wang et al. 2016). The WRKY transcription factor binds to the W-box in the promoter of many defence-related genes, enhancing the gene transcription for defence purposes (Liu et al. 2019; Yang et al. 2020), while the trihelix transcription factor binds to the GT-1 motif to activate the defence mechanism (Park et al. 2004; Xu et al. 2018). Defence against pathogens often involves the activation of the jasmonic acid signalling pathway in plants (Zhang et al. 2018; Yang et al. 2019). Methyl jasmonate (MeJA), a derivative of jasmonic acid, has been shown to induce the expression of defence genes in many plants (Benevenuto et al. 2019; Repka et al. 2004). The CGTCA-motif responsive to MeJA is widely distributed in the promoter regions of many pathogenesis-related genes (Hussain et al. 2018; Jiang et al. 2014; Kaur et al. 2017; Fang et al. 2019). Hence, the above-mentioned defence-related motifs in the 5′ flanking regions of *EgCAD1* and *EgCAD2* imply that the genes are associated with plant defence. This agrees with the induced *EgCAD1* gene expression observed by Goh et al. (2018) in oil palm seedlings infected by *Ganoderma boninense*. Together, these observations lead us to propose that both *bona fide EgCAD* genes are responsive to pathogen infection.

Apart from biotic stress, we also anticipate that the expression of *EgCAD1* might be inducible by gibberellin due to the presence of TATC-box (gibberellin responsiveness element) in the 5′ flanking region of *EgCAD1* (Table 2). Previously, other studies also reported the presence of TATC-box in the promoter regions of lignin biosynthetic genes in other plants (Jin et al. 2014; Sun et al. 2020; Li et al. 2020). It is well known that gibberellin acts as a regulator of plant growth and development (Wang et al. 2015; Binenbaum et al. 2018; Hedden and Sponsel 2015). The presence of TATC-box in the 5′ flanking region of *EgCAD1* hints that there might be crosstalk between gibberellin and *EgCAD1* in controlling oil palm development. Perhaps the endogenous gibberellin levels in a particular oil palm tissue could affect the lignin content of the tissue by altering the expression level of *EgCAD1*. A recent study by Falcioni et al. (2018) demonstrated that the concentration of gibberellin positively affected the total lignin content of the tobacco plant. Recently, Wang et al. (2020) reported increased expressions of lignin biosynthetic genes and enhanced lignin accumulation in carrot tissues treated with exogenous gibberellin. In roots, gibberellin is biosynthesized and accumulated exclusively in the endodermal cells of the root elongation zone (Barker et al. 2021; Shani et al. 2013). Considering the root-selective expression of *EgCAD1*, it is most likely that the gene is responsible for the biosynthesis of monolignols required for cell wall construction during root development and elongation in response to gibberellin. Judging by the expression patterns and the *cis*-acting elements in the regulatory regions of *EgCAD1* and *EgCAD2*, it is suggested that the two genes play different but overlapping roles in oil palm during plant growth and in response to various abiotic and biotic stresses. The potential roles of *EgCAD1* and *EgCAD2* genes in oil palm were summarized in Fig. 7.

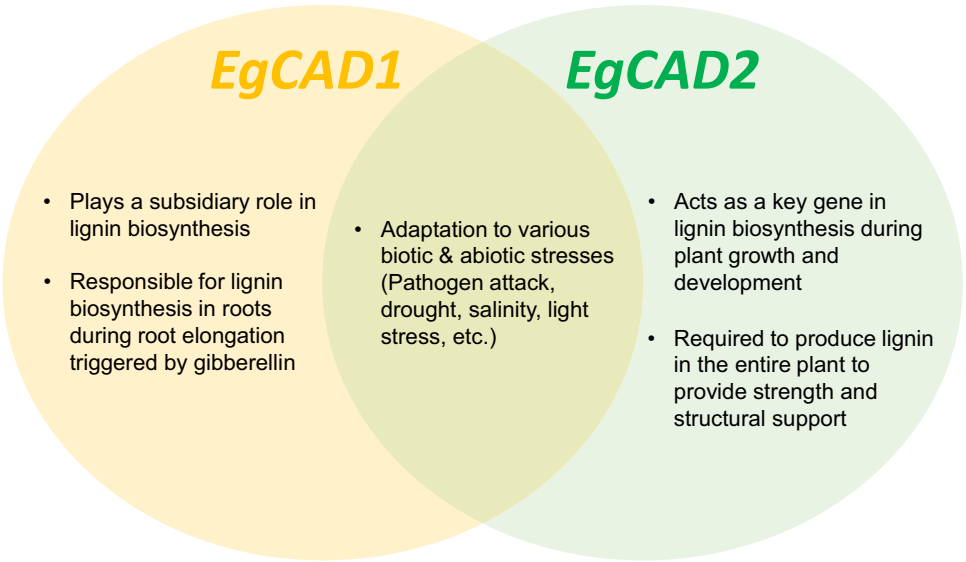

Fig. 7 Potential roles of *EgCAD1* and *EgCAD2* genes in oil palm judging by their expression patterns and the *cis*-acting elements in their respective regulatory regions

Alternative transcription initiation occurs in *EgCAD2*

Identification of TSS provides a clear picture of the structure and the regulatory region of a gene. In addition to the Cap Analysis of Gene Expression (CAGE) method and its derivatives, the 5′ Rapid Amplification of cDNA Ends (RACE) through template switching approach is widely used in the determination of gene TSS nowadays (Liu et al. 2018a). Through 5′ RACE, we discovered that the *EgCAD2* gene possesses multiple TSSs, allowing alternative transcription initiation (ATI) to take place. Consequently, *EgCAD2* transcripts with varying lengths of 5′ UTR were produced from the same gene locus. Previously, ATI has been reported in *Arabidopsis* (Garcia and Sanchez-Puerta 2021). Apart from plants, ATI is also found in protozoans, animals and humans (Markus et al. 2021; Nepal et al. 2020; Anvar et al. 2018). In certain cases, ATI altered the translational activities of mRNAs derived from a gene and affected the protein yields (Rojas-Duran and Gilbert 2012). In the gene with multiple TSSs, the selection of TSS during transcription varied at different developmental stages (Zhang et al. 2017). It can also be influenced by external factors such as light and carbon sources (Kurihara et al. 2018; Inoue et al. 2020). In light of the previous findings of ATI, we speculated that the presence of alternative TSS would allow the regulation of *EgCAD2* at the translation level by using different TSSs for transcription initiation at different developmental stages or tissues and in response to external factors. This further implies the importance of *EgCAD2* in oil palm for plant development and adaptation.

Despite many studies on ATI, the role of ATI is still unclear and remains controversial. Previously, studies showed that ATI was important in the transcriptional regulation of the enolase-encoding gene in *Aspergillus oryzae* under different environmental conditions (Inoue et al. 2020). Besides, Persson et al. (2016) reported that ATI regulates the retrotransposon activity in the genome. However, Xu et al. (2019) suggested that ATI occurs primarily due to molecular errors. Therefore, a comprehensive study is required before concluding the involvement and the importance of ATI in regulating the *EgCAD2* activity in oil palm.

It is well known that a TATA-box is required to initiate gene transcription. However, most of the promoters in plants are TATA-less (Molina and Grotewold 2005; Francki et al. 2009). Although there are five TSSs in the 5′ flanking region of *EgCAD2*, there is only a single TATA-box being identified (Fig. 5). Judging by its position, this TATA-box is expected to participate in the gene transcription when the most downstream TSS (A^{195}) is used. Meaning that, other types of core promoter elements take over the role of TATA-box when other TSSs are utilized to initiate transcription. The presence of TC microsatellite and TC-rich sequences in the 5′ flanking region of *EgCAD2* (Fig. 5) prompts us to speculate that these TC-rich sequences might play a role like TATA-box in the initiation of transcription when those TSSs that lack a TATA-box at their 5′ upstream regions are used during transcription. This agrees with the findings reported by several previous studies (Bernard et al. 2010; Francki et al. 2009; Yamamoto et al. 2007; Zuo and Li 2011; Tokizawa et al. 2017).

Importance of this study

Manipulation of lignin content in plant has been an approach employed by many researchers to enhance disease resistance, improve wood quality and facilitate the conversion of lignocellulosic biomass to biodiesel. To

manipulate the lignin content in oil palm, the target genes must be first identified. Previously, our team has reported the studies of two important lignin biosynthetic genes in oil palm, i.e., *EgPAL* and *Eg4CL* genes (Yusuf et al. 2018a, b). In this study, the *bona fide EgCAD* gene candidates were identified in oil palm, and their potential roles were proposed. The study of lignin biosynthetic genes in oil palm is important because it is a prerequisite for manipulating lignin biosynthesis and lignin content in oil palm. Identification of the molecular switch of lignification in oil palm would allow one to manipulate the lignin content either in a specific organ of a plant or throughout the plant. Therefore, it would be possible to improve the plant defence by increasing the lignin content in oil palm roots and reducing the lignin content in fruit bunch and leaf to improve the efficiency of biofuel production concurrently. Apart from that, identifying *bona fide EgCAD* genes in oil palm would also benefit genetic improvement and plant breeding activities, particularly the development of new oil palm varieties with modified lignin content and composition through marker-assisted breeding or genetic engineering.

Despite comprehensive information that has been obtained through in silico sequence analysis and expression data in this study, further functional analysis, a gene knockout experiment in particular, is still required to validate the roles of *EgCAD* genes in oil palm. However, producing transgenic oil palm is a great challenge as the plant showed unsatisfactory transformation and regeneration efficiencies under in vitro conditions. The long regeneration time of oil palm explant further aggravates the problem. Hence, this study sheds some light on the roles of the *bona fide CAD* genes in oil palm. The findings of this study will be valuable for the study and improvement of oil palm lignin content in future. Study of the promoter will provide valuable information on the regulations of the genes and further reveal the involvement of these genes in lignin biosynthesis in oil palm.

Conclusions

In conclusion, seven *EgCAD* genes were identified in the oil palm genome. However, we cannot rule out the possibility of the existence of other unidentified *EgCAD* genes in the oil palm genome, especially within the regions not covered by the sequencing project. Our results indicated that *EgCAD1* and *EgCAD2* are the *bona fide CAD* gene candidates in oil palm. Albeit both *EgCAD1* and *EgCAD2* seem to be involved in lignin biosynthesis, we suggest that the two genes play different, but overlapping roles in oil palm during plant growth and in response to various abiotic and biotic stresses. Apparently, *EgCAD2* is the key gene in lignin biosynthesis for growth and development, while *EgCAD1* plays a subsidiary role in lignin biosynthesis for growth and development.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s13205-022-03208-0.

Acknowledgements Assistance in sample collection by the officers from GanoDROP Unit, MPOB is highly appreciated.

Author contribution MPA and IAS conceived and designed the research. MPA advised on the technical aspects and contributed research equipment. CYLY, NSN and NAAMT conducted the research. CYLY and MPA analysed the results and prepared the manuscript. All authors read and approved the manuscript.

Funding This study was supported by Fundamental Research Grant Scheme (Reference Code: FRGS/1/2019/STG05/UITM/02/1) from the Malaysian Ministry of Higher Education (MOHE).

Data availability The gene and promoter sequences of *EgCAD1* and *EgCAD2* have been submitted to NCBI as mentioned in the result section. The sequences will be opened to the public after the manuscript is published.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This study does not involve any animal or human.

References

Anvar SY, Allard G, Tseng E, Sheynkman GM, de Klerk E, Vermaat M, Yin RH, Johansson HE, Ariyurek Y, den Dunnen JT, Turner SW, 't Hoen PAC (2018) Full-length mRNA sequencing uncovers a widespread coupling between transcription initiation and mRNA processing. Genome Biol 19(1):46. https://doi.org/10.1186/s13059-018-1418-0

Barker R, Fernandez Garcia MN, Powers SJ, Vaughan S, Bennett MJ, Phillips AL, Thomas SG, Hedden P (2021) Mapping sites of gibberellin biosynthesis in the Arabidopsis root tip. New Phytol 229(3):1521–1534. https://doi.org/10.1111/nph.16967

Baucher M, Bernard-vailhé MA, Chabbert B, Besle J-M, Opsomer C, Van Montagu M, Botterman J (1999) Down-regulation of cinnamyl alcohol dehydrogenase in transgenic alfalfa (*Medicago sativa* L.) and the effect on lignin composition and digestibility. Plant Mol Biol 39(3):437–447. https://doi.org/10.1023/A:1006182925584.

Benevenuto RF, Seldal T, Hegland SJ, Rodriguez-Saona C, Kawash J, Polashock J (2019) Transcriptional profiling of methyl jasmonate-induced defense responses in bilberry (*Vaccinium myrtillus* L.). BMC Plant Biol 19(1):70. https://doi.org/10.1186/s12870-019-1650-0

Bernard V, Brunaud V, Lecharny A (2010) TC-motifs at the TATA-box expected position in plant genes: a novel class of motifs involved

in the transcription regulation. BMC Genomics 11(1):166. https://doi.org/10.1186/1471-2164-11-166
Bewg WP, Poovaiah C, Lan W, Ralph J, Coleman HD (2016) RNAi downregulation of three key lignin genes in sugarcane improves glucose release without reduction in sugar production. Biotechnol Biofuels 9(1):270. https://doi.org/10.1186/s13068-016-0683-y
Bi M, Li X, Yan X, Liu D, Gao G, Zhu P, Mao H (2021) Chrysanthemum WRKY15–1 promotes resistance to Puccinia horiana Henn. via the salicylic acid signaling pathway. Horticult Res 8(1):6–6. https://doi.org/10.1038/s41438-020-00436-4
Binenbaum J, Weinstain R, Shani E (2018) Gibberellin localization and transport in plants. Trends Plant Sci 23(5):410–421. https://doi.org/10.1016/j.tplants.2018.02.005
Bouvier d'Yvoire M, Bouchabke-Coussa O, Voorend W, Antelme S, Cézard L, Legée F, Lebris P, Legay S, Whitehead C, McQueen-Mason SJ, Gomez LD, Jouanin L, Lapierre C, Sibout R (2013) Disrupting the *cinnamyl alcohol dehydrogenase 1* gene (*BdCAD1*) leads to altered lignification and improved saccharification in *Brachypodium distachyon*. Plant J 73(3):496–508. https://doi.org/10.1111/tpj.12053
Bryant ND, Pu Y, Tschaplinski TJ, Tuskan GA, Muchero W, Kalluri UC, Yoo CG, Ragauskas AJ (2020) Transgenic poplar designed for biofuels. Trends Plant Sci 25(9):881–896. https://doi.org/10.1016/j.tplants.2020.03.008
Bukh C, Nord-Larsen PH, Rasmussen SK (2012) Phylogeny and structure of the cinnamyl alcohol dehydrogenase gene family in *Brachypodium distachyon*. J Exp Bot 63(17):6223–6236. https://doi.org/10.1093/jxb/ers275
Carrari F, Frankel N, Lijavetzky D, Benech-Arnold R, Sánchez R, Iusem ND (2001) The tata-less promoter of VP1, a plant gene controlling seed germination. DNA Sequence 12(2):107–114. https://doi.org/10.3109/10425170109047563
Chanoca A, de Vries L, Boerjan W (2019) Lignin engineering in forest trees. Front Plant Sci. https://doi.org/10.3389/fpls.2019.00912
Chao N, Liu S-X, Liu B-M, Li N, Jiang X-N, Gai Y (2014) Molecular cloning and functional analysis of nine cinnamyl alcohol dehydrogenase family members in *Populus tomentosa*. Planta 240(5):1097–1112. https://doi.org/10.1007/s00425-014-2128-9
Chen C, Baucher M, Holst Christensen J, Boerjan W (2001) Biotechnology in trees: towards improved paper pulping by lignin engineering. Euphytica 118(2):185. https://doi.org/10.1023/A:1004176714883
Chen F, Tobimatsu Y, Havkin-Frenkel D, Dixon RA, Ralph J (2012a) A polymer of caffeyl alcohol in plant seeds. Proc Natl Acad Sci USA 109(5):1772–1777. https://doi.org/10.1073/pnas.1120992109
Chen W, VanOpdorp N, Fitzl D, Tewari J, Friedemann P, Greene T, Thompson S, Kumpatla S, Zheng P (2012b) Transposon insertion in a cinnamyl alcohol dehydrogenase gene is responsible for a brown midrib1 mutation in maize. Plant Mol Biol 80(3):289–297. https://doi.org/10.1007/s11103-012-9948-4
Chen F, Tobimatsu Y, Jackson L, Nakashima J, Ralph J, Dixon RA (2013) Novel seed coat lignins in the Cactaceae: structure, distribution and implications for the evolution of lignin diversity. Plant J 73(2):201–211. https://doi.org/10.1111/tpj.12012
Cheng H, Li L, Xu F, Cheng S, Cao F, Wang Y, Yuan H, Jiang D, Wu C (2013) Expression patterns of a cinnamyl alcohol dehydrogenase gene involved in lignin biosynthesis and environmental stress in *Ginkgo biloba*. Mol Biol Rep 40(1):707–721. https://doi.org/10.1007/s11033-012-2111-0
del Río JC, Rencoret J, Gutiérrez A, Elder T, Kim H, Ralph J (2020) Lignin monomers from beyond the canonical monolignol biosynthetic pathway: another brick in the wall. ACS Sustain Chem Eng 8(13):4997–5012. https://doi.org/10.1021/acssuschemeng.0c01109
Deng W-W, Zhang M, Wu J-Q, Jiang Z-Z, Tang L, Li Y-Y, Wei C-L, Jiang C-J, Wan X-C (2013) Molecular cloning, functional analysis of three cinnamyl alcohol dehydrogenase (CAD) genes in the leaves of tea plant, *Camellia sinensis*. J Plant Physiol 170(3):272–282. https://doi.org/10.1016/j.jplph.2012.10.010
Eudes A, Pollet B, Sibout R, Do CT, Séguin A, Lapierre C, Jouanin L (2006) Evidence for a role of AtCAD 1 in lignification of elongating stems of *Arabidopsis thaliana*. Planta 225(1):23–39. https://doi.org/10.1007/s00425-006-0326-9
Falcioni R, Moriwaki T, de Oliveira DM, Andreotti GC, de Souza LA, dos Santos WD, Bonato CM, Antunes WC (2018) Increased gibberellins and light levels promotes cell wall thickness and enhance lignin deposition in xylem fibers. Front Plant Sci. https://doi.org/10.3389/fpls.2018.01391
Fang L-J, Qin R-L, Liu Z, Liu C-R, Gai Y-P, Ji X-L (2019) Expression and functional analysis of a PR-1 Gene, MuPR1, involved in disease resistance response in mulberry (*Morus multicaulis*). J Plant Interact 14(1):376–385. https://doi.org/10.1080/17429145.2019.1640295
Fornalé S, Capellades M, Encina A, Wang K, Irar S, Lapierre C, Ruel K, Joseleau J-P, Berenguer J, Puigdomènech P, Rigau J, Caparrós-Ruiz D (2012) Altered lignin biosynthesis improves cellulosic bioethanol production in transgenic maize plants down-regulated for cinnamyl alcohol dehydrogenase. Mol Plant 5(4):817–830. https://doi.org/10.1093/mp/ssr097
Francki M, Cíváň PC, Švec MŠ (2009) Genome-wide analysis of rice (*Oryza sativa* L. subsp. japonica) TATA box and Y Patch promoter elements. Genome 52(3):294–297
Fu C, Xiao X, Xi Y, Ge Y, Chen F, Bouton J, Dixon RA, Wang Z-Y (2011) Downregulation of cinnamyl alcohol dehydrogenase (CAD) leads to improved saccharification efficiency in switchgrass. BioEnergy Res 4(3):153–164. https://doi.org/10.1007/s12155-010-9109-z
Garcia LE, Sanchez-Puerta MV (2021) Transcriptional landscape and splicing efficiency in *Arabidopsis* mitochondria. Cells 10(8):2054
Goh KM, Dickinson M, Supramaniam CV (2018) Morphological and transcript changes in the biosynthesis of lignin in oil palm (Elaeis guineensis) during Ganoderma boninense infections in vitro. Physiol Plant 162(3):274–289. https://doi.org/10.1111/ppl.12645
Gorea EA, Godwin ID, Mudge AM (2020) Ganoderma infection of oil palm – a persistent problem in Papua New Guinea and Solomon Islands. Australas Plant Pathol 49(1):69–77. https://doi.org/10.1007/s13313-019-00673-9
Gubler F, Jacobsen JV (1992) Gibberellin-responsive elements in the promoter of a barley high-pI alpha-amylase gene. Plant Cell 4(11):1435–1441
Hatton D, Sablowski R, Yung M-H, Smith C, Schuch W, Bevan M (1995) Two classes of *cis* sequences contribute to tissue-specific expression of a *PAL2* promoter in transgenic tobacco. Plant J 7(6):859–876. https://doi.org/10.1046/j.1365-313X.1995.07060859.x
Hedden P, Sponsel V (2015) A century of Gibberellin research. J Plant Growth Regul 34(4):740–760. https://doi.org/10.1007/s00344-015-9546-1
Hirano K, Aya K, Kondo M, Okuno A, Morinaka Y, Matsuoka M (2012) *OsCAD2* is the major *CAD* gene responsible for monolignol biosynthesis in rice culm. Plant Cell Rep 31(1):91–101. https://doi.org/10.1007/s00299-011-1142-7
Hodgson-Kratky K, Papa G, Rodriguez A, Stavila V, Simmons B, Botha F, Furtado A, Henry R (2019) Relationship between sugarcane culm and leaf biomass composition and saccharification efficiency. Biotechnol Biofuels 12(1):247. https://doi.org/10.1186/s13068-019-1588-3

Hussain RMF, Sheikh AH, Haider I, Quareshy M, Linthorst HJM (2018) *Arabidopsis* WRKY50 and TGA transcription factors synergistically activate expression of PR1. Front Plant Sci 9:930. https://doi.org/10.3389/fpls.2018.00930

Inoue T, Toji H, Tanaka M, Takama M, Hasegawa-Shiro S, Yamaki Y, Shintani T, Gomi K (2020) Alternative transcription start sites of the enolase-encoding gene enoA are stringently used in glycolytic/gluconeogenic conditions in Aspergillus oryzae. Curr Genet 66(4):729–747. https://doi.org/10.1007/s00294-020-01053-3

Jiang Y, Duan Y, Yin J, Ye S, Zhu J, Zhang F, Lu W, Fan D, Luo K (2014) Genome-wide identification and characterization of the Populus WRKY transcription factor family and analysis of their expression in response to biotic and abiotic stresses. J Exp Bot 65(22):6629–6644. https://doi.org/10.1093/jxb/eru381

Jin Y, Zhang C, Liu W, Qi H, Chen H, Cao S (2014) The cinnamyl alcohol dehydrogenase gene family in melon (*Cucumis melo* L.): bioinformatic analysis and expression patterns. PLoS One 9(7):e101730. https://doi.org/10.1371/journal.pone.0101730

Jung HG, Allen MS (1995) Characteristics of plant cell walls affecting intake and digestibility of forages by ruminants. J Anim Sci 73(9):2774–2790. https://doi.org/10.2527/1995.7392774x

Kaur A, Pati PK, Pati AM, Nagpal AK (2017) *In-silico* analysis of cis-acting regulatory elements of pathogenesis-related proteins of *Arabidopsis thaliana* and *Oryza sativa*. PLoS ONE 12(9):e0184523. https://doi.org/10.1371/journal.pone.0184523

Kim Y-H, Huh G-H (2019) Overexpression of cinnamyl alcohol dehydrogenase gene from sweetpotato enhances oxidative stress tolerance in transgenic *Arabidopsis*. In Vitro Cell Dev Biol Plant 55(2):172–179. https://doi.org/10.1007/s11627-018-09951-5

Kim S-J, Kim M-R, Bedgar DL, Moinuddin SGA, Cardenas CL, Davin LB, Kang C, Lewis NG (2004) Functional reclassification of the putative cinnamyl alcohol dehydrogenase multigene family in *Arabidopsis*. Proc Natl Acad Sci USA 101(6):1455–1460. https://doi.org/10.1073/pnas.0307987100

Kim YH, Bae JM, Huh GH (2010) Transcriptional regulation of the cinnamyl alcohol dehydrogenase gene from sweet potato in response to plant developmental stage and environmental stress. Plant Cell Rep 29(7):779–791. https://doi.org/10.1007/s00299-010-0864-2

Kropat J, Tottey S, Birkenbihl RP, Depège N, Huijser P, Merchant S (2005) A regulator of nutritional copper signaling in Chlamydomonas is an SBP domain protein that recognizes the GTAC core of copper response element. Proc Natl Acad Sci USA 102(51):18730–18735. https://doi.org/10.1073/pnas.0507693102

Kurihara Y, Makita Y, Kawashima M, Fujita T, Iwasaki S, Matsui M (2018) Transcripts from downstream alternative transcription start sites evade uORF-mediated inhibition of gene expression in *Arabidopsis*. Proc Natl Acad Sci USA 115(30):7831–7836. https://doi.org/10.1073/pnas.1804971115

Ladeira Ázar RIS, Bordignon-Junior SE, Laufer C, Specht J, Ferrier D, Kim D (2020) Effect of lignin content on cellulolytic saccharification of liquid hot water pretreated sugarcane bagasse. Molecules 25(3):623

Lee M-H, Jeon HS, Kim SH, Chung JH, Roppolo D, Lee H-J, Cho HJ, Tobimatsu Y, Ralph J, Park OK (2019) Lignin-based barrier restricts pathogens to the infection site and confers resistance in plants. EMBO J 38(23):e101948. https://doi.org/10.15252/embj.2019101948

Lescot M, Déhais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouzé P, Rombauts S (2002) PlantCARE, a database of plant *cis*-acting regulatory elements and a portal to tools for *in silico* analysis of promoter sequences. Nucleic Acids Res 30(1):325–327. https://doi.org/10.1093/nar/30.1.325

Li H, Huang Y (2017) Expression of brown-midrib in a spontaneous sorghum mutant is linked to a 5′-UTR deletion in lignin biosynthesis gene SbCAD2. Sci Rep 7(1):11664. https://doi.org/10.1038/s41598-017-10119-1

Li M, Cheng C, Zhang X, Zhou S, Li L, Yang S (2019) Overexpression of pear (*Pyrus pyrifolia*) *CAD2* in tomato affects lignin content. Molecules 24(14):2595. https://doi.org/10.3390/molecules24142595

Li L, Yang K, Wang S, Lou Y, Zhu C, Gao Z (2020) Genome-wide analysis of laccase genes in moso bamboo highlights PeLAC10 involved in lignin biosynthesis and in response to abiotic stresses. Plant Cell Rep 39(6):751–763. https://doi.org/10.1007/s00299-020-02528-w

Liu F, Zheng K, Chen H-C, Liu Z-F (2018a) Capping-RACE: a simple, accurate, and sensitive 5′ RACE method for use in prokaryotes. Nucleic Acids Res 46(21):e129–e129. https://doi.org/10.1093/nar/gky739

Liu Q, Luo L, Zheng L (2018b) Lignins: biosynthesis and biological functions in plants. Int J Mol Sci 19(2):335

Liu G, Zeng H, Li X, Wei Y, Shi H (2019) Functional analysis of MaWRKY24 in transcriptional activation of autophagy-related gene 8f/g and plant disease susceptibility to soil-borne *Fusarium oxysporum* f. sp. *cubense*. Pathogens 8(4):264. https://doi.org/10.3390/pathogens8040264

Liu W, Jiang Y, Wang C, Zhao L, Jin Y, Xing Q, Li M, Lv T, Qi H (2020) Lignin synthesized by CmCAD2 and CmCAD3 in oriental melon (*Cucumis melo* L.) seedlings contributes to drought tolerance. Plant Mol Biol 103(6):689–704. https://doi.org/10.1007/s11103-020-01018-7

Liu X, Van Acker R, Voorend W, Pallidis A, Goeminne G, Pollier J, Morreel K, Kim H, Muylle H, Bosio M, Ralph J, Vanholme R, Boerjan W (2021) Rewired phenolic metabolism and improved saccharification efficiency of a *Zea mays cinnamyl alcohol dehydrogenase 2* (*zmcad2*) mutant. Plant J. https://doi.org/10.1111/tpj.15108

Liu Z-Q, Shi L-P, Yang S, Qiu S-S, Ma X-L, Cai J-S, Guan D-Y, Wang Z-H, He S-L (2021) A conserved double-W box in the promoter of CaWRKY40 mediates autoregulation during response to pathogen attack and heat stress in pepper. Mol Plant Pathol 22(1):3–18. https://doi.org/10.1111/mpp.13004

Lockhart J (2015) Altering lignin composition to improve biofuel production. Plant Cell 27(8):2082–2082. https://doi.org/10.1105/tpc.15.00668

Ma Q-H (2010) Functional analysis of a cinnamyl alcohol dehydrogenase involved in lignin biosynthesis in wheat. J Exp Bot 61(10):2735–2744. https://doi.org/10.1093/jxb/erq107

Ma D, Xu C, Alejos-Gonzalez F, Wang H, Yang J, Judd R, Xie D-Y (2018a) Overexpression of *Artemisia annua* cinnamyl alcohol dehydrogenase increases lignin and coumarin and reduces artemisinin and other sesquiterpenes. Front Plant Sci 9:828–828. https://doi.org/10.3389/fpls.2018.00828

Ma Q-H, Zhu H-H, Qiao M-Y (2018) Contribution of both lignin content and sinapyl monomer to disease resistance in tobacco. Plant Pathol 67(3):642–650. https://doi.org/10.1111/ppa.12767

Mahlia TMI, Ismail N, Hossain N, Silitonga AS, Shamsuddin AH (2019) Palm oil and its wastes as bioenergy sources: a comprehensive review. Environ Sci Pollut Res 26(15):14849–14866. https://doi.org/10.1007/s11356-019-04563-x

Markus BM, Waldman BS, Lorenzi HA, Lourido S (2021) High-resolution mapping of transcription initiation in the asexual stages of *Toxoplasma gondii*. Front Cell Infect Microbiol. https://doi.org/10.3389/fcimb.2020.617998

Martin AF, Tobimatsu Y, Kusumi R, Matsumoto N, Miyamoto T, Lam PY, Yamamura M, Koshiba T, Sakamoto M, Umezawa T (2019) Altered lignocellulose chemical structure and molecular assembly in cinnamyl alcohol dehydrogenase-deficient rice. Sci Rep 9(1):17153. https://doi.org/10.1038/s41598-019-53156-8

Molina C, Grotewold E (2005) Genome wide analysis of *Arabidopsis* core promoters. BMC Genomics 6(1):25. https://doi.org/10.1186/1471-2164-6-25

Mutuku JM, Cui S, Hori C, Takeda Y, Tobimatsu Y, Nakabayashi R, Mori T, Saito K, Demura T, Umezawa T, Yoshida S, Shirasu K (2019) The structural integrity of lignin is crucial for resistance against *Striga hermonthica* parasitism in rice. Plant Physiol 179(4):1796–1809. https://doi.org/10.1104/pp.18.01133

Nepal C, Hadzhiev Y, Balwierz P, Tarifeño-Saldivia E, Cardenas R, Wragg JW, Suzuki A-M, Carninci P, Peers B, Lenhard B, Andersen JB, Müller F (2020) Dual-initiation promoters with intertwined canonical and TCT/TOP transcription start sites diversify transcript processing. Nat Commun 11(1):168. https://doi.org/10.1038/s41467-019-13687-0

Park HC, Kim ML, Kang YH, Jeon JM, Yoo JH, Kim MC, Park CY, Jeong JC, Moon BC, Lee JH, Yoon HW, Lee S-H, Chung WS, Lim CO, Lee SY, Hong JC, Cho MJ (2004) Pathogen- and NaCl-induced expression of the SCaM-4 promoter is mediated in part by a GT-1 box that interacts with a GT-1-Like transcription factor. Plant Physiol 135(4):2150–2161. https://doi.org/10.1104/pp.104.041442

Park HL, Kim TL, Bhoo SH, Lee TH, Lee SW, Cho MH (2018) Biochemical characterization of the rice cinnamyl alcohol dehydrogenase gene family. Molecules. https://doi.org/10.3390/molecules23102659

Paterson RRM (2019) *Ganoderma boninense* disease of oil palm to significantly reduce production after 2050 in Sumatra if projected climate change occurs. Microorganisms 7(1):24. https://doi.org/10.3390/microorganisms7010024

Persson J, Steglich B, Smialowska A, Boyd M, Bornholdt J, Andersson R, Schurra C, Arcangioli B, Sandelin A, Nielsen O, Ekwall K (2016) Regulating retrotransposon activity through the use of alternative transcription start sites. EMBO Reports 17(5):753–768. https://doi.org/10.15252/embr.201541866

Poovaiah CR, Nageswara-Rao M, Soneji JR, Baxter HL, Stewart CN Jr (2014) Altered lignin biosynthesis using biotechnology to improve lignocellulosic biofuel feedstocks. Plant Biotechnol J 12(9):1163–1173. https://doi.org/10.1111/pbi.12225

Preisner M, Kulma A, Zebrowski J, Dymińska L, Hanuza J, Arendt M, Starzycki M, Szopa J (2014) Manipulating cinnamyl alcohol dehydrogenase (CAD) expression in flax affects fibre composition and properties. BMC Plant Biol 14(1):50. https://doi.org/10.1186/1471-2229-14-50

Qiao X, Li Q, Yin H, Qi K, Li L, Wang R, Zhang S, Paterson AH (2019) Gene duplication and evolution in recurring polyploidization–diploidization cycles in plants. Genome Biol 20(1):38. https://doi.org/10.1186/s13059-019-1650-2

Qiu W, Song X, Han X, Liu M, Qiao G, Zhuo R (2018) Overexpression of *Sedum alfredii* cinnamyl alcohol dehydrogenase increases the tolerance and accumulation of cadmium in *Arabidopsis*. Environ Exp Bot 155:566–577. https://doi.org/10.1016/j.envexpbot.2018.08.003

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(3):1051–1071. https://doi.org/10.1104/pp.103.026484

Ren R, Wang H, Guo C, Zhang N, Zeng L, Chen Y, Ma H, Qi J (2018) Widespread whole genome duplications contribute to genome complexity and species diversity in angiosperms. Mol Plant 11(3):414–428. https://doi.org/10.1016/j.molp.2018.01.002

Repka V, Fischerová I, Šilhárová K (2004) Methyl Jasmonate is a potent elicitor of multiple defense responses in grapevine leaves and cell-suspension cultures. Biol Plant 48(2):273–283. https://doi.org/10.1023/B:BIOP.0000033456.27521.e5

Rojas-Duran MF, Gilbert WV (2012) Alternative transcription start site selection leads to large differences in translation activity in yeast. RNA 18(12):2299–2305. https://doi.org/10.1261/rna.035865.112

Rong W, Luo M, Shan T, Wei X, Du L, Xu H, Zhang Z (2016) A wheat cinnamyl alcohol dehydrogenase TaCAD12 contributes to host resistance to the sharp eyespot disease. Front Plant Sci. https://doi.org/10.3389/fpls.2016.01723

Rouster J, Leah R, Mundy J, Cameron-Mills V (1997) Identification of a methyl jasmonate-responsive region in the promoter of a lipoxygenase 1 gene expressed in barley grain. Plant J 11(3):513–523. https://doi.org/10.1046/j.1365-313X.1997.11030513.x

Rubio-Somoza I, Martinez M, Abraham Z, Diaz I, Carbonero P (2006) Ternary complex formation between HvMYBS3 and other factors involved in transcriptional control in barley seeds. Plant J 47(2):269–281. https://doi.org/10.1111/j.1365-313X.2006.02777.x

Rushton PJ, Torres JT, Parniske M, Wernert P, Hahlbrock K, Somssich IE (1996) Interaction of elicitor-induced DNA-binding proteins with elicitor response elements in the promoters of parsley PR1 genes. EMBO J 15(20):5690–5700

Saathoff AJ, Sarath G, Chow EK, Dien BS, Tobias CM (2011a) Down-regulation of cinnamyl-alcohol dehydrogenase in switchgrass by RNA silencing results in enhanced glucose release after cellulase treatment. PLoS ONE 6(1):e16416. https://doi.org/10.1371/journal.pone.0016416

Saathoff AJ, Tobias CM, Sattler SE, Haas EJ, Twigg P, Sarath G (2011b) Switchgrass contains two cinnamyl alcohol dehydrogenases involved in lignin formation. BioEnergy Res 4(2):120–133. https://doi.org/10.1007/s12155-010-9106-2

Saballos A, Ejeta G, Sanchez E, Kang C, Vermerris W (2009) A Genomewide analysis of the cinnamyl alcohol dehydrogenase family in sorghum [*Sorghum bicolor* (L.) Moench] identifies *SbCAD2* as the *Brown midrib6* Gene. Genetics 181(2):783–795. https://doi.org/10.1534/genetics.108.098996

Séguin A, Laible G, Leyva A, Dixon RA, Lamb CJ (1997) Characterization of a gene encoding a DNA-binding protein that interacts in vitro with vascular specific cis elements of the phenylalanine ammonia-lyase promoter. Plant Mol Biol 35(3):281–291. https://doi.org/10.1023/A:1005853404242

Shani E, Weinstain R, Zhang Y, Castillejo C, Kaiserli E, Chory J, Tsien RY, Estelle M (2013) Gibberellins accumulate in the elongating endodermal cells of *Arabidopsis* root. Proc Natl Acad Sci USA 110(12):4834–4839. https://doi.org/10.1073/pnas.1300436110

Sibout R, Eudes A, Pollet B, Goujon T, Mila I, Granier F, Séguin A, Lapierre C, Jouanin L (2003) Expression pattern of two paralogs encoding cinnamyl alcohol dehydrogenases in Arabidopsis. Isolation and characterization of the corresponding mutants. Plant Physiol 132(2):848–860. https://doi.org/10.1104/pp.103.021048

Sibout R, Eudes A, Mouille G, Pollet B, Lapierre C, Jouanin L, Séguin A (2005) *cinnamyl alcohol dehydrogenase-C* and *-D* are the primary genes involved in lignin biosynthesis in the floral stem of *Arabidopsis*. Plant Cell 17(7):2059–2076. https://doi.org/10.1105/tpc.105.030767

Simpson SD, Nakashima K, Narusaka Y, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (2003) Two different novel *cis*-acting elements of *erd1*, a *clpA* homologous *Arabidopsis* gene function in induction by dehydration stress and dark-induced senescence. Plant J 33(2):259–270. https://doi.org/10.1046/j.1365-313X.2003.01624.x

Singh R, Ong-Abdullah M, Low E-TL, Manaf MAA, Rosli R, Nookiah R, Ooi LC-L, Ooi SE, Chan K-L, Halim MA, Azizi N, Nagappan J, Bacher B, Lakey N, Smith SW, He D, Hogan M, Budiman MA, Lee EK, DeSalle R, Kudrna D, Goicoechea JL, Wing RA, Wilson RK, Fulton RS, Ordway JM, Martienssen RA, Sambanthamurthi R (2013) Oil palm genome sequence reveals

divergence of interfertile species in old and new worlds. Nature 500(7462):335–339. https://doi.org/10.1038/nature12309

Sun S-C, Xiong X-P, Zhang X-L, Feng H-J, Zhu Q-H, Sun J, Li Y-J (2020) Characterization of the *Gh4CL* gene family reveals a role of *Gh4CL7* in drought tolerance. BMC Plant Biol 20(1):125. https://doi.org/10.1186/s12870-020-2329-2

Tang L, Nie S, Li W, Fan C, Wang S, Wu F, Pan K (2019) Wheat straw increases the defense response and resistance of watermelon monoculture to Fusarium wilt. BMC Plant Biol 19(1):551. https://doi.org/10.1186/s12870-019-2134-y

Thum KE, Kim M, Morishige DT, Eibl C, Koop HU, Mullet JE (2001) Analysis of barley chloroplast psbD light-responsive promoter elements in transplastomic tobacco. Plant Mol Biol 47(3):353–366

Tobias CM, Chow EK (2005) Structure of the cinnamyl-alcohol dehydrogenase gene family in rice and promoter activity of a member associated with lignification. Planta 220(5):678–688. https://doi.org/10.1007/s00425-004-1385-4

Tokizawa M, Kusunoki K, Koyama H, Kurotani A, Sakurai T, Suzuki Y, Sakamoto T, Kurata T, Yamamoto YY (2017) Identification of *Arabidopsis* genic and non-genic promoters by paired-end sequencing of TSS tags. Plant J 90(3):587–605. https://doi.org/10.1111/tpj.13511

Trabucco GM, Matos DA, Lee SJ, Saathoff AJ, Priest HD, Mockler TC, Sarath G, Hazen SP (2013) Functional characterization of cinnamyl alcohol dehydrogenase and caffeic acid O-methyltransferase in *Brachypodium distachyon*. BMC Biotechnol 13(1):61. https://doi.org/10.1186/1472-6750-13-61

Tronchet M, Balagué C, Kroj T, Jouanin L, Roby D (2010) Cinnamyl alcohol dehydrogenases-C and D, key enzymes in lignin biosynthesis, play an essential role in disease resistance in *Arabidopsis*. Mol Plant Pathol 11(1):83–92. https://doi.org/10.1111/j.1364-3703.2009.00578.x

Tsuruta S-i, Ebina M, Kobayashi M, Akashi R, Kawamura O (2010) Structure and expression profile of the cinnamyl alcohol dehydrogenase gene and its association with lignification in the sorghum (*Sorghum bicolor* (L.) Moench) *bmr*-6 mutant. Breed Sci 60(4):314–323. https://doi.org/10.1270/jsbbs.60.314

Tu M, Wang X, Yin W, Wang Y, Li Y, Zhang G, Li Z, Song J, Wang X (2020) Grapevine VlbZIP30 improves drought resistance by directly activating VvNAC17 and promoting lignin biosynthesis through the regulation of three peroxidase genes. Horticult Res 7(1):150. https://doi.org/10.1038/s41438-020-00372-3

Van Acker R, Déjardin A, Desmet S, Hoengenaert L, Vanholme R, Morreel K, Laurans F, Kim H, Santoro N, Foster C, Goeminne G, Légée F, Lapierre C, Pilate G, Ralph J, Boerjan W (2017) Different routes for conifer- and sinapaldehyde and higher saccharification upon deficiency in the dehydrogenase CAD1. Plant Physiol 175(3):1018–1039. https://doi.org/10.1104/pp.17.00834

Vanholme R, Demedts B, Morreel K, Ralph J, Boerjan W (2010) Lignin biosynthesis and structure. Plant Physiol 153(3):895–905. https://doi.org/10.1104/pp.110.155119

Wang Q, Yuan F, Pan Q, Li M, Wang G, Zhao J, Tang K (2010) Isolation and functional analysis of the Catharanthus roseus deacetylvindoline-4-O-acetyltransferase gene promoter. Plant Cell Rep 29(2):185–192. https://doi.org/10.1007/s00299-009-0811-2

Wang Y, Chantreau M, Sibout R, Hawkins S (2013) Plant cell wall lignification and monolignol metabolism. Front Plant Sci. https://doi.org/10.3389/fpls.2013.00220

Wang G-L, Que F, Xu Z-S, Wang F, Xiong A-S (2015) Exogenous gibberellin altered morphology, anatomic and transcriptional regulatory networks of hormones in carrot root and shoot. BMC Plant Biol 15(1):290. https://doi.org/10.1186/s12870-015-0679-y

Wang Z, Liu Q, Wang H, Zhang H, Xu X, Li C, Yang C (2016) Comprehensive analysis of trihelix genes and their expression under biotic and abiotic stresses in *Populus trichocarpa*. Sci Rep 6(1):36274. https://doi.org/10.1038/srep36274

Wang G-L, An Y-H, Wang Y-H, Liu J-X, Wang J-Z, Sun M, Xiong A-S (2020) Gibberellin-induced alterations to the expression of cell wall-related genes in the xylem of carrot root. J Plant Growth Regul. https://doi.org/10.1007/s00344-020-10143-y

Wen W, Wang R, Su L, Lv A, Zhou P, An Y (2021) MsWRKY11, activated by MsWRKY22, functions in drought tolerance and modulates lignin biosynthesis in alfalfa (*Medicago sativa* L.). Environ Exp Bot 184:104373. https://doi.org/10.1016/j.envexpbot.2021.104373

Xie M, Zhang J, Tschaplinski TJ, Tuskan GA, Chen J-G, Muchero W (2018) Regulation of lignin biosynthesis and its role in growth-defense tradeoffs. Front Plant Sci. https://doi.org/10.3389/fpls.2018.01427

Xu X, Chen C, Fan B, Chen Z (2006) Physical and functional interactions between pathogen-induced *Arabidopsis* WRKY18, WRKY40, and WRKY60 transcription factors. Plant Cell 18(5):1310–1326. https://doi.org/10.1105/tpc.105.037523

Xu Y, Thammannagowda S, Thomas TP, Azadi P, Schlarbaum SE, Liang H (2013) LtuCAD1 is a cinnamyl alcohol dehydrogenase ortholog involved in lignin biosynthesis in *Liriodendron tulipifera* L., a basal angiosperm timber species. Plant Mol Biol Report 31(5):1089–1099. https://doi.org/10.1007/s11105-013-0578-z

Xu H, Shi X, He L, Guo Y, Zang D, Li H, Zhang W, Wang Y (2018) *Arabidopsis thaliana* trihelix transcription factor AST1 mediates salt and osmotic stress tolerance by binding to a novel AGAG-box and some GT motifs. Plant Cell Physiol 59(5):946–965. https://doi.org/10.1093/pcp/pcy032

Xu C, Park J-K, Zhang J (2019) Evidence that alternative transcriptional initiation is largely nonadaptive. PLoS Biol 17(3):e3000197. https://doi.org/10.1371/journal.pbio.3000197

Yamamoto YY, Ichida H, Abe T, Suzuki Y, Sugano S, Obokata J (2007) Differentiation of core promoter architecture between plants and mammals revealed by LDSS analysis. Nucleic Acids Res 35(18):6219–6226. https://doi.org/10.1093/nar/gkm685

Yamamoto M, Tomiyama H, Koyama A, Okuizumi H, Liu S, Vanholme R, Goeminne G, Hirai Y, Shi H, Takata N, Ikeda T, Uesugi M, Kim H, Sakamoto S, Mitsuda N, Boerjan W, Ralph J, Kajita S (2020) A century-old mystery unveiled: Sekizaisou is a natural lignin mutant. Plant Physiol 182(4):1821–1828. https://doi.org/10.1104/pp.19.01467

Yang J, Duan G, Li C, Liu L, Han G, Zhang Y, Wang C (2019) The crosstalks between jasmonic acid and other plant hormone signaling highlight the involvement of jasmonic acid as a core component in plant response to biotic and abiotic stresses. Front Plant Sci. https://doi.org/10.3389/fpls.2019.01349

Yang J, Wang Q, Luo H, He C, An B (2020) HbWRKY40 plays an important role in the regulation of pathogen resistance in *Hevea brasiliensis*. Plant Cell Rep 39(8):1095–1107. https://doi.org/10.1007/s00299-020-02551-x

Yoon J, Choi H, An G (2015) Roles of lignin biosynthesis and regulatory genes in plant development. J Integr Plant Biol 57(11):902–912. https://doi.org/10.1111/jipb.12422

Youn B, Camacho R, Moinuddin SG, Lee C, Davin LB, Lewis NG, Kang C (2006) Crystal structures and catalytic mechanism of the *Arabidopsis* cinnamyl alcohol dehydrogenases AtCAD5 and AtCAD4. Org Biomol Chem 4(9):1687–1697. https://doi.org/10.1039/b601672c

Yusuf CYL, Abdullah JO, Shaharuddin NA, Abu Seman I, Abdullah MP (2018a) Characterization of promoter of *EgPAL1*, a novel *PAL* gene from the oil palm *Elaeis guineensis* Jacq. Plant Cell Rep 37(2):265–278. https://doi.org/10.1007/s00299-017-2228-7

Yusuf CYL, Abu Seman I, Nor Aini AS, Mohd Nor MN, Abdullah MP (2018) Cloning and analysis of the Eg4CL1 gene and its promoter from oil palm (*Elaeis guineensis* Jacq.). Sains Malaysiana 47(8):1709–1723

Zhang P-J, He Y-C, Zhao C, Ye Z-H, Yu X-P (2018) Jasmonic acid-dependent defenses play a key role in defending tomato against *Bemisia tabaci* nymphs, but not adults. Front Plant Sci. https://doi.org/10.3389/fpls.2018.01065

Zhang Y, Wu L, Wang X, Chen B, Zhao J, Cui J, Li Z, Yang J, Wu L, Wu J, Zhang G, Ma Z (2019) The cotton laccase gene *GhLAC15* enhances Verticillium wilt resistance via an increase in defence-induced lignification and lignin components in the cell walls of plants. Mol Plant Pathol 20(3):309–322. https://doi.org/10.1111/mpp.12755

Zhang P, Dimont E, Ha T, Swanson DJ, Hide W, Goldowitz D, the FC (2017) Relatively frequent switching of transcription start sites during cerebellar development. BMC Genomics 18(1):461. https://doi.org/10.1186/s12864-017-3834-z

Zhao Q, Tobimatsu Y, Zhou R, Pattathil S, Gallego-Giraldo L, Fu C, Jackson LA, Hahn MG, Kim H, Chen F, Ralph J, Dixon RA (2013) Loss of function of cinnamyl alcohol dehydrogenase 1 leads to unconventional lignin and a temperature-sensitive growth defect in *Medicago truncatula*. Proc Natl Acad Sci USA 110(33):13660–13665. https://doi.org/10.1073/pnas.1312234110

Zhong R, Ye Z-H (2009) Transcriptional regulation of lignin biosynthesis. Plant Signal Behav 4(11):1028–1034. https://doi.org/10.4161/psb.4.11.9875

Zhou J, Lee C, Zhong R, Ye Z-H (2009) MYB58 and MYB63 are transcriptional activators of the lignin biosynthetic pathway during secondary cell wall formation in *Arabidopsis*. Plant Cell 21(1):248–266. https://doi.org/10.1105/tpc.108.063321

Zuo Y-C, Li Q-Z (2011) Identification of TATA and TATA-less promoters in plant genomes by integrating diversity measure GC-Skew and DNA geometric flexibility. Genomics 97(2):112–120. https://doi.org/10.1016/j.ygeno.2010.11.002