RESEARCH ARTICLE

Cloning and *in silico* analysis of a cinnamyl alcohol dehydrogenase gene in *Pennisetum purpureum*

RAN TANG, XIANG-QIAN ZHANG, YOU-HAN LI and XIN-MING XIE*

College of Agriculture, South China Agricultural University, Guangzhou 510642, People's Republic of China

Abstract

Lignin is a major constituent of plant cell walls and indispensable to the normal growth of a plant. However, the presence of lignin complicates the structure of the plant cell walls and negatively influences pulping industry, lignocellulose utilization as well as forage properties. Cinnamyl alcohol dehydrogenase (CAD), a key enzyme involved in lignin biosynthesis, catalyses the last step in monolignol synthesis and has a major role in genetic regulation of lignin production. In the present study, a 1 342-bp cDNA fragment of *CAD* gene, named *PpCAD*, was isolated from *Pennisetum purpureum* using strategies of homologous clone and rapid amplification of cDNA end. It was translated into an intact protein sequence including 366 amino acid residues by ORF Finder. The genomic full-length DNA of *PpCAD* was a 3 738-bp sequence containing four exons and three introns, among which the 114-bp exon was considered to be a conserved region compared with other *CADs*. Basic bioinformatic analysis presumed that the PpCAD was a nonsecretory and hydrophobic protein with five possible transmembrane helices. The phylogenetic analysis indicated that the PpCAD belonged to the class of bona fide CADs involved in lignin synthesis and it showed a high similarity (nearly 90%) with CAD protein sequences of *Sorghum bicolor*, *Panicum virgatum* and *Zea mays* in Gramineae. Furthere, PpCAD amino acid sequence was demonstrated to have some conserved motifs such as Zn-binding site, Zn-catalytic centre and NADP(H) binding domain after aligning with other bona fide CADs. Three-dimensional homology modelling of PpCAD showed that the protein had some exclusive features of bona fide CADs.

Introduction

Lignin, a complex three-dimensional phenolic polymer derived mainly from hydroxycinnamyl alcohol, plays a crucial role in plant adaptation to terrestrial environments (Peter and Neale 2004; Guo *et al.* 2010; Weng and Chapple 2010). It is formed by three main subunits (monolignols), *p*-coumaryl (H), coniferyl (G) and sinapyl (S) alcohols and provides plant with structural rigidity, a protective barrier against pathogens as well as a system for water and nutrient conduction (Boerjan *et al.* 2003; Barakat *et al.* 2009; Weng and Chapple 2010). In short, lignin is a major component existing in plant secondary cell wall, which ensures the proper functioning of plant systems (Eudes *et al.* 2012). However, lignin in plant cell walls is hard to remove and, as such, is a problem in pulp manufacturing, conversion of lignocellulosic biomass and dry matter digestibility of forages (Ragauskas *et al.* 2006;

Hisano *et al.* 2009; Fu *et al.* 2011). Genetically modifying elite germplasm may diminish recalcitrance of undesirable lignin to some extent (Hisano *et al.* 2009; Fu *et al.* 2009, 2011), and most of the genes involved in lignin biosynthetic pathway have been successfully cloned and characterized (Dixon *et al.* 2001; Boerjan *et al.* 2003). Studies on the lignin metabolic pathway help to understand the plant development and modify plants for meeting the industrial and livestock needs (Baucher *et al.* 1996; Hisano *et al.* 2009).

CAD (EC 1.1.1.195) is a multifunctional enzyme that catalyses the last step in lignin monolignol biosynthesis and the conversion of cinnamyl aldehydes to cinnamyl alcohols (Saballos *et al.* 2009; Ma 2010; Weng and Chapple 2010). It is the first enzyme purified from lignin biosynthetic pathway (Mansell *et al.* 1974). When the first CAD gene was isolated from *Nicotiana tabacum* stems in 1992 (Halpin *et al.* 1992; Knight *et al.* 1992; Kim *et al.* 2004), the relationship between *CAD* genes and their functions was of great importance. Since the important role in regulation of lignin content and composition, more and more *CAD* genes and their

---

*For correspondence. E-mail: xiexmbs@scau.edu.cn.

Keywords. lignin biosynthesis; cinnamyl alcohol dehydrogenase; clone; *in silico* analysis; *Pennisetum purpureum*.

homologues (bona fide *CAD* and *CAD*-like genes) have been isolated from different species, such as *Pinus taeda* (O'Malley *et al.* 1992), *Medicago sativa* (Van Doorsselaere *et al.* 1995; Brill *et al.* 1999), *Lolium perenne* (Lynch *et al.* 2002) and *Sorghum bicolor* (Tsuruta *et al.* 2007) etc. It was shown that *CAD* genes existed in plants in the form of a multi-gene family by means of expression analyses of *CADs* from *Populus* (Barakat *et al.* 2009), *Triticum aestivum* (Ma 2010), *Brachypodium distachyon* (Bukh *et al.* 2012) and enzyme kinetic characterization studies of *Panicum virgatum* CAD1 (PviCAD1) (Saathoff *et al.* 2012). Further, different gene family members had distinct expression characters during plant development, indicating that they might play diversified roles for ensuring the normal growth of the plant (Barakat *et al.* 2009; Guo *et al.* 2010; Saathoff *et al.* 2012).

In terms of fully sequenced plant genomes, about 17 *CAD* genes have been annotated from *Arabidopsis*, but only nine of them (AtCAD1–AtCAD9) show affinity with the cinnamyl alcohol substrates (Tavares *et al.* 2000; Sibout *et al.* 2003; Kim *et al.* 2004). *Oryza sativa* and *Populus trichocarpa* gene families consist of 12 and 15 *CAD* gene members, respectively (Tobias and Chow 2005; Barakat *et al.* 2009). On the other hand, a homologue of CAD, named sinapyl alcohol dehydrogenase (SAD), was first isolated from *Populus tremuloides* (PotSAD) (Li *et al.* 2001). It was shown to be a key enzyme involved in the sinapyl monolignol biosynthesis. Most recently, almost 80 *CAD* genes from more than 35 plants have been identified (Guo *et al.* 2010). They have arbitrarily been assigned roles in monolignols and lignin formation, and the functions of individual *CAD* genes are different from species to species (Sibout *et al.* 2003).

Due to the important role of CAD in determining lignin content and composition, the regulation of this enzyme in lignin polymer biosynthesis has been studied in several plants, especially in woody plants (*Populus*, *Pinus*) (Baucher *et al.* 1996; MacKay *et al.* 1997), energy plants (*Panicum virgatum*, *Zea mays*) (Fu *et al.* 2009; Fornalé *et al.* 2012), and forage plants (*M. sativa, Festuca arundinacea*) (Baucher *et al.* 1999; Chen *et al.* 2003; Jackson *et al.* 2008). Downregulating *CAD* gene is very meaningful to these three kinds of plants. Likewise, downregulating *CAD* expression in *Populus* caused a change in the number of aldehyde units in the lignin, and alkaline pulping experiments on the three-month-old *CAD* downregulated trees showed a reduction of the kappa number without affecting the degree of cellulose degradation (Baucher *et al.* 1996), implying that reducing the CAD activity in woody plants is beneficial to pulp and paper industry. Similarly, the decrease in *CAD* expression level of RNA interference (RNAi) transgenic *Panicum virgatum*, caused not only a reduction in lignin content, but also a change in lignin composition. The modification of lignin biosynthesis exerts a positive impact on the improvement of sugar release and the increase of ethanol yield (Fu *et al.* 2009). Antisense approach had been used to downregulate *M. sativa CAD* for improving forage crop digestibility because of the change of lignin composition (Baucher *et al.* 1999).

All these researches just suggest that *CAD* is a key gene for lignin genetic modification.

In recent years, bona fide *CAD* genes involved in the monolignol biosynthesis have been identified in many species (Guo *et al.* 2010), but few reports are focused on monocots (Fornalé *et al.* 2012). The functions of other members (*CAD*-like genes) are still elusive (Ma 2010), and their detailed biochemical characterization with regard to encoded protein is also enigmatic (Pandey *et al.* 2011). Meanwhile, bona fide *CAD* genes and *CAD*-like genes belong to the same gene family, though the similarity between them is low. For instance, three *CAD* genes had been isolated from *Camellia sinensis*, but they showed only 20–54% identities at nucleotide level (Deng *et al.* 2011). Thus, it is difficult to assign unique functions to individual *CAD* genes. Only when the functions of all members involved in *CAD* gene family had been explored thoroughly, could *CAD* gene be a target for effectively regulating lignin synthesis by gene engineering.

Elephant grass (*P. purpureum* Schum.) is a kind of fast-growing C4 grass with high biomass production (de Morais *et al.* 2011),which grows well in humid and semihumid tropic regions (Xie *et al.* 2009; Videira *et al.* 2012). It can be used as nonwood fibre (Madakadze *et al.* 2010), green energy plant as well as forage grass (Matthews-Amune and Kakulu 2012). The content and composition of lignin determine the utilizability of this plant. In this work, we present the isolation and *in silico* characterization of a *CAD* gene from *P. purpureum* (named as PpCAD), including predicted structure of the protein. We have also carried out a detailed phylogenetic analysis of PpCAD with a wide variety of plants, aiming to find out the critical relationships between PpCAD and other identified CAD proteins, and to lay the foundations for further expressional analysis and gene regulation. To date, *PpCAD* is the first gene cloned and analysed for *P. purpureum*, and this is the first report about this key gene involved in *P. purpureum* lignin biosynthesis pathway.

Materials and methods

Plant material

P. purpureum cv. Huanan was used for all experiments. Plants were grown in the Forage Introduction Garden of South China Agricultural University under natural conditions. Harvested stems and leaves were frozen in liquid nitrogen stored at −80°C until used.

Synthesis of cDNA and extraction of genomic DNA

Total RNA was extracted from stem using Trizol reagent (Takara, Dalian, China). Reverse transcription into cDNA was carried out by AMV reverse transcriptase kit (Takara). Leaf tissues from mature plants were used for genomic DNA extraction according to Doyle and Doyle (1987) with slight modifications.

Cloning of partial CAD gene of P. purpureum

Homologous clone strategy was used to clone partial *CAD* sequence (EST) of *P. purpureum*. The known *CAD* gene sequences from some gramineous plants were retrieved from GenBank (http://www.ncbi.nlm.nih.gov/) and then aligned to find out their conserved nucleotide regions using ClustalW2 (http://www.ebi.ac.uk/Tools/clustalw2/index.html). According to these conserved regions, two primers (CF1: 5′-AAC ACAGGCCCTGAAGATGT-3′, CR2: 5′-GACGTCGACGA CGAAGCGGTA-3′) were designed for the amplification of target fragment. PCR reaction mixture, 50 μL, included 5 μL 10× Ex *Taq* buffer (with Mg^{2+}), 4 μL dNTP (2.5 mM), 1 μL each of the primers (10 mM), 5–10 ng template cDNA, 0.5 μL Ex *Taq* (5 U/μL, Takara), finally adding sterile double distilled water to the total volume. PCR mixture was subjected to 35 cycles of denaturation at 95°C for 30 s, annealing at 52°C for 35 s, and extension at 72°C for 2 min. Target products amplified (EST fragments) were detected in a 1% (m/v) agarose gel containing 5% (v/v) Golden View (Newprobe, Beijing, China) and visualized under ultraviolet light.

The positive PCR products were sequenced by Invitrogen (Shanghai, China). Then, the nested upstream primers, outer1 (5′-AGAGGATCTGGTCCTACAACGA-3′) and inner1 (5′-GCACTTCGGTCTGACATCCCCC-3′) were designed for cloning 3′ UTR of *CAD* gene based on the obtained EST fragment. The cloning of 3′ end was performed using the 3′ Full RACE Core Set ver. 2.0 kit (Takara) following manufacturer's instructions. The remaining 5′ sequence (5′ EST) was cloned using the primers CF2 (5′-G(A/G)ATGGG(G/C)A(A/G)CCTGGCGTC-3′) and CR2 (5′-AAGACAGGCGCTGAAGATGT-3′). Reaction system and procedures were the same as the amplification of EST with a different annealing temperature at 58°C.

Full sequence cloning of P. purpureum CAD gene

Based on the obtained EST fragment of CAD gene, three pairs of specific primers CADF1 (5′-GATAAGACAGTGGT GGGATG-3′) and CADR1 (5′-TTCGTGGGAGTTGAATA GAG-3′); CADF2 (5′-GAAGATGTGGTGGTGAAGGTGC T-3′) and CADR2 (5′-CAGTGATGAGTGGTGTTAGAGG-3′); CADF3 (5′-TGCAACAAGAGGATCTGGTCCTA-3′) and CADR3 (5′-TGTCCATGATGTAGTCCAGGGAG-3′) were designed for cloning full-length sequence of *PpCAD*. The PCR mixture contained 5–10 ng genomic DNA, 1 μL each of the specific primers, 25 μL 2× Premix Ex *Taq* PCR (Takara) and double distilled water to a final volume of 50 μL. After amplification by 35 cycles of 95°C for 30 s, 55°C for 35 s, 72°C for 2 min, three target fragments were sequenced and spliced.

Parameter analysis

ORF Finder of NCBI was used to search the open reading frames (ORFs) of *PpCAD* and to deduce its amino acid sequence. The homologies of *PpCAD* gene and its encoded protein were analysed by BLASTN and BLASTP, respectively. The general characters of PpCAD protein were obtained by ProtParam online program (http://web. expasy.org/protparam/, Gasteiger *et al.* 2005), which included molecular mass, theoretical isoelectric point (pI), amino acid composition, atomic composition, instability index as well as grand average of hydropathicity (GRAVY). TMpred (http://www.ch.embnet.org/software/ TMPRED_form.html, Hofmann and Stoffel 1993) was used to predict membrane-spanning regions of PpCAD protein. Prediction of signal peptidesof PpCAD was performed by TargetP (http://www.cbs.dtu.dk/services/TargetP/, Emanuelsson *et al.* 2000) and SignalP (http://www.cbs.dtu. dk/services/SignalP/, Petersen *et al.* 2011).

Alignment and phylogenetic analysis

The CAD sequences used in alignment and phylogenetic analysis were retrieved from three major databases: GenBank, TIGR (The Institute of Genomic Research Database-TDB: http://www.tigr.org) and JGI (Joint Genome Institute:
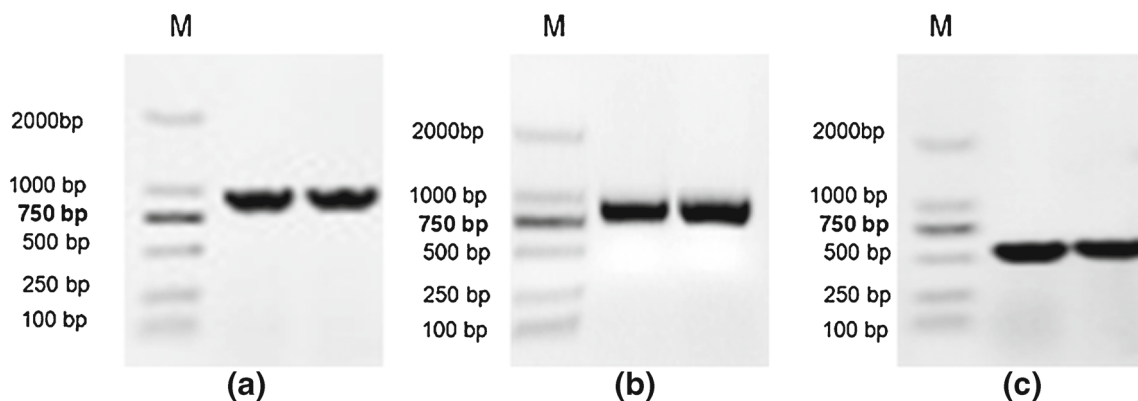
Figure 1. cDNA cloning from *P. purpureum CAD* gene. (a) 940-bp EST isolated; (b) 791-bp 3′ UTR isolated; (c) 514-bp 5′ EST isolated.

http://www.jqi.doe.gov). Multiple alignments of CAD protein sequences were performed by DNAMAN as well as ClustalW2 and figured using ESPript (http://espript.ibcp.fr/ESPript/cgi-bin/ESPript.cgi, Gouet *et al.* 1999). The experimentally proved or putative 62 CAD sequences from 25 species (including PpCAD) were used to reconstruct a Neighbour-joining (NJ) phylogenetic tree in MEGA 5 software based on amino acid sequences, Jones-Taylor- Thornton (JTT) model was selected for analysis and 500 bootstrap replicates were used to estimate branch support.

Structural modelling of PpCAD

Secondary structure prediction of PpCAD was performed in the NPS@ web server (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_server.html), and annotated with aligned results through ESPript. Pfam of Sanger Centre (http://pfam.sanger.ac.uk/search, Finn *et al.* 2010) was used to analyse the functional domains of PpCAD. Three-dimensional structure was predicted using SWISS-MODEL (http://swissmodel.expasy.org/, Arnold *et al.* 2006). Autheprot 3D software (http://antheprot-pbil.ibcp.fr/) was used to generate predicted three-dimensional figure of PpCAD protein. The spatial structure of PpCAD conserved domains and the structural comparisons of PpCAD and the other CADs were analysed using PyMOL software.

Results

Cloning of CAD from P. purpureum

An expected 940-bp EST fragment was amplified using primers CF1 and CF2 based on standard RT-PCR procedure
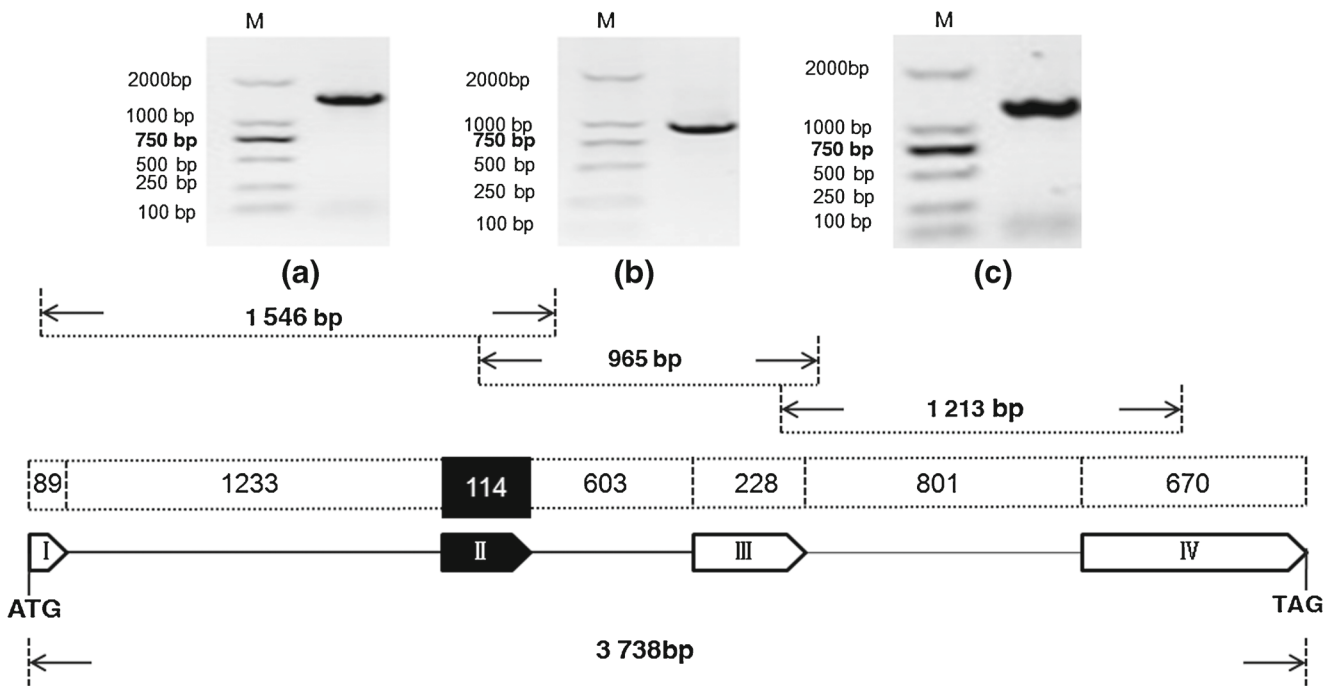

Figure 2. Cloning of *PpCAD* full-length sequence showing the intron–exon structure. (a) Isolating of the first intron using CADF1 and CADR1 within a 1 546-bp product. (b) Isolating of the second intron using CADF2 and CADR2 within a 965-bp product. (c) Isolating of the third intron using CADF3 and CADR3 within a 1 213-bp product. The order of cloning was (c), (b), then (a). Boxes represent exons, lines represent introns. Black box indicates conserved 114-bp sequence identified from other species. Numbers are the size of each fragment measured in nucleotides. ATG and TAG represent initiation and termination codes, respectively.

Table 1. Predicted PpCAD membrane-spanning regions.

| Type | Possible transmembrane helices (aa) | | Centre | Score |
	From	To		
Inside-to-outside (i-o)	176 (176)	200 (195)	186	437
	272 (272)	291 (289)	281	676
Outside-to-inside (o-i)	85 (87)	103 (103)	95	788
	175 (177)	197 (193)	185	657
	273 (273)	291 (289)	281	976

The sequence positions in brackets denominate the core region.
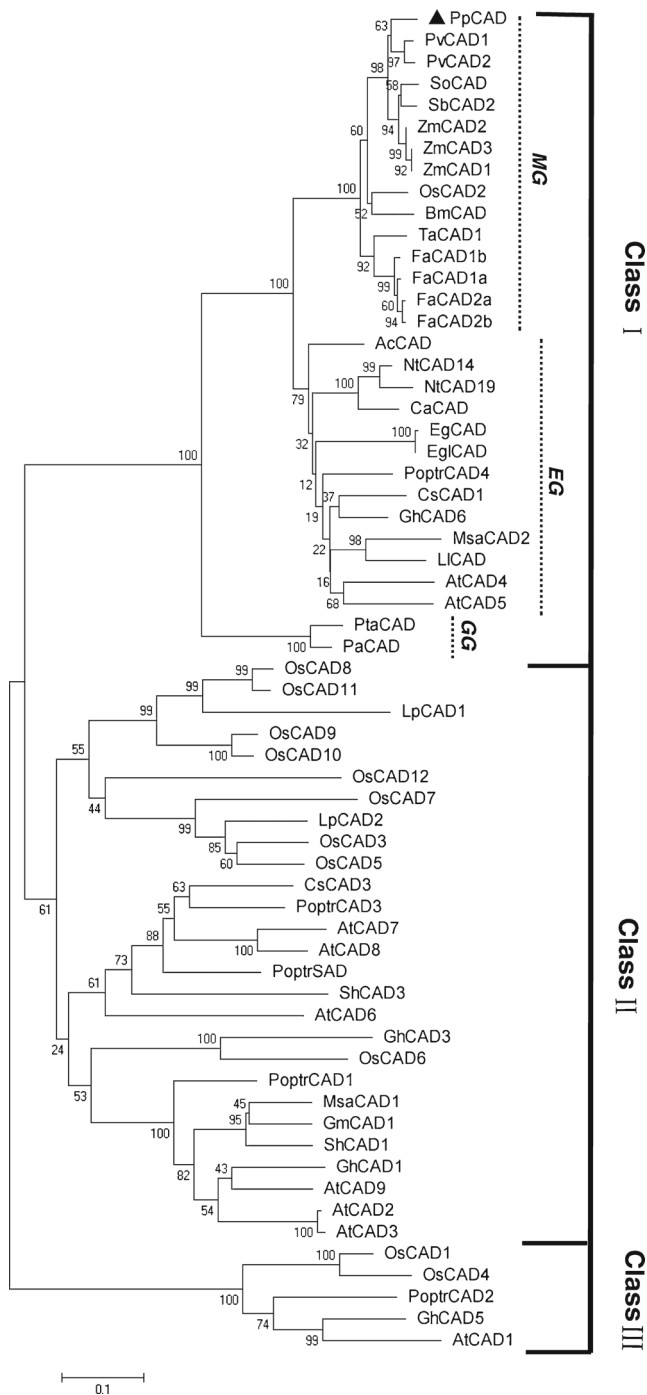
Figure 3. Neighbour-joining phylogenetic tree of 63 CAD protein sequences from various plants. The tree was constructed using MEGA 5 software with Jones-Taylor-Thornton model. Numbers above branches denote bootstrap value with 500 replications. The value of 0.1 represents the evolution distance. PpCAD protein sequence is marked by (▲). Brackets highlight the three classes of CAD. MG, EG and GG represent three groups subdivided from class I (monocot, eudicot and gymnosperm). Sequences used in phylogenetic analysis were retrieved mainly from GenBank, TIGR and JGI. Sequence names and their accession numbers are provided in table 3.

(figure 1a). The similarities between the target EST fragment and *Panicum virgatum PvCAD1*, *Saccharum officinarum SoCAD*, *Zea mays ZmCAD2* were 92%, 91%, 90%, respectively. When the target EST fragment was aligned with *Hordeum vulgare HvCAD, Triticum aestivum TaCAD1, Festuca arundinacea FaCAD2b, Bambusa multiplex BmCAD* and *Lolium perenne LpCAD1*, the similarities were more than 85% either. All of these confirmed the facticity of 940-bp PCR product as a fragment of *PpCAD*.

PCR approach, 3′ RACE, was used for isolating 3′ end of *PpCAD* gene with nested primers outer1 and inner1. The length of obtained 3′ end was 791-bp (figure 1b), which had the highest similarity with *SoCAD* (92%), *PvCAD1* (92%) and *ZmCAD2* (91%). After amplifying the third part of *CAD* cDNA with primers CF2, CR2 and spliced by DNAMAN, we got a 514-bp fragment of 5′ EST (figure 1c), and finally obtained a 1 342-bp sequence containing complete poly (A) tail. The *PpCAD* sequence was found most similar to that from *Sorghum bicolor, Panicum virgatum, Saccharum officinarum* and *Zea mays*, with the identities more than 90% and it could be translated into an intact protein sequence encoding 366 amino acid residues using ORF Finder. This sequence had been submitted to NCBI with the nucleotide accession number HQ840705 and amino acid accession number ADW94628.

A 3 977-bp full-length DNA sequence of *PpCAD* was obtained using three pairs of specific primers and spliced by DNAMAN (figure 2). This sequence was also submitted to NCBI (acc. no. JQ794485). Analysis of gene structure showed that the full-length *PpCAD* contained four exons and three introns. The length of I, II, III and IV exons were 89, 114, 228 and 670-bp, respectively, all introns were in accord with 'GT–AG' rule. In addition, it was worth noting that a 114-bp exon (figure 2, black box) could also be found in most other *CAD* genes. The research of *Arabidopsis AtCAD4* and *AtCAD5* demonstrated this 114-bp sequence encoded the putative binding site I for the monolignol substrate (Youn *et al.* 2006).

General characters of P. purpureum CAD protein

ProtParam was used to predict basic physical and chemical parameters of PpCAD. The results showed that the protein formula encoded by *PpCAD* was $C_{1718}H_{2733}N_{471}O_{519}S_{20}$ with molecular mass of 38.9 kDa, theoretical pI of 6.00, and a grand average of hydropathicity (GRAVY) value of 0.071 (positive value for hydrophobicity, the negative for hydrophilicity). Therefore, PpCAD was supposed as a hydrophobic protein. Hydrophobicity, to some extent, affects the three-dimensional conformation of protein. Meanwhile, PpCAD was presumed to be a stable protein due to the instability index of 26.34. Analysis of amino acid composition indicated that PpCAD was mainly made of hydrophobic (nonpolar) amino acids (50.9%), and the rest was polar amino acids. More specifically, the percentages of polar uncharged amino acids, polar positive amino acids

(basic amino acid) and polar negative amino acids (acidic amino acid) were 25.6%, 12.6%, 11.0%, respectively. In the analysis of SignalP and TargetP, no signal peptides could be detected in PpCAD, and it was hypothesized to be a nonsecretory protein.

Prediction of PpCAD membrane-spanning regions was performed by TMpred online program (TM-helix length: 17–33). The results showed that PpCAD included five possible transmembrane helices (two for inside-to-outside helix (i-o), three for outside-to-inside helix (o-i), table 1). The centres of two i-o helices were predicted at the location of I^{186} and L^{281}, with the score 437 and 676, respectively. Similarly, there were three centres existing in o-i helices, which were located at V^{95}, G^{185}, L^{281} and the scores of them were 788, 657 and 976. As a result, PpCAD were predicted to be a transmembrane protein since the scores of four membrane-spanning regions were above 500, which is a standard value considered to be significant.

Phylogenetic analysis of CAD proteins

The PpCAD amino acid sequence was used to construct a phylogenetic NJ tree, along with 61 CAD sequences from model plants *Arabidopsis, Oryza, Populus,* as well as 21 other species (figure 3). Phylogenetic analysis showed that CADs were divided into three major classes based on their amino acid sequences. The distribution of these three classes was supported by relatively high bootstrap value. Class I was comprised of the sequences from monocots, eudicots and gymnosperms, which could be further subdivided into three groups. Class II and class III contained CAD sequences from monocots and eudicots of angiosperm. A majority of sequences from class I were previously proved to be associated with lignin biosynthesis, such as OsCAD2, SbCAD2, AtCAD5 and PoptrCAD4. It was speculated that AtCAD4 of *Arabidopsis CAD* gene family functions mainly in formation of *p*-coumaryl and coniferyl alcohol, whereas AtCAD5 could be used to normally generate all three monolignols, including sinapyl alcohol (Kim *et al.* 2004). Class II included *CAD* sequences only from angiosperm. AtCAD7 and AtCAD8 from class II were

clustered into a small branch, obtaining the support of 100% bootstrap value. They had been preliminarily proved to be probably involved in lignin biosynthesis (Kim *et al.* 2007). *OsCAD7* in class II had been indicated to be a causal gene of *flexible culm 1* (*fc1*), and it had been shown to affect the culm strength and lignin content of *O. sativa* (Li *et al.* 2009). It is noteworthy that sinapyl alcohol dehydrogenase of *P. trichocarpa* (*PoptrSAD*, previously named as *PoptrCAD10*) was also a member of this class (Li *et al.* 2001; Barakat *et al.* 2009). Interestingly, SAD activity could not be detected in *Arabidopsis* and *Oryza* (Barakat *et al.* 2009). Class III represented CAD sequences from *O. sativa* (*OsCAD1* and *OsCAD4*), *P. trichocarpa* (*PoptrCAD2*), *Arabidopsis* (*AtCAD1*) and *Gossypium* (*GhCAD5*). However, the functions of class III members are poorly understood.

PpCAD was grouped in class I of monocot cluster. The closest *in silico* matches with PpCAD protein were *S. bicolor, P. virgatum* and *Z. mays*, which showed an identity of nearly 90% at the amino acid level (table 2). The identity between PpCAD and OsCAD2 was 87%, whereas the identity to AtCAD5 and PtaCAD was 70 and 68%, respectively. It was worth mentioning that OsCAD2, AtCAD5 and PtaCAD were regarded as typical bona fide CADs in monocots, eudicots and gymnosperms and were clustered into same class with PpCAD. This close relationship might be indicative of similar physiological roles for this class. Thus, the PpCAD was predicted as a key enzyme responsible for *P. purpureum* lignin synthesis by phylogenetic analysis (table 3).

Alignment analysis of CAD proteins

Some amino acid sequences of bona fide CADs from class I were aligned together with PpCAD (figure 4), obtaining the overall identity of 28.4%. It was not difficult to find that PpCAD contained a highly conserved motif, GLGGVG, which was named NADP(H) cofactor-binding domain and located at amino acids 188–193. A structure of α-helix was formed behind this motif because the procedure of binding $NADP^+$ needed a helical dipole (Youn *et al.* 2006). Two other conserved motif patterns of CADs, Zn-catalytic centre ^{68}GHE(X)VG(X)VX3G(X)2V^{82}, and Zn-binding site

Table 2. Sequence similarities between different CADs.

	PpCAD	PvCAD1	SbCAD2	ZmCAD1	OsCAD2	TaCAD1	AtCAD5	PtaCAD	PoptrSAD	AtCAD7
PvCAD1	91									
SbCAD2	90	92								
ZmCAD1	90	91	95							
OsCAD2	87	86	86	87						
TaCAD1	86	85	85	86	87					
AtCAD5	70	70	72	73	73	73				
PtaCAD	68	67	66	67	66	66	65			
PoptrSAD	50	49	49	50	50	51	52	54		
AtCAD7	49	48	48	49	48	49	45	51	70	
tCAD1	41	41	41	41	40	42	43	46	48	45

Numbers represent the percentages of CAD amino acids similarity.

Table 3. *CAD* genes used in phylogenetic and alignment analysis.

| Gene	Accession no.	Species	Database
AcCAD	D13991	*Aralia cordata*	GenBank
AtCAD1	AY288079	*Arabidopis thaliana*	GenBank
AtCAD2	AY302077	*A. thaliana*	GenBank
AtCAD3	AY302078	*A. thaliana*	GenBank
AtCAD4	AY302081	*A. thaliana*	GenBank
AtCAD5	AY302082	*A. thaliana*	GenBank
AtCAD6	AY302075	*A. thaliana*	GenBank
AtCAD7	AY302079	*A. thaliana*	GenBank
AtCAD8	AY302080	*A. thaliana*	GenBank
AtCAD9	AY302076	*A. thaliana*	GenBank
BmCAD	GU985522	*Bambusa multiplex*	GenBank
CaCAD	EU616553	*Capsicum annuum*	GenBank
CsCAD1	HQ880207	*Camellia sinensis*	GenBank
CsCAD3	HQ880208	*C. sinensis*	GenBank
EgCAD	X65631	*Eucalyptus gunnii*	GenBank
EglCAD	AF038561	*E. globulus*	GenBank
FaCAD1a	AF188292	*Festuca arundinacea*	GenBank
FaCAD1b	AF188293	*F. arundinacea*	GenBank
FaCAD2a	AF188294	*F. arundinacea*	GenBank
FaCAD2b	AF188295	*F. arundinacea*	GenBank
GhCAD1	EU281304	*Gossypium hirsutum*	GenBank
GhCAD3	FJ376601	*G. hirsutum*	GenBank
GhCAD5	FJ848868	*G. hirsutum*	GenBank
GhCAD6	EU281305	*G.hirsutum*	GenBank
GmCAD1	HQ651807	*Glycine max*	GenBank
LlCAD	FJ008133	*Leucaena leucocephala*	GenBank
LpCAD1	AF472591	*Lolium perenne*	GenBank
LpCAD2	AF472592	*L.perenne*	GenBank
MsaCAD1	AF083333	*Medicago sativa*	GenBank
MsaCAD2	AF083332	*M. sativa*	Genbank
NtCAD14	X62343	*Nicotiana tabacum*	GenBank
NtCAD19	X62344	*N. tabacum*	GenBank
OsCAD1	Os10g11810	*Oryza sativa*	TBD
OsCAD10	Os09g23550	*O. sativa*	TBD
OsCAD11	Os09g23560	*O. sativa*	TBD
OsCAD12	Os03g12270	*O. sativa*	TBD
OsCAD2	Os02g09490	*O. sativa*	TBD
OsCAD3	Os10g29470	*O. sativa*	TBD
OsCAD4	Os11g40690	*O. sativa*	TBD
OsCAD5	Os08g16910	*O. sativa*	TBD
OsCAD6	Os04g15920	*O. sativa*	TBD
OsCAD7	Os04g52280	*O. sativa*	TBD
OsCAD8	Os09g23530	*O. sativa*	TBD
OsCAD9	Os09g23540	*O. sativa*	TBD
PaCAD	X72675	*Picea abies*	GenBank
PoptrCAD1	estExt_fgenesh4_pg.C_LG_I2533	*Populus trichocarpa*	JGI
PoptrCAD2	estExt_fgenesh4_pg.C_LG_XVI0159	*P. trichocarpa*	JGI
PoptrCAD3	estExt_fgenesh4_pm.C_LG_VI0462	*P. trichocarpa*	JGI
PoptrCAD4	estExt_Genewise1_v1.C_LG_IX2359	*P. trichocarpa*	JGI
PoptrSAD	grail3.0004034803	*P. trichocarpa*	JGI
PpCAD	HQ840705	*Pennisetum purpureum*	GenBank
PtaCAD	Z37991	*Pinus taeda*	GenBank
PvCAD1	GU045611	*Panicum virgatum*	GenBank
PvCAD2	GU045612	*P. virgatum*	GenBank
SbCAD2	JF754973	*Sorghum bicolor*	GenBank
ShCAD1	L36823	*Stylosanthes humilis*	GenBank
ShCAD3	L36456	*S. humilis*	GenBank
SoCAD	AJ231135	*Saccharum officinarum*	GenBank
TaCAD1	GU563724	*Triticum aestivum*	GenBank
ZmCAD1	AJ005702	*Zea mays*	GenBank
ZmCAD2	Y13733	*Z. mays*	GenBank
ZmCAD3	NM_001112184	*Z. mays*	GenBank
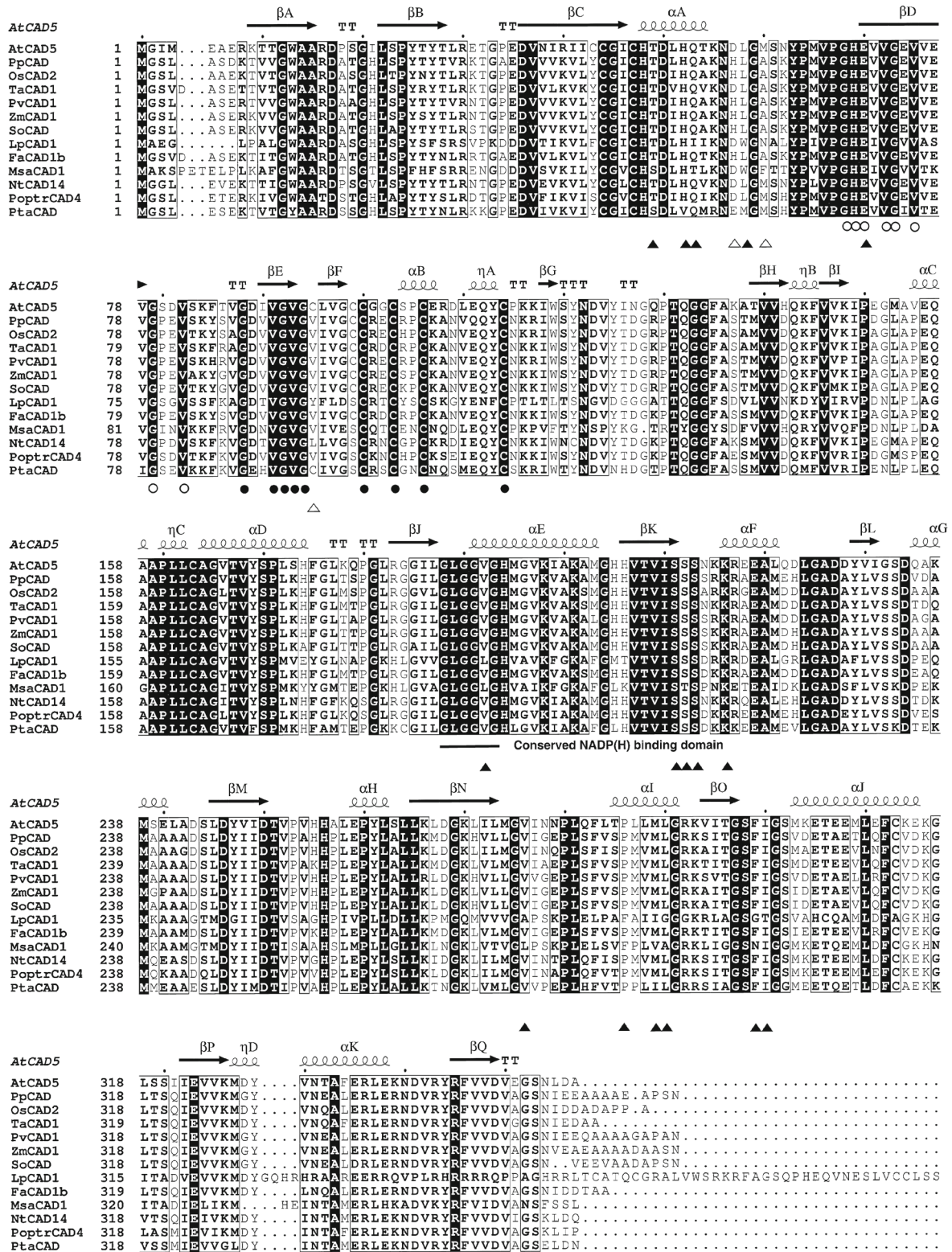
Figure 4. Amino acid sequence alignment of PpCAD and other experimentally proved CADs from class I. Multiple alignments were performed in DNAMAN and figured by ESpript. Elements of secondary structure are indicated at the top of the alignment referring to protein structure of *Arabidopsis* AtCAD5 (PBD number: 2cf5). Allows represent β-sheet (extended strand), helices represent α-helix and TT represent turns. Dark regions indicate identical residues and dots. Hollow boxes indicate similar residue arrangement. The conserved NADP(H) binding domain, GLGGV(L)G, is highlighted by a bold line. Other conserved motif patterns are indicated below the alignment: Zn-catalytic centre is marked by ◯, and Zn-binding site by ●. Triangles represent similar key amino acids between PpCAD and AtCAD5, ▲ indicate complete similarity and △ indicate distinction.
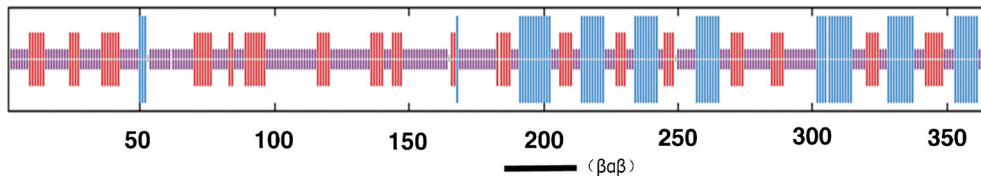
Figure 5. The predicted secondary structure of PpCAD. Numbers below show the position of amino acid. Blue rectangles represent helices structure of α-helix, and red rectangles represent β-sheet. A standard Rossmann foldable structure is underlined.
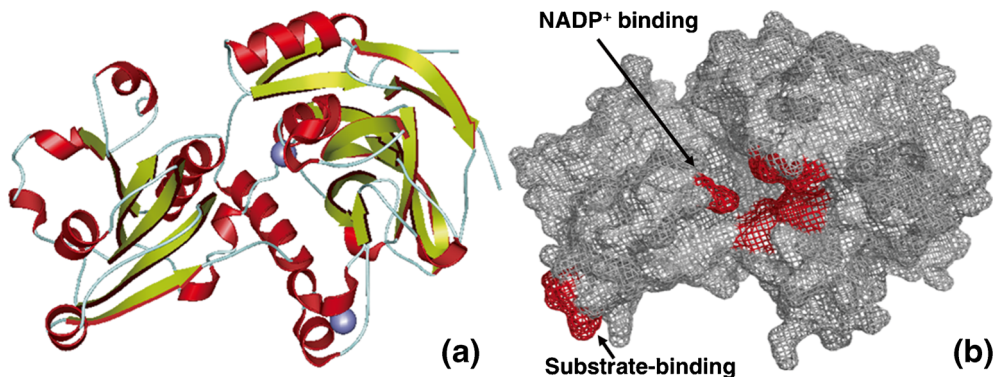
Figure 6. *In silico* modelling of three-dimensional structure of PpCAD. Protein in PDB format was analysed by SWISS-MODEL using AtCAD5 (PBD number: 2c5f) as a template, and modelled using Antheprot 3D and PyMOL softwares. (a) The predicted three-dimensional structure of PpCAD. Red helices represent α-helix, yellow arrows represent β-sheet, and blue filaments represent random coils. Positions of Zn binding are showed as grey balls. (b) A mesh model showing the surface morphology of PpCAD protein. Red areas denote putative substrate-binding and NADP$^+$ binding cavity.
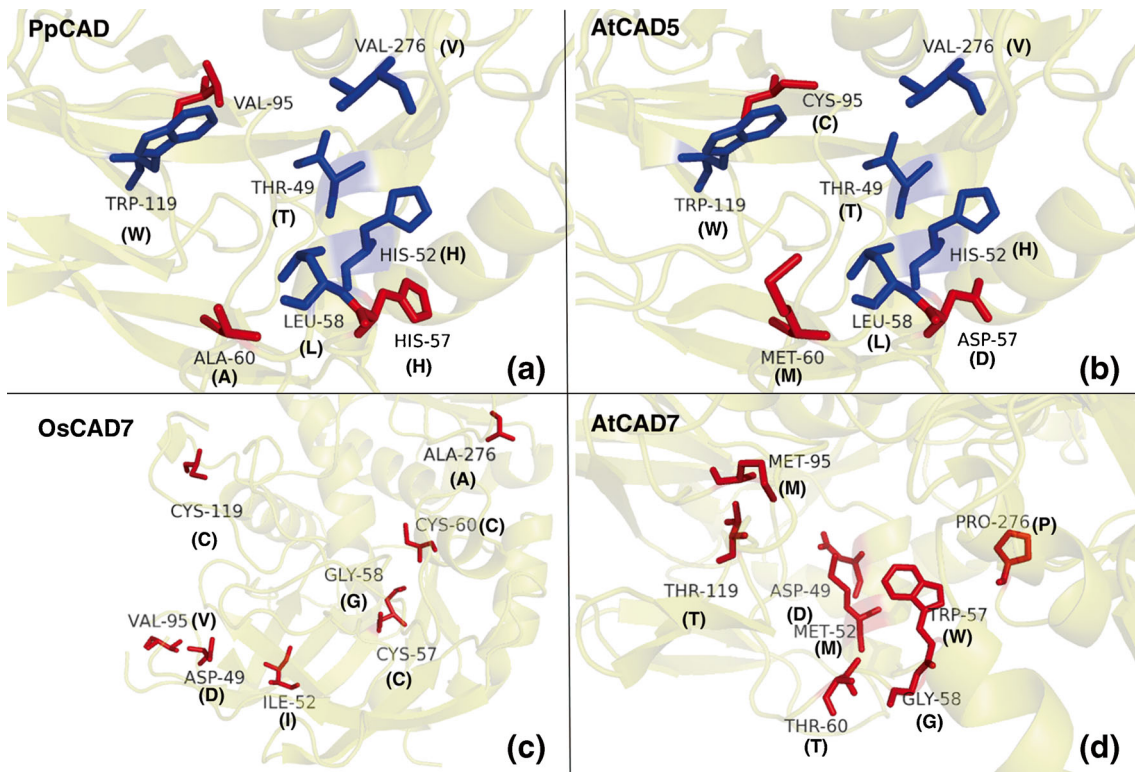
Figure 7. Structural comparison of predicted active residues of PpCAD (a), AtCAD5 (b), AtCAD7 (c) and OsCAD7 (d). The template for AtCAD7 and OsCAD7 was PotSAD (PDB number: 1yqd). The differences between PpCAD and AtCAD5 are shown by red residues in (a) and (b). Only residues that are proved be involved in substrate-binding pocket are shown.

^{88}G(X)$_2$VGVG(X)$_5$C(X)$_2$C(X)$_2$C(X)$_7$C^{114} were identified in the amino acid sequence of *PpCAD* (figure 4). In addition, lignifying CADs require a specific set of key residues for efficient activity against monolignols (Saathoff *et al.* 2012). Take AtCAD5 for example, T^{49} was related to H-shuttle and C^{95} participating in substrate-binding (Youn *et al.* 2006; Bukh *et al.* 2012). Other key amino acid residues of AtCAD5 were H^{52}, Q^{53}, D^{57}, L^{58}, M^{60}, E^{70}, W^{119}, V^{192}, S^{211}, S^{212}, S^{213}, K^{216}, V^{276}, P^{286}, M^{289}, L^{290}, F^{299} and I^{230} (Kim *et al.* 2004; Youn *et al.* 2006). The key residues of CADs showed high similarity throughout the alignment. Their locations are shown in figure 4.

Predicted secondary and three-dimensional structures of PpCAD

The predicted secondary structure of PpCAD was performed in the NPS@ web server. The results showed that α-helix and β-sheet (extended strand) were the main elements of PpCAD (figures 4 and 5), with the percentage of 23.2% and 26.8%, respectively. And they had been interlaced with random coils and turns (figures 5 and 6a), which was similar to the secondary structure prediction of the TaCAD1 from *T. aestivum* (Ma 2010). After analysis by Pfam, *PpCAD* was predicted to belong to a NADP-binding Rossmann fold superfamily and a standard Rossmann foldable structure of $\beta\alpha\beta$ (Rossmann *et al.* 1974) could be detected from the 182 to 212

residues of PpCAD (figure 5), including conserved sequence of NADP(H) binding domain.

AtCAD5 (PDB number: 2cf5) was the most homologous template for PpCAD three-dimensional modelling (figure 6a). Amino acids 6 to 356 of PpCAD showed 71.2% sequence identity with AtCAD5 sequence. Surface morphology of PpCAD protein was modelled in a mesh style. The putative substrate-binding and NADP$^+$ binding cavities, which were identified as two critical structures of bona fide CAD are denoted in red (figure 6b). The key residues of PpCAD were in general very similar to the residues of AtCAD5, and most residues had a similar orientation (figure 7, a&b). A residue of W^{119} resided in both PpCAD and AtCAD5 was likely to stabilize the aromatic ring of cinnamaldehydes through pi-bonding (Saathoff *et al.* 2011). However, three residues were apparently different. In PpCAD, the D^{57}, M^{60} and C^{95} were replaced by H^{57}, A^{60} and V^{95} when compared with AtCAD5. Two CAD-like proteins of AtCAD7 and OsCAD7 were modelled to investigate the distinctions between PpCAD and nonbona fide CADs. In examining the binding pocket models for PpCAD, AtCAD7 and OsCAD7, the differences were more obvious (figure 7, a, c&d). The bona fide CADs always have a H^{57}/L^{58} or D^{57}/L^{58} motif in the binding pocket, which is more variable at the equivalent position in other CAD-like sequences (Saathoff *et al.* 2011). The motif was H^{57}/L^{58} in PpCAD

PpCAD

AtCAD7

Structural alignment
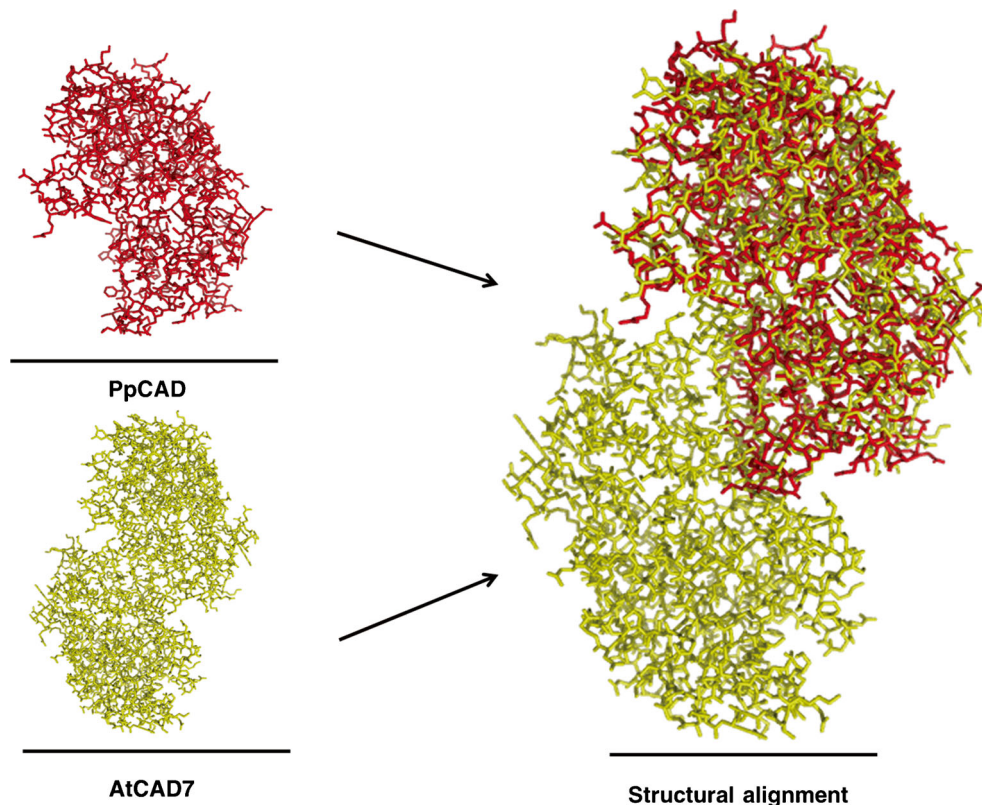
Figure 8. Structural alignment of PpCAD (red) and AtCAD7 (yellow) in the stick models. PDB format was generated by SWISS-MODEL and the structures were modelled and aligned using PyMOL software.

while W^{57}/G^{58} in AtCAD7 and C^{57}/G^{58} in OsCAD7. Significantly, the residues of AtCAD7 and OsCAD7 showed completely irregular orientations in the binding pocket when compared with that of PpCAD (figure 7, a, c&d). The structural alignment of PpCAD and AtCAD7 further showed the similarities and differences between these two kinds of CADs. The major structural difference between PpCAD and AtCAD7 was an additional part in the C-terminal end of AtCAD7 (figure 8).

Discussion

The characters of CADs have been described in many researches, and bona fide CAD is considered as key significance in the lignin biosynthesis (Weng and Chapple 2010). However, no study has been reported on the multipurpose grass *P. purpureum* till now. The aim of the study was to isolate *P. purpureum CAD* gene which could be identified as bona fide *CAD* by bioinformatics analysis.

Characterization of CAD gene in P. purpureum

Previous studies have demonstrated that the *CAD* gene coding sequences of the most plant species range from 1044 to 1131-bp in length (Guo *et al.* 2010). In this research, a 1 342-bp cDNA fragment with 1 011-bp ORF was isolated from *P. purpureum*. It could be translated into an intact protein sequence encoding 366 amino acid residues with a calculated molecular mass of 38.9 kDa, which also varied from 35 to 50 kDa in most other species (Goffner *et al.* 1998; Li *et al.* 2001; Sibout *et al.* 2003). There is such an exception as the molecule mass of OsCAD5 is only 31.4 kDa owing to the absent of 52 amino residues in C-terminal (Tobias and Chow 2005).

After full-length cloning of *PpCAD*, a 3 977-bp DNA sequence with four exons and three introns was obtained. Similar intron–exon pattern could be detected in *OsCAD2* (Tobias and Chow 2005). A very recent report on the evolution of the *CAD/SAD* gene family revealed that *CAD* gene has fewer than six introns in land plants and the patterns of intron–exon structure have great variation between different species (Barakat *et al.* 2009; Guo *et al.* 2010). A 114 bp sequence (exon II) of *PpCAD*, highlighted in figure 2, was proved to be a highly conserved gene structure with function of binding monolignol substrate (Youn *et al.* 2006). *P. trichocarpa* have 16 *CAD* genes and all of them contain the 114-bp sequence (Barakat *et al.* 2009). In 12 *CAD* genes of *Oryza*, the conserved structure can also be detected from all of them (Tobias and Chow 2005), although the position of the sequence located in gene is uncertain. In *Brachypodium distachyon*, the sequence (exon II) is exactly identical in *BdCAD2, BdCAD3, BdCAD4, BdCAD5* and *BdCAD8*, but only a part of the exon in *BdCAD1* and *BdCAD7* (Bukh *et al.* 2012).

Phylogenetic analysis of CAD gene family

The diversity of *CAD* gene family determines the character of multi-function in CAD protein (Ma 2010; Deng *et al.* 2011). A reliable phylogeny of a gene family is expected to give important clues for understanding its evolutionary history and functional differentiation (Guo *et al.* 2010). CAD gene family was divided into three main classes based on their amino acid sequences (figure 3). Class I was so-called bona fide CADs and PpCAD was clustered into this class, suggesting that this protein may have the function of regulating lignification of *P. purpureum*. Class II and class III contained only the sequences from monocot and eudicot angiosperms, indicating that the evolution of these two classes happened before the differentiation of monocots and eudicots. Similar situation had been found by Barakat *et al.* (2009).

In addition, class I could be further divided into three groups, including monocot group (MG), eudicot group (EG) and gymnosperm group (GG), which is consistent with a recent phylogenetic analysis of CADs (Guo *et al.* 2010). The reason of this subdivision may be attributed to the distinct enzyme features between gymnosperms and angiosperms. Gymnosperm CAD has only one form and presents more affinity for coniferyl aldehyde than sinapyl aldehyde, while angiosperm CAD has more isoforms and shows high catalytic activity towards both of them (Tobias and Chow 2005; Ma 2010). Recent study of *Oryza gh2* gene showed that gh2 protein had strong CAD and SAD activities (Zhang *et al.* 2006), which demonstrated again that the same *CAD* gene products in angiosperm plants can synthesize both coniferyl and sinapyl alcohol. The reason of gymnosperm CAD highly specific for the degradation of coniferyl aldehyde is probably due to the existence of SAD, which have been proved specifically to participate in the degradation of sinapaldehyde and can hardly be detected in some other angiosperm plants such as *Arabidopsis* and *Oryza* (Li *et al.* 2001; Baraket *et al.* 2009). The fact that gymnosperm plants contain mainly G monomers whereas angiosperm plants consist of G and S monomers may be a consequence of this distinction (Galliano *et al.* 1993; MacKay *et al.* 1997; Baucher *et al.* 1999), but more specific reasons are underexplored.

Arabidopsis and *Oryza* are two well-studied model plants, and the identification of complete sets of *CAD* genes from them have been reported (Raes *et al.* 2003; Tobias and Chow 2005). AtCAD1 to AtCAD9, and OsCAD1 to OsCAD12 were completely included in our phylogenetic analysis, aiming to annotate general functions of CADs from each class. OsCAD7 of *Oryza* in class II affected the culm strength and lignin content (Li *et al.* 2009). An *Arabidopsis* CAD double mutant (*AtCAD4* and *AtCAD5*) resulted in a phenotype with a limp floral stem at maturity as well as modifications in the pattern of lignin staining (Sibout *et al.* 2005). This deficiency could be partially restored by *AtCAD1* of class III (Eudes *et al.* 2006). The data of previous studies had preliminary indicated that *AtCAD7* and *AtCAD8* may be involved to some extent in lignin biosynthesis (Kim *et al.* 2007), which

were grouped into a small branch in class II. Meanwhile, overexpression of *AtCAD7* could be detected in *AtCAD2* and *AtCAD6* double mutant (Sibout *et al.* 2005), suggesting that the absent expression of some *CAD* gene family members would be compensated by others. And it is also indicated that lignin formation are not dependent on a single *CAD*, due to the presence of functionally redundant and complexly metabolic networks in different tissues, organs and cell types in a plant (Kim *et al.* 2004). Additionally, the mechanism of lignin biosynthesis might differ among different species (Hirano *et al.* 2012).

AtCAD7 has been characterized as a benzyl alcohol dehydrogenase with low catalytic activity against monolignol compounds in an early study (Somssich *et al.* 1996), and it shows a high similarity with PoptrSAD in amino acid sequence (table 2). *AtCAD2, AtCAD3, AtCAD6* and *AtCAD9* from class III may have the function of encoding mannitol dehydrogenases (Sibout *et al.* 2005). This result confirmed that some *CAD*-like genes may participate in lignin biosynthesis and perform bona fide *CAD* gene functions. Phylogenetic analysis showed that all the other CADs from *Arabidopsis* and *Oryza*, except bona fide CAD (AtCAD4, AtCAD5 and OsCAD2), were clustered into classes II and III. It is necessary to investigate the evolutionary relationships between the bona fide *CAD* genes and the *CAD*-like genes, and their functional intercoordination in plant growth, especially in lignin biosynthesis.

PpCAD: a potential enzyme responsible for lignin synthesis

The amino acid sequence of PpCAD showed a high similarity with some CADs in class I, especially with sequences from monocot cluster (such as PvCAD1, SbCAD2, ZmCAD1, OsCAD2 and TaCAD1; table 2). Meanwhile, the alignment of PpCAD protein sequence showed a high identity with 12 other selected CADs from class I. Three key conserved motifs have been detected in PpCAD sequences according to the alignment: Zn-catalytic centre of GHE(X)VG(X)VX$_3$G(X)$_2$V, Zn-binding site of G(X)$_2$VGVG(X)$_5$C(X)$_2$C(X)$_2$C(X)$_7$C, and NADP(H) cofactor binding domain of GLGGV(L)G (Youn *et al.* 2006; Saballos *et al.* 2009; Bukh *et al.* 2012). The three domains could be found in all aligned sequences (figure 4). It is known that conserved domains of amino acid sequences decide the function of proteins. The pattern of NADP(H) binding domain in some researches can be further identified as G(X)$_3$G(X)$_2$GLGG(X)GH(X)$_2$VK(X)$_2$K(X)$_2$G(X)VTV(X)S (X)S(X)$_2$K (Bukh *et al.* 2012), which is located between β-sheet-J and β-sheet-K in our alignment. The position is known for Rossmann foldable structure, connected with an α-helix (αF) used to combine with its cofactor NADP(H) (Rossmann *et al.* 1974, figure 5), which can also be detected in PpCAD amino acid sequence.

The key amino acid residues in the catalytic and substrate-binding site, respectively, were highly conserved in PpCAD and other aligned sequences. Every substrate-binding pocket has its own critical effects. In our analysis, the key amino acid residues of PpCAD were very similar to the residues found in AtCAD5, although three of them (H^{57}, A^{60} and V^{95}) were different (figure 7, a&b). Similar situation had been found in other monocot CADs (Saballos *et al.* 2009), and the distinctions might be related to the different substrate specificity between eudicot and monocot CADs (Ma 2010).

When nonbona fide CAD sequences including AtCAD7 and OsCAD7 were modelled and compared with PpCAD, the differences of their active residues were obvious. A motif of H^{57}/L^{58} or D^{57}/L^{58}, which was conserved in monocot or eudicot CADs, could not be detected in OsCAD7 and AtCAD7 protein sequences. The W^{119} was identified as an invariant residue in PpCAD, but not conserved in both OsCAD7 and AtCAD7 (figure 7, c&d). It was evident that the structure of PpCAD was quite different from nonbona fide CADs (AtCAD7) after structural alignment (figure 8). PpCAD protein had the necessary features of bona fide CADs, but showed critical differences with nonbona fide CADs in both key residues and three-dimensional structures. Thus, we can preliminarily prove that the PpCAD is an enzyme responsible for lignin synthesis.

P. purpureum, a multi-purpose grass, possesses the properties of fast growth and high biomass production (Xie *et al.* 2009). Diminishing lignin recalcitrance and improving lignin extractability are of great benefits to utilization of this kind of plant. In this research, a *CAD* gene encoding cinnamyl alcohol dehydrogenase was successfully cloned and fully sequenced, and it have been proved as a 'lignification' gene of *P. purpureum* by bioinformatics analysis. Nevertheless, more in-depth investigations are required to determine *in planta* functions as well as *in vitro* biochemical characters of PpCAD. The functional studies of CADs have great significance, which is not only embodied in understanding the formation mechanism of plant lignin, but also in providing plants with expected characters by transgenic methods. At the same time, it is necessary to isolate other *CAD* gene family members from *P. purpureum* because different family members may perform distinct functions in the process of plant growth, and all of them have more or less participated in lignin biosynthesis, which should be considered thoroughly in lignin genetic modification.

References

Arnold K., Bordoli L., Kopp J. and Schwede T. 2006 The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. *Bioinformatics* **22**, 195–201.

Barakat A., Bangniewska-Zadworna A., Choi A., Plakkat U., Diloreto D. S., Yellanki P. *et al.* 2009 The cinnamyl alcohol dehydrogenase gene family in *Populus*: phylogeny, organization, and expression. *BMC Plant Biol.* **9**, 26.

Baucher M., Chabbert B., Pilate G., Van Doorsselaere J., Tollier M. T., Petit-Conil M. *et al.* 1996 Red xylem and higher lignin extractability by down-regulating a cinnamyl alcohol dehydrogenase in poplar. *Plant Physiol.* **112**, 1479–1490.

Baucher M., Bernard-Vailhe M. A., Chabbert B., Besle J. M., Opsomer C., Van Montagu M. and 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**, 437–447.

Boerjan W., Ralph J. and Baucher M. 2003 Lignin biosynthesis. *Annu. Rev. Plant Biol.* **54**, 519–546.

Brill E. M., Abrahams S., Hayes C. M., Jenkins C. L. D. and Watson J. M. 1999 Molecular characterization and expression of a wound-inducible cDNA encoding a novel cinnamyl-alcohol dehydrogenase enzyme in lucerne. *Plant Mol. Biol.* **41**, 279–291.

Bukh C., Nord-Larsen P. H. and Rasmussen S. K. 2012 Phylogeny and structure of the cinnamyl alcohol dehydrogenase gene family in *Brachypodium distachyon*. *J. Exp. Bot.* **63**, 6223–6236.

Chen L., Auh C. K., Dowling P., Bell J., Chen F., Hopkins A. *et al.* 2003 Improved forage digestibility of tall fescue (*Festuca arundinacea*) by transgenic down regulation of cinnamyl alcohol dehydrogenase. *Plant Biotechnol. J.* **1**, 437–449.

Deng W. W., Zhang M., Wu J. Q., Jiang Z. Z., Tang L., Li Y. Y. *et al.* 2011 Molecular cloning, functional analysis of three cinnamyl alcohol dehydrogenase (CAD) genes in the leaves of tea plant, *Camellia sinensis*. *J. Plant Physiol.* **170**, 272–282.

de Morais R. F., Quedsds D. M., Reis V. M., Urquiaga S., Alves B. J. R. and Boddet R. M. 2011 Contribution of biological nitrogen fixation to elephant grass (*Pennisetum purpureum* Schum.) *Plant Soil* **356**, 23–34.

Dixon R. A., Chen F., Guo D. and Parvathi K. 2001 The biosynthesis of monolignols: a 'metabolic grid', or independent pathways to guaiacyl and syringyl units? *Phytochemistry* **57**, 1069–1084.

Doyle J. and Doyle J. L. 1987 A rapid DNA isolation method for small quantities of fresh tissues. *Phytochem. Bull.* **19**, 11–15.

Emanuelsson O., Nielsen H., Brunak S. and von Heijne G. 2000 Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* **300**, 1005–1016.

Eudes A., Pollet B., Sibout R., Do C. T., Séguin A., Lapierre C. *et al.* 2006 Evidence for a role of AtCAD 1 in lignification of elongating stems of Arabidopsis thaliana. *Planta* **225**, 23–39.

Eudes A., George A., Mukerjee P., Kim J. S., Pollet B., Benke P. I. *et al.* 2012 Biosynthesis and incorporation of side-chain-truncated lignin monomers to reduce lignin polymerization and enhance saccharification. *Plant Biotechnol. J.* **10**, 609–620.

Finn R. D., Mistry J., Tate P., Coggill P., Hegar A., Pollington J. E. *et al.* 2010 The pfam protein families database. *Nucleic Acids Res.* **38**, 211–222.

Fu C., Xiao X., Xi Y., Ge Y., Chen F., Bouton J. *et al.* 2009 Down-regulation of cinnamyl alcohol dehydrogenase (CAD) leads to improved saccharification efficiency in switchgrass. *Bioenerg. Res.* **4**, 153–164.

Fu C, Mielenz J. R., Xiao X., Ge Y., Hamilton C. Y., Rodriguez M. *et al.* 2011 Genetic manipulation of lignin reduces recalcitrance and improves ethanol production from switchgrass. *Proc. Natl. Acad. Sci. USA* **108**, 3803–3808.

Fornalé S., Capellades M., Encina A., Wang K., Irar S., Lapierre C. *et al.* 2012 Altered lignin biosynthesis improves cellulosic bioethanol production in transgenic maize plants down-regulated for cinnamyl alcohol dehydrogenase. *Mol. Plant* **5**, 817–830.

Galliano H., Cabane M., Eckerskorn C., Lottspeich F., Sandermann H. and Ernst D. 1993 Molecular cloning, sequence analysis and elicitor-/ozone-induced accumulation of cinnamyl alcohol dehydrogenase from Norway spruce (*Picea abies* L.) *Plant Mol. Biol.* **23**, 145–156.

Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D. *et al.* 2005 *The proteomics protocols handbook.* pp. 571–607. Humana Press, New York, USA.

Goffner D., Van Doorsselaere J., Yahiaoui N., Samaj J., Grima-Pettenati J. and Boudet A. M. 1998 A novel aromatic alcohol dehydrogenase in higher plants: Molecular cloning and expression. *Plant Mol. Biol.* **36**, 755–765.

Gouet P., Courcelle E., Stuart D. I. and Métoz F. 1999 ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15**, 305–308.

Guo D. M., Ran J. H. and Wang X. Q. 2010 Evolution of the cinnamyl/sinapyl alcohol dehydrogenase(CAD/SAD) gene family: The emergence of real Lignin is associated with the origin of bona fide CAD. *J. Mol. Evol.* **71**, 202–218.

Halpin C., Knight M. E., Grima-Pettenati J., Goffner D., Boudet A. and Schuch W. 1992 Purification and characterization of cinnamyl alcohol dehydrogenase from tobacco stems. *Plant Physiol.* **98**, 12–16.

Hirano K., Aya K., Kondo M., Okuno A., Morinak Y. and Matasuoka M. 2012 OsCAD2 is the major CAD gene responsible for monolignol biosynthesis in rice culm. *Plant Cell Rep.* **31**, 91–101.

Hisano H., Nandakumar R. and Wang Z. Y. 2009 Genetic modification of lignin biosynthesis for improved biofuel production. *In Vitro Cell. Dev. Biol.-Plant* **45**, 306–313.

Hofmann K. and Stoffel W. 1993 TMbase-A database of membrane spanning protein segments. *Biol. Chem. Hoppe-Seyler* **374**, 166.

Jackson L. A., Shadle G. L., Zhou R., Nakashima J., Chen F. and Dixon C. R. 2008 Improving saccharification efficiency of alfalfa stems through modification of the terminal stages of monolignol biosynthesis. *Bioenerg. Res.* **1**, 180–192.

Kim S. J., Kim M. R., Bedgar D. L., Moinuddin S. G. A., Cardenas, C. L., Davin L. B. *et al.* 2004 Functional reclassification of the putative cinnamyl alcohol dehydrogenase multigene family in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **101**, 1455–1460.

Kim S. J., Kim K. W., Cho M. H., Franceschi V. R., Davin L. B. and Lewis N. G. 2007 Expression of cinnamyl alcohol dehydrogenase and their putative homologues during *Arabidopsis thaliana* growth and development: lessons for database annotation? *Phytochemistry* **68**, 1957–1974.

Knight M. E., Halpin C. and Schuch W. 1992 Identification and characterization of cDNA clones encoding cinnamyl alcohol dehydrogenase from tobacco. *Plant Mol. Biol.* **19**, 739–810.

Li L. G., Cheng X. F., Leshkevich J., Umezawa T., Harding S. A. and Chiang V. L. 2001 The last step of syringyl monolignol biosynthesis in angiosperms is regulated by a novel gene encoding sinapyl alcohol dehydrogenase. *Plant Cell* **13**, 1567–1585.

Li X., Yang Y., Yao J., Chen G., Li X. H., Zhang Q. *et al.* 2009 FLEXIBLE CULM 1 encoding a cinnamyl-alcohol dehydrogenase controls culm mechanical strength in rice. *Plant Mol. Biol.* **69**, 685–697.

Lynch D., Lidgett A., Mclnnes R., Huxley H., Jones E., Mahoney N. *et al.* 2002 Isolation and characterisation of three cinnamyl alcohol dehydrogenase homologue cDNAs from perennial ryegrass (*Lolium perenne* L.) *J. Plant Physiol.* **159**, 653–660.

Ma Q. H. 2010 Functional analysis of a cinnamyl alcohol dehydrogenase involved in lignin biosynthesis in wheat. *J. Exp. Bot.* **61**, 2735–2744.

MacKay J. J., O'Malley D. M, Presnell T., Booker F. L., Campbell M. M., Whetten R. W. *et al.* 1997 Inheritance, gene expression, and lignin characterization in a mutant pine deficient in cinnamyl alcohol dehydrogenase. *Proc. Natl. Acad. Sci. USA* **94**, 8255–8260.

Madakadze I. C., Masamvu T. M., Radiotis T., Li J. and Smith D. L. 2010 Evaluation of pulp and paper making characteristics of

elephant grass (*Pennisetum purpureum* Schum.) and switchgrass (*Panicum virgatum* L.) *Afr. J. Environ. Sci. Techonol.* **4**, 465–470.

Matthews-Amune O. C. and Kakulu S. 2012 Determination of heavy metals in forage grasses (carpet grass (*Axonopus ompressus*), guinea grass (Panicum maximum) and elephant grass (*Pennisetum purpureum*)) in the vicinity of Itakpe Iron Ore Mine, Nigeria. *Int. J. Pure Appl. Sci. Technol.* **13**, 16–25.

Mansell R. L., Gross G. G., Stoeckigt J., Franke H. and Zenk M. H. 1974 Purification and properties of cinnamyl alcohol dehydrogenase from higher plants involved in lignin biosynthesis. *Phytochemistry* **13**, 2427–2436.

O'Malley D. M., Porter S. and Sederoff R. R. 1992 Purification, characterization, and cloning of cinnamyl alcohol dehydrogenase in loblolly pine (*Pinus taeda* L). *Plant Physiol.* **98**, 1364–1371.

Pandey B., Pandey V. P. and Dwivedi U. N. 2011 Cloning, expression, functional validation and modeling of cinnamyl alcohol dehydrogenase isolated from xylem of *Leucaena leucocephala*. *Protein Expr. Purif.* **79**, 197–203.

Peter G. and Neale D. 2004 Molecular basis for the evolution of xylem lignification. *Curr. Opin. Plant Biol.* **7**, 737–742.

Petersen T. N., Brunak S., von Heijine G. and Nielson H. 2011 SignalP 4.0: discriminating signal peptides from transmembrane region. *Nat. Methods* **8**, 785–786.

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

Ragauskas A., Williams C., Davison B., Britovsek G., Cairney J. and Eckert C. A. 2006 The path forward for biofuels and biomaterials. *Science* **311**, 484–489.

Rossmann M. G., Moras D. and Olsen K. W. 1974 Chemical and biological evolution of a nucleotide-binding protein. *Nature* **250**, 194–199.

Saathoff A. J., Hargrove M. S., Haas E. J., Tobias C. M., Twigg P., Sattler S. *et al.* 2012 Switchgrass PviCAD1: Understanding residues important for substrate preferences and activity. *Appl. Biochem. Biotechnol.* **168**, 1086–1100.

Saathoff A. J., Tobias C. M., Sattler S. E, Haase J., Twigg P. and Sarath G. 2011 Switchgrass contains two cinnamyl alcohol dehydrogenases involved in lignin formation. *Bioenerg. Res.* **4**, 120–133.

Saballos A., Ejeta G., Sanchez E., Kang C. and Vermerris W. 2009 A genome-wide analysis of the cinnamyl alcohol dehydrogenase family in sorghum [*Sorghum bicolor* (L.) Moench] identifies SbCAD2 as the *brown midrib6* gene. *Genetics* **181**, 783–795.

Sibout R., Eudes A., Pollet B., Goujon T., Mila I., Granier F. *et al.* 2003 Expression pattern of two paralogs encoding cinnamyl

alcohol dehydrogenases in *Arabidopsis*. Isolation and characterization of the corresponding mutants. *Plant Physiol.* **132**, 848–860.

Sibout R., Eudes A., Mouille G., Pollet B., Lapierre C., Jouanin L. *et al.* 2005 CINNAMYL ALCOHOL DEHYDROGENASE-C and -D are the primary genes involved in lignin biosynthesis in the floral stem of *Arabidopsis*. *Plant Cell* **17**, 2059–2076.

Somssich I. E., Wernert P., Kiedrowski S. and Hahlbrock K. 1996 *Arabidopsis thaliana* defense-related protein ELI3 is an aromatic alcohol: NADP$^+$ oxidoreductase. *Proc. Natl. Acad. Sci. USA* **93**, 14199–14203.

Tavares C. M., Aubourg S., Lecharny A. and Kreis M. 2000 Organization and structural evolution of four multigene families in *Arabidopsis thaliana*: AtLCAD, AtLGT, AtMYST and AtHD-GL2. *Plant Mol. Biol.* **42**, 703–717.

Tobias C. M. and Chow E. K. 2005 Structure of the cinnamyl-alcohol dehydrogenase gene family in rice and promoter activity of a member associated with lignification. *Planta* **220**, 678–688.

Tsuruta S., Ebina M., Nakagawa H., Kawamura O. and Akashi R. 2007 Isolation and characterization of cDNA encoding cinnamyl alcohol dehydrogenase (CAD) in sorghum (*Sorghum bicolor* (L.) Moench). *Grassland Sci.* **53**, 103—109.

Van Doorsselaere J., Baucher M., Feuillet C., Boudet A. M., Van Montagu M. and Inzé D. 1995 Isolation of cinnamyl alcohol dehydrogenase cDNAs from two important economic species: alfalfa and poplar. Demonstration of a high homology of the gene within angiosperms. *Plant Physiol. Biochem.* **33**, 105–109.

Videira S. S., Oliveira D. M. D, de Morais R. F., Borges W. L., Baldani V. L. and Baldani J. I. 2012 Genetic diversity and plant growth promoting traits of diazotrophic bacteria isolated from two *Pennisetum purpureum* Schum. genotypes grown in the field. *Plant Soil* **356**, 51–66.

Weng J. K. and Chapple C. 2010 The origin and evolution of lignin biosynthesis. *New Phytol.* **187**, 273–285.

Xie X. M., Zhou F, Zhang X. Q. and Zhang J. M. 2009 Genetic variability and relationship between MT-1 elephant grass and closely related cultivars assessed by SRAP marker. *J. Genet.* **88**, 281–290.

Youn B., Camacho R., Moinuddin S. G. A., Lee C., Davin L. B., Lewis N. G. *et al.* 2006 Crystal structures and catalytic mechanism of the *Arabidopsis* cinnamyl alcohol dehydrogenases AtCAD5 and AtCAD4. *Org. Biomol. Chem.* **4**, 1687–1697.

Zhang K., Qian Q., Huang Z., Wang Y., Li M., Hong L. *et al.* 2006 *GOLD HULL AND INTERNODE2* encodes a primarily multifunctional cinnamyl alcohol dehydrogenase in rice. *Plant Physiol.* **140**, 972–983.