Phylogeny and structure of the cinnamyl alcohol dehydrogenase gene family in Brachypodium distachyon

Christian Buhk*, Pia Haugaard Nord-Larsen*, Søren K. Rasmussen†

Department of Plant and Environmental Sciences, University of Copenhagen, DK-1871 Frederiksberg C, Denmark

* These authors contributed equally to this work.
† To whom correspondence should be addressed. E-mail: skr@life.ku.dk

Abstract

Cinnamyl alcohol dehydrogenase (CAD) catalyses the final step of the monolignol biosynthesis, the conversion of cinnamyl aldehydes to alcohols, using NADPH as a cofactor. Seven members of the CAD gene family were identified in the genome of Brachypodium distachyon and five of these were isolated and cloned from genomic DNA. Semi-quantitative reverse-transcription PCR revealed differential expression of the cloned genes, with BdCAD5 being expressed in all tissues and highest in root and stem while BdCAD3 was only expressed in stem and spikes. A phylogenetic analysis of CAD-like proteins placed BdCAD5 on the same branch as bona fide CAD proteins from maize (ZmCAD2), rice (OsCAD2), sorghum (SbCAD2) and Arabidopsis (AtCAD4, 5). The predicted three-dimensional structures of both BdCAD3 and BdCAD5 resemble that of AtCAD5. However, the amino-acid residues in the substrate-binding domains of BdCAD3 and BdCAD5 are distributed symmetrically and BdCAD5 is similar to that of poplar sinapyl alcohol dehydrogenase (PotSAD). BdCAD3 and BdCAD5 expressed and purified from Escherichia coli both showed a temperature optimum of about 50 °C and molar weight of 49kDa. The optimal pH for the reduction of coniferyl aldehyde were pH 5.2 and 6.2 and the pH for the oxidation of coniferyl alcohol were pH 8 and 9.5, for BdCAD3 and BdCAD5 respectively. Kinetic parameters for conversion of coniferyl aldehyde and coniferyl alcohol showed that BdCAD5 was clearly the most efficient enzyme of the two. These data suggest that BdCAD5 is the main CAD enzyme for lignin biosynthesis and that BdCAD3 has a different role in Brachypodium. All CAD enzymes are cytosolic except for BdCAD4, which has a putative chloroplast signal peptide adding to the diversity of CAD functions.

Key words: Brachypodium distachyon (Bd21-3), coniferyl aldehyde, Cinnamyl alcohol dehydrogenase (CAD), gene structure, lignocellulose, recalcitrance, signal peptide.

Introduction

Utilization of lignocellulosic plant material for biofuel production has regained importance in society. In order to be an economically viable solution to biofuel production, it is necessary to develop strategies to overcome the recalcitrance of lignin, which is a limiting factor in the degradation of cellulose into sugars. The current understanding of lignin biosynthesis has been obtained from research in different areas. Improved yields in the pulp and paper industry promoted research in lignocellulose in woody species and, during the 1980s, the phenylalanine pathway providing the monolignols, the building blocks of lignin, was studied as an important part of the plant defence towards pathogens. Thus, manipulation of the lignin biosynthesis pathway has been proposed as a possible solution to reduce recalcitrance. Brown midrib mutants in maize were identified in the 1920s (Jorgenson, 1931) but it was much later that their potential for improving digestibility was realized. Existing mutants in maize and sorghum, known as brown midrib mutants, with altered lignin biosynthesis, have been shown to have increased digestibility (for
review see Barriere et al., 2004; Guillaumie et al., 2007; Sattler et al., 2010). The brown midrib phenotype has been linked to mutations in a cinnamyl alcohol dehydrogenase (CAD) gene in bm6 sorghum (Sattler et al., 2010). In maize, a link between the expression of CAD genes and the brown midrib phenotype was found, but as no CAD mutants have been identified it is speculated that the mutation reside in a transcription factor (Guillaumie et al., 2007). The CAD-like gene family has been investigated in a number of plant species such as sorghum (Saballos et al., 2009; Sattler et al., 2009), rice (Tobias and Chow, 2005; Li et al., 2009), switchgrass (Fu et al., 2011; Saathoff et al., 2011), poplar (Lapiere et al., 2000, 2004; Barakat et al., 2009, 2010), pine (MacKay et al., 1995), white spruce (Bedon et al., 2009), Arabidopsis (Kim et al., 2004, 2007; Sibout et al., 2003, 2005), tobacco (Halpin et al., 1994; Damiani et al., 2005), sweet potato (Kim et al., 2010), wheat (Ma, 2010), and maize (Halpin et al., 1998; Marita et al., 2003; Guillaumie et al., 2007).

The biosynthetic pathway of lignin monomers and the genes involved therein have been thoroughly reviewed (Dixon et al., 2001; Boerjan et al., 2003; Vanholme et al., 2008). Lignin is a complex heteropolymer formed by coupling of different monolignol subunits, and p-coumaryl alcohol (H), coniferyl alcohol (G), and sinapyl alcohol (S) are the three most frequent (Boerjan et al., 2003). Lignin content and composition vary between plant species, between different tissues as well as between the different layers of the plant cell wall (Campbell and Sederoff, 1996; Hisano et al., 2009). The lignin of most dicotyledonous angiosperms consists mainly of G- and S-monolignols, whereas monocots also contain H-monolignol (Dixon et al., 2001; Bonawitz and Chapple, 2010). Ten enzymes are believed to be required for the biosynthesis of the three monolignols, with phenylalanine ammonia lyase as the initial enzyme of the pathway and CAD catalysing the last step of the reduction of the three aldehydes to alcohols (Li et al., 2008; Hisano et al., 2009).

Modification of the lignin concentration and/or composition, in order to increase digestibility and saccharification efficiency, has been frequently attempted by manipulating the expression of the different genes in the lignin biosynthesis pathway (for recent reviews see Li et al., 2008; Vanholme et al., 2008). These genes provide the monomeric building blocks and they are the primary target in many research projects aiming at reducing the recalcitrance of lignocellulosic material. This is often referred to as the low-hanging fruit strategy. The most pronounced effect on lignin content is obtained by suppressing genes in the early steps of the pathway, whereas reducing CAD activity rather affects lignin composition according to Kim et al. (2002) and Li et al. (2008). Decreasing the lignin content in order to improve digestibility can result in plants with impaired growth (Chabannes et al., 2001; Chen and Dixon, 2007). However, it appears that manipulating enzymes in the later steps of the pathway, e.g. CAD genes, has less, if any, effect on the plant biomass production (Bonawitz and Chapple, 2010).

CAD catalyses the final step of the monolignol biosynthesis, the conversion of cinnamyl aldehydes to alcohols, using NADPH as a cofactor (Sattler et al., 2010). Early reports on the identification of CAD enzymes suggested that only one form exists in most plant species. However, in Eucalyptus gunnii two CAD isoforms were isolated and named EgCAD1 and EgCAD2 (Goffner et al., 1992) and recent studies have indicated that angiosperms contain a family of CAD-like genes, with nine members in Arabidopsis (Kim et al., 2004), 14 members in sorghum (Saballos et al., 2009), and 12 members in rice (Tobias and Chow, 2005). However, when comparing biochemical properties of EgCAD1 and EgCAD2, EgCAD2 was considered to be the most likely to be involved in lignification (Goffner et al., 1992) and is thus regarded to be a bona fide CAD. Furthermore, the sequence of EgCAD1 (accession CAA61275) isolated from Eucalyptus is not conserved in amino-acid residues which are believed to be essential and characteristic for CAD function, e.g. in the zinc-binding domains (Youn et al., 2006). Based on its amino-acid sequence, EgCAD1 is actually more similar to the cinnamoyl-CoA reductase family acting in the middle of the lignin biosynthetic pathway (Goffner et al., 1998).

In the early stage of identifying CAD-like genes, the sequences ELI-1 (At CAD7) and ELI3-2 (At CAD8) from Arabidopsis were identified but not biochemically characterized. However an ELI3 homologue from celery was isolated and described as a mannitol dehydrogenase (Stoop and Pharr, 1992). The annotation was later changed to benzyl alcohol dehydrogenase showing low catalytic activity against monolignol compounds according to biochemical analysis (Somsisch et al., 1996). The initial mannitol dehydrogenase annotation has thus resulted in misleading annotations in databases because these were not substantiated by reliable biochemical experiments of homologues to CAD-like enzymes (Kim et al., 2007). Proteins with both proven mannitol dehydrogenase activity and known protein structure are short-length dehydrogenases of 265–280 residues and lack Zn-binding sites (see Nüss et al., 2010). CAD-like enzymes are typically mid-length dehydrogenases of 350 amino acids. This illustrates that biochemical characterization of enzymes encoded by cloned genes in order to confirm their putative function (Kim et al., 2004) is needed.

This study identifies seven CAD-like genes and isolated five in Brachypodium distachyon, the model plant for temperate grasses. The relative expressions of these genes in different plant organs as well as in vitro biochemical characterization of the enzymes BdCAD3 and BdCAD5 are presented. As a first step to fully assign functions to the five CAD genes in developing Brachypodium plants, the function of BdCAD3 and BdCAD5 as well as BdCAD4, which has a chloroplast-targeting signal peptide, are discussed. It is anticipated that detailed knowledge of these genes will support a strategy to reduce recalcitrance by modifying lignin composition with minor biomass yield penalty, which subsequently can be handle by plant breeding.

Materials and methods

Plant material

The B. distachyon genotype Bd21-3 was used for all experiments. Plants were grown in a naturally lit greenhouse with standard irrigation and fertilization. Plants were harvested at the seed-filling stage. Harvested plant material were immediately frozen in liquid nitrogen and stored at –80 °C until use.

Isolation of gDNA and RNA and synthesis of cDNA

RNA was extracted from different Brachypodium tissues using a RNeasy Kit (Qiagen, UK), according to the manufacturer’s protocol. RNA was treated with RQ1 RNase-Free DNase (Promega, USA) before
reverse transcription into cDNA using iScript cDNA Synthesis Kit (Bio-Rad, USA) or M-MuLV Reverse Transcriptase RNase H- (Finnzymes, FI) according to the manufacturer’s protocol using a p(dT)₆ primer. DNA was extracted at the immature seed stage from Brachypodium leaf tissue using a DNeasy Plant Mini Kit, according to the manufacturer’s protocol.

Cloning of Brachypodium CAD genes

The Brachypodium genome (www.brachypodium.org) 8x release (version 1.2) was screened in silico for putative CAD sequences using known CAD sequences from Arabidopsis, rice, sorghum, maize, wheat, and perennial ryegrass as queries. Gateway-specific primers with attB1 and attB2 sites were designed (Supplementary Table S1, available at JXB online) for amplification of the open reading frames of putative Brachypodium CAD, including flanking nucleotides from the 5'- and 3'-UTR respectively to maintain the correct frame. PCR was conducted on both gDNA and cDNA for amplification of the different CAD sequences using LeFaq (Takara, Japan) and buffer [GCI/II buffer (Takara) was used for CAD2 from gDNA and CAD7 from cDNA], according to the manufacturer’s protocol, and a three-step amplification program (Supplementary Table S1). All products were cloned into the pDONR201 vector, propagated in Escherichia coli Top10 (Invitrogen, USA) and inserts were confirmed by sequencing (MWG, Germany). Sequence analyses were performed using CLC Main Workbench version 6.6 (CLC bio, Aarhus, Denmark). Sequences were deposited at GenBank [accession numbers: JQ768796 (BdCAD3) and JQ768797 (BdCAD5)].

Expression of CAD genes in different Brachypodium tissues

Semi-quantitative reverse-transcription PCR was used to determine expression levels in root, stem, leaf, and spike tissues, respectively. Gene-specific primer pairs (Supplementary Table S1) flanking introns were designed to discern between amplification resulting from cDNA and residual gDNA. cDNA concentrations were adjusted to 100 ng μl⁻¹. Parallel samples were run using the primer pair RT-BdU14 as a control. PCR products were sequenced to confirm primer specificity.

Expression and purification of recombinant enzymes

The coding region of BdCAD5 and BdCAD3 was transferred from the donor vector pDONR201 into the expression vector pDEST17 by the LR recombination reaction (Gateway; Invitrogen), according to the manufacturer’s protocol. Cultures (400 ml) were grown at 37 °C until an A₆₀₀ of approximately 1 was obtained. The cells were cooled to 16 °C and IPTG was added to a final concentration of 0.5 mM to induce expression of the His-tagged CAD protein and grown for another 26 h at 16 °C. Cells were harvested at 4000 g and 4 °C for 15 min and then stored at −20 °C until use.

The frozen pellets were thawed on ice and resuspended in lysis buffer (50 mM NaH₂PO₄, 300 mM NaCl with 10 mM imidazole, pH 8.0). The cells were disrupted by sonication on ice and the extracts were cleared by centrifugation at 12,000 g and 4 °C for 30 min. To the cleared lysates, 1 ml Ni-NTA agarose (Qiagen) was added, according to the manufacturer’s protocol, and poured into a Ni-NTA superflow columns (Qiagen). After washing the column using lysis buffer (starting with 10 mM then followed by 20 mM imidazole), the His-tagged BdCAD5 were eluted with 250 mM imidazole in lysis buffer.

After desalting, the BdCAD5 fraction was applied to a MonoQ HR 5/5 anion exchange column (GE Healthcare, UK) equilibrated in buffer A at a flow rate of 1 ml min⁻¹. Proteins were eluted by a linear NaCl gradient from 0 to 1.0 M and collected in 1 ml fractions. Fractions showing CAD activity were pooled and concentrated.

The final purification step of BdCAD5 and BdCAD3 was size exclusion chromatography using a Sephacryl 200 HR 20/60 column (GE Healthcare) at 0.5 ml min⁻¹, equilibrated with 20 mM TRIS-HCl, 150 mM NaCl, pH 7.5. Collected fractions showing CAD activity were concentrated and stored in 20 mM TRIS-HCl, 150 mM NaCl, pH 7.5, 5% ethylene glycol, 5 mM DTT at −20 °C until analysis. Presence and purity of BdCAD5 and BdCAD3 was analysed by SDS-PAGE using PageRuler Prestained Protein ladder (Fermentas, CA, USA) and bands were visualized by Coomassie Brilliant Blue G-250 (Candiano et al., 2004). Protein was quantified at 280 nm using a NanoDrop 2000 (Thermo) and bovine serum albumin as standard.

Phylogenetic tree

The protein sequences for 86 putative CAD enzymes were obtained from GenBank, Ensembl Genomes project, UniProt, Phylotome, and WheatEST and they are described in Supplementary Table S2. Sequences were aligned with MUSCLE version 3.7 (Edgar, 2004) and ambiguous regions (i.e. containing gaps and/or poorly aligned) were removed with Gblocks version 0.91b (Talavera and Castresana, 2007) using the default settings and a minimum length of a block after gap cleaning of 10. A phylogenetic tree was reconstructed using the maximum-likelihood method implemented in the PhyML program version 3.0 aLRT (Anisimova and Gascuel, 2006) on the Phylogeny.fr platform (Dereeper et al., 2008) with the LG substitution model selected and using the default settings. Reliability for internal branch was assessed using the bootstrapping method with 100 bootstrap replicates. Graphical representation and edition of the phylogenetic tree were performed with TreeDyn version 198.3 (Chevenet et al., 2006).

Determination of the molecular weight by size exclusion chromatography

BdCAD5 was analysed for dimer conformations by size exclusion chromatography at 0.5 ml min⁻¹, equilibrated with 20 mM TRIS-HCl, 150 mM NaCl, pH 7.5 and detected at 220 and 280 nm. A combination of the Low and the High Molar Weight calibration kits (GE Healthcare), containing ferritin (440,000 Da), aldolase (158,000 Da), albumin (67,000 Da), ovalbumin (43,000 Da), and ribonuclease A (13,700 Da), was used as standard and Blue dextran (2000 kDa) for determining void volume.

pH dependence for coniferyl aldehyde

The pH-dependent change in molar absorption of coniferyl aldehyde was obtained by measuring the UV-Vis spectra at different pH on a Shimatzu UV-3600 spectrophotometer. Coniferyl aldehyde (~6 mg) was...
solubilized in 1 ml 2-methoxyethanol (Sigma-Aldrich, USA) and further diluted in a multi-pH buffer (50 mM sodium citrate, 50 mM Na₂HPO₄, and 50 mM sodium borate titrated with 4 M NaOH) to 0.03479 mM. The buffer was used in the range pH 3.5–11.5.

**Enzyme assays**

The enzymatic reduction of coniferyl aldehyde was determined by the decrease in absorbance at 340 nm due to the reduction of coniferyl or sinapyl aldehyde and NADPH (Wyrambik and Grisebach, 1975). The incubation mixture (final volume 1.6 ml) contained 0–1200 µM coniferyl alcohol, 0–1400 µM of conifer aldehyde (Wyrambik and Grisebach, 1975). The incubation measured by the increase in absorbance at 400 nm due to the formation similar in value.

In time. The molar absorptions of coniferyl and sinapyl aldehydes were 0–30 µM NADPH with 10 µl diluted protein in 200 mM potassium phosphate buffer, pH 6.25 (BdCAD5). The protein was diluted in dilution buffer (20 mM TRIS-HCl, % ethylene glycol, and 5 mM DTT, pH 7.5) prior to measurements, to have a satisfactory activity with linearity in time. The molar absorptions of coniferyl and sinapyl aldehydes were similar in value.

For the oxidation of coniferyl alcohol, the enzymatic reaction was measured by the increase in absorbance at 400 nm due to the formation of conifer aldehyde (Wyrambik and Grisebach, 1975). The incubation mixture (1.6 ml) contained 0–1200 µM coniferyl alcohol, 0–1400 µM NADP⁺, 10 µl diluted protein in 100 mM potassium phosphate buffer, pH 8.8 (BdCAD3) or 0–100 µM conifer alcohol, 0–70 µM NADP⁺, 10 µl diluted protein in 100 mM TRIS-HCl, pH 8.8 (BdCAD5). Protein was diluted in dilution buffer prior to measurements. The steady-state parameters were determined by fitting the rates to the Michaelis–Menten equation using the nlrwr package (Ritz and Streibig, 2008) in R (R Development Core Team, 2011).

**The pH optimum of catalytic activity**

The pH profile of the catalytic activity of BdCAD5 and BdCAD3 was determined at pH 4.5–10 using coniferyl aldehyde and NADPH or conifer alcohol and NADP⁺ as substrates and cofactors, respectively, in multi-pH buffer in the range pH 3.5–11.5.

**Temperature optimum of catalytic activity**

The temperature profile of the catalytic activity of BdCAD3 and BdCAD5 was determined using coniferyl aldehyde and NADPH as substrates in a multi-pH buffer pH 5.2 (BdCAD3) or 100 mM sodium phosphate buffer pH 6.40 (BdCAD5), according to the optimum pH. Buffer was heated to desired temperature before coniferyl aldehyde, NADPH, and finally enzyme were added and the activity was measured instantly.

**Tryptic peptide mapping**

The purified BdCAD3 and BdCAD5 proteins were analysed by Alphalyse A/S (Odense, Denmark) by MALDI-TOF tryptic peptide mass fingerprinting and MALDI-TOF/TOF peptide sequencing after treatment of the protein with trypsin.

**Results**

**In silico characterization of the B. distachyon CAD genes**

**In silico screening of the Brachypodium genome (www.brachypodium.org) identified seven putative CAD genes numbered according to Guo et al. (2010): BdCAD1: Bradi5g04130.1; BdCAD2: Bradi5g21550.1; BdCAD3: Bradi4g29770.1; BdCAD4: Bradi4g29780.1; BdCAD5: Bradi3g06480.1; BdCAD7: Bradi3g17920.1; and BdCAD8: Bradi3g22980.1. However, in a previous study on the evolution of the CAD/SAD gene family, Guo et al. (2010) annotated Bradi3g10580.1 as BdCAD6. Bradi3g10580.1 was not considered further in this study as the deduced protein sequence lacks characteristics important for CAD activity. For further details, see Supplementary Table S3. The CAD genes were distributed on three of the five chromosomes of Brachypodium: BdCAD1 and BdCAD2 on chromosome 5, BdCAD3 and BdCAD4 next to each other on chromosome 4, and BdCAD5, BdCAD7 and BdCAD8 on chromosome 3.

Structural analysis of the isolated Brachypodium CAD genes revealed different intron–exon patterns both in relation to position and number of introns, which ranged from one to four per gene (Fig. 1). Furthermore, substantial differences in the size between the exons were observed. Comparison of the positions of the splicing sites (intron locations) between the seven CAD sequences showed conserved domains without splicing sites. A 114-bp sequence (black boxes in Fig. 1) was found in all seven sequences and encoded the putative binding site I for the monolignol substrate according to the model proposed by Youn et al. (2006) (see also annotations on the alignment in Fig. 4). The 114-bp sequence was identical to the second exon flanked by introns in BdCAD2, -3, -4, -5, and -8. Relative to these, the 114-bp sequence was merged with exon1 and exon3 in BdCAD1 and merged with exon2 in BdCAD7.

The last intron in BdCAD2, -7, and -8 was located in the middle of the helix structured monolignol-binding site II. This splitting site was conserved relative to the amino-acid sequences of binding site II AFAL\|VGG, AFAL\|VAK, and ATLN\|LGA in BdCAD2, 7, and 8, respectively, approximately 200 nucleotides upstream from the stop codon. The first exon of BdCAD4 contained additional 177 nucleotides encoding a putative 59 amino-acid signal peptide. No signal peptides could be predicted in any of the other six CAD genes.

**Fig. 1.** Intron–exon structure of Brachypodium distachyon CAD genes. Boxes represent exons, black indicating the splicing-site-free box coding for binding site I; numbers are the size of the exons measured in nucleotides.
In vitro characterization of the BdCAD genes

Five genes, BdCAD1, -2, -3, -4, -5, and -7, were isolated from genomic DNA from the Brachypodium genotype Bd21-3. BdCAD3 and BdCAD5 were also isolated from cDNA. Sequencing of the isolated genes revealed minor sequence differences between the genotypes Bd21 (database) and Bd21-3 (this study). For BdCAD1 the sequence of the isolated cDNA aligned perfectly with the cDNA predicted from the database. For BdCAD5 three nucleotide differences between the database-predicted cDNA and the isolated cDNA were identified. None of these resulted in changes in the deduced amino-acid sequence.

BdCAD gene expression in tissues

The expression patterns of five BdCAD genes were analysed by semi-quantitative reverse-transcription PCR in four Brachypodium tissues at the seed-filling stage: root, stem, leaf, and spike (Fig. 2). BdCAD5 was the only gene to be expressed in all tissues and the relative expression varied across tissues, with the highest expression in root and stem. BdCAD2 was expressed in the root, stem, and spike, with the highest expression in the stem. BdCAD3 was expressed in stem and spike and BdCAD4 was solely expressed in the stem. BdCAD7 was expressed in all tissues except the root, showing highest expression in the leaf.

Characterization of the B. distachyon CAD proteins

The BdCAD5 protein sequence was found most similar to CAD proteins from Festuca arundinacea, wheat, and Lolium perenne, with identities close to 90%. Further phylogenetic analysis revealed that BdCAD5 clustered with AtCAD5, SlCAD2, and ZmCAD2 (group I, Fig. 3), all of which have been characterized as bona fide CAD genes involved in lignin biosynthesis. The identity between BdCAD5 and AtCAD5 was 73% whereas the identity to SlCAD2 and ZmCAD2 was 86%. BdCAD3 in group IIIb had the highest sequence identities to CAD proteins from maize, sorghum, rice, and L. perenne, in the range of 78–83%. BdCAD2, BdCAD3 or BdCAD4, and BdCAD7 did not group with the bona fide CADs, as indicated by identities to ZmCAD2 in the range of 46–51%. BdCAD2 in group IIIa had the highest sequence identities with sorghum, maize, rice, and L. perenne, in the range of 68–74%. BdCAD4 in group IIIa had the highest sequence identity to CAD proteins from rice and sorghum, in the range of 80–84%. For BdCAD7 in group IIIa, the highest identities were observed with L. perenne, maize, rice, and sorghum, in the range of 82–85%. When comparing 86 plant CAD sequences from database searches, five groups were found: groups I–V. BdCAD5 was positioned in the bona fide group I, whereas BdCAD2, -3, -4, and -7 were located in group IIIb, BdCAD1 in group IV and BdCAD8 in group V. Sequences from C3 plants (maize, rice, Festuca, Brachypodium, and barley) fell in the same subgroups splitting from sequences from C4 plants and sequences from monocot and dicot species were present in all five subgroups.

The protein sequences of selected bona fide CAD sequences (from group I) were aligned together with BdCAD5, BdCAD3, and SlCAD8-4 (Fig. 4) and the overall identity for the selected sequences was 37% (135 invariant residues out of a total of 368). The cofactor- and zinc-binding sequences conserved in alcohol dehydrogenases ADHs were present in all BdCADs. The Zn1-binding motif and structural Zn2 consensus regions were located at amino acids 48, 70–71, 164, and 101, 104, 107, and 115, respectively. The NADP-binding domain was identified as residues 193, 212–217, and 276 and the sequences contained also a GLGGV(L)G sequence in the loop between β-sheet 10 and α-helix 5 (Fig. 4); known as a Rossmann fold (Rossmann et al., 1974). Of the seven invariant Cys, six were related to the Zn-binding site. The residues related to binding of cofactor NADP(H) and aldehyde substrate showed high degree of similarity throughout the alignment.

The protein sequences of the seven putative BdCADs found by in silico screening of the Brachypodium genome were aligned (Supplementary Fig. S1). Out of 379 amino-acid residues, 89 were invariant (23% identity). The amino-acid residues 58HL59 in the BdCAD5 sequence, believed to be part of the catalytic mechanism as a proton donor, were not found in the other BdCADs that contain either EW or DW dipeptide, except for BdCAD8 which contains QH indicating a change from E to Q as a result of a G→C base transversion. This change in one of the proton-donating residues according to the model of Youn et al. (2006) may lead to an inactive BdCAD8.

The protein sequence of BdCAD4 was predicted to have a chloroplast or mitochondrial signal peptide by TargetP, ChloroP, and SignalP algorithms (Emanuelsson et al., 2007) and also OsCAD8B, OsCAD8C, SlCAD4-4, SlCAD8-1 and HvCAD8B contain these putative signal peptides. Mannitol dehydrogenase from Zea mays (ZmMTD, accession ACG28500) showed 50% identity to BdCAD4 and it contains a similar chloroplast signal peptide and all residues known to be critical for CAD activity. The reliability class for the prediction in TargetP was the lowest of all tested sequences with a value of 2. Besides an obvious need for ADH activity in these organelles, there is no hint in...
the literature to any detectable CAD activity in chloroplast or mitochondria.

**Optimization of the enzyme assay**

Since the absorbance of coniferyl aldehyde was found to be strongly pH dependent, due to the phenol–phenolate equilibrium, the UV-Vis spectra in the region 200–550 nm of the diluted solutions of coniferyl aldehyde were recorded (Supplementary Fig. S2A). The changes in absorbance correlated the increase in pH resulted in a decrease at 221, 238, and 338 nm and an increase in absorbance at 268 and 401. A total of four isosbestic points at 213, 251, 278, and 360 nm were observed in the spectra. The molar absorbance coefficients at 340 nm and 400 nm were plotted
as function of the pH value, and the pKₐ of coniferyl aldehyde was determined to be 8.09 ± 0.04 (Table 1) by fitting the function \( \varepsilon_{\text{total}} = (\varepsilon_{\text{acid}} + \varepsilon_{\text{base}} \times 10^{-(pK_a - \text{pH})})/(1 + 10^{-(pK_a - \text{pH})}) \) to the spectra of coniferyl aldehyde at different pH values (Supplementary Fig. S2B). The used function was derived by relating the law of Lambert-Beers to the mass balance and the acid constant relations.

**Characterization of the recombinant BdCAD3 and BdCAD5 enzymes**

Recombinant BdCAD3 and BdCAD5 were expressed in *E. coli* as 6×His-tagged proteins and purified to homogeneity in three steps: Ni-affinity capture, followed by anion exchange chromatography, and finally by size exclusion chromatography. The relative molar masses were determined by SDS-PAGE to be 48.7 and 49.0 kDa for BdCAD3 and BdCAD5, respectively (Supplementary Fig. S3). The theoretical molar masses of BdCAD3 and BdCAD5 were calculated to be 37.7 and 38.5 kDa, respectively (Supplementary Table S3).

Using size exclusion chromatography, the molar mass of BdCAD5 was determined to be 40.3 kDa. The elution profile (data not shown) was almost symmetric with a shoulder at 97.4 kDa. The total area of the peak could be deconvoluted into two peaks covering 76% of the area at 40.3 kDa and 24% at 97.4 kDa. This observation indicated that active soluble BdCAD5 existed as equilibrium between a monomer and homo-dimer form.

Fig. 4. Alignment of the protein sequences of BdCAD3 and BdCAD5 and other putative lignin CADs. The alignment also includes sequences from rice (OsCAD2), switchgrass (PvCAD1), maize (ZmCAD), sorghum (SbCAD2 and SbCAD8-4), eucalyptus (EgCAD2), Arabidopsis (AtCAD5). Populus tremuloides (PotSAD) and tobacco (NtCAD19). Secondary structure elements extracted from the protein structure of AtCAD5 (accession 2CF5) are indicated at the top and those from PotSAD (accession 1YQD) at the bottom. Arrows indicate β-sheet structure, helices indicate helical structure motifs, and TT indicate turns. The amino acids believed to be part of the active sites as discussed in Youn et al. (2006) are shown below the alignment with following codes: ⋆ = NADP⁺ binding; ○ = monolignol-binding site I; ● = monolignol-binding site II; ▲ = Zn-binding catalytic; ▲ = Zn-binding structural.
Table 1. The parameters of the function ε_{total} = ε_{acid} + ε_{base} \times 10^{(pK_a-pH)/(1 + 10^{(pK_a-pH)})} fitted to molar absorption at 340 and 400 nm as a function of pH

| Wavelength (nm) | ε_{acid} (M^{-1} cm^{-1}) | ε_{base} (M^{-1} cm^{-1}) | pK_a |
|----------------|---------------------------|---------------------------|------|
| 340            | 23165.7 ± 207.8           | 5626.7 ± 267.1            | 8.09 ± 0.029 |
| 400            | 495.8 ± 108.4             | 3693.4 ± 138.2            | 8.07 ± 0.010 |

Data are presented in Supplementary Fig. S2. The function was derived from the expression of the acid constant, the mass balance, and the law of Lambert-Beers and describes the pH dependence of coniferyl aldehyde with respect to the pK&sub;a of the phenol group. ε_{acid} and ε_{base} are the molar absorption coefficients of the protonated (phenol-form) and the unprotonated (phenolate form) of coniferyl aldehyde, respectively. The molar absorption coefficient to any pH needed can be calculated using the formula and the constants.

The purified BdCAD3 and BdCAD5 proteins were analysed by MALDI-TOF peptide mass fingerprinting and MALDI-TOF/TOF peptide sequencing after treatment with trypsin. The peptides were blasted against both the NCBI database of general non-redundant sequences and against a database of protein sequences covering 31,029 protein sequences at www.brachypodium.org. The significant best-matching protein based on a probability-scoring algorithm with BdCAD3, with sequence coverage 37%, was Bradi4g29770.1 and with BdCAD5, with 58% sequence coverage was Bradi3g06480.1, both found in the Brachypodium.org database.

The optimal pH for BdCAD3 and BdCAD5 were investigated for both reducing and oxidizing catalytic activity (Fig. 5A). The optimal specific activity for the reduction process for BdCAD3 and BdCAD5 were determined to be pH 5.2 and 6.1, respectively. The oxidation of coniferyl alcohol showed maximal activity at pH 8.0 and 9.7, respectively. For both the reduction and oxidation process, the specific activity of BdCAD5 was significantly higher than for BdCAD3.

The reduction activity was maximal at 50–55 °C for both BdCAD3 and BdCAD5 (Fig. 5B) and as for the pH dependency, BdCAD5 showed significantly higher specific activity compared with BdCAD3.

The stability of purified BdCAD3 and BdCAD5 was monitored by the specific activity using coniferyl alcohol as substrate. There was no drop in activity over a period of 12 days after dilution at 4 °C (data not shown).

The steady-state parameters for the purified BdCAD3 and BdCAD5 were determined using coniferyl aldehyde and coniferyl alcohol as substrates (Table 2). BdCAD3 and BdCAD5 differed substantially in their overall catalytic properties. The efficiency of the reduction of coniferyl aldehyde by BdCAD5, expressed as the values k_{cat}/K_m, differs by a factor of 50 compared with BdCAD3. For coniferyl aldehyde, the K_m value for BdCAD5 was 10-times lower than for BdCAD3, which could indicate a difference in choice of primary substrate. The same was observed when comparing the ability to oxidize coniferyl alcohol. BdCAD5 showed a performance 100-times higher than that determined for BdCAD3. However, the maximal velocity of the conversion of coniferyl alcohol by BdCAD3 was found to be 4.6-times higher than for BdCAD5, but the overall efficiency was 100-times lower than for BdCAD5 because of a very high K_m value. For comparison, the conversion of sinapyl aldehyde was tested as well. The enzyme batch had been in the freezer for 6 months, resulting in a slight reduction in activity. All measurements were performed on enzyme batches stored under the same conditions and time for reliable comparison. For BdCAD3, the values were K_m 17.1 vs. 6.1 µM, V_max 92.6 vs. 25.4 nkat (mg protein)^{-1}, k_{cat} 5.2 vs. 1.4 s^{-1}, and k_{cat}/K_m 0.30 vs. 0.23 s^{-1} µM^{-1}. For BdCAD5, the values were K_m 4.9 vs. 6.5 µM, V_max 513 vs. 589
Characterization of BdCADs

| Enzyme  | Substrate          | $K_m$ (µM) | $V_{max}$ (nkat mg$^{-1}$) | $k_{cat}$ (s$^{-1}$) | $k_{cat}/K_m$ (s$^{-1}$ µM$^{-1}$) |
|---------|--------------------|------------|----------------------------|---------------------|-----------------------------------|
| BdCAD3  | Coniferyl aldehyde | 33.3 ± 3.80| 154.3 ± 6.86               | 8.66                | 0.26                              |
|         | NADPH              | 1.30 ± 0.12| 114.9 ± 2.04               | 6.45                | 0.55                              |
|         | Coniferyl alcohol* | 509.1 ± 134.0| 611.5 ± 72.1              | 34.33               | 0.067                             |
|         | NADP*              | 66.0 ± 11.79| 93.9 ± 3.31               | 5.24                | 0.079                             |
| BdCAD5  | Coniferyl aldehyde | 3.1 ± 0.44 | 492.3 ± 19.16              | 38.60               | 12.87                             |
|         | NADPH              | 7.2 ± 1.7  | 487.1 ± 32.1               | 38.19               | 5.30                              |
|         | Coniferyl alcohol* | 1.44 ± 0.24| 131.8 ± 4.55              | 10.35               | 7.19                              |
|         | NADP*              | 0.72 ± 0.10| 125.9 ± 3.55              | 9.87                | 13.71                             |

The parameters were calculated by fitting the Michaelis–Menten equation on initial rate experimental data using non-linear fitting ($n = 3$; except *, $n = 1$) using OriginPro (Originlab). Reactions were carried out at 23 °C and the protein concentrations were 0.42 (BdCAD3) and 0.58 µg/ml (BdCAD5).

| Substrate          | $K_m$ (µM) | $V_{max}$ (nkat mg$^{-1}$) | $k_{cat}$ (s$^{-1}$) | $k_{cat}/K_m$ (s$^{-1}$ µM$^{-1}$) |
|--------------------|------------|----------------------------|---------------------|-----------------------------------|
| Coniferyl aldehyde | 33.3 ± 3.80| 154.3 ± 6.86               | 8.66                | 0.26                              |
| NADPH              | 1.30 ± 0.12| 114.9 ± 2.04               | 6.45                | 0.55                              |
| Coniferyl alcohol* | 509.1 ± 134.0| 611.5 ± 72.1              | 34.33               | 0.067                             |
| NADP*              | 66.0 ± 11.79| 93.9 ± 3.31               | 5.24                | 0.079                             |
| Coniferyl aldehyde | 3.1 ± 0.44 | 492.3 ± 19.16              | 38.60               | 12.87                             |
| NADPH              | 7.2 ± 1.7  | 487.1 ± 32.1               | 38.19               | 5.30                              |
| Coniferyl alcohol* | 1.44 ± 0.24| 131.8 ± 4.55              | 10.35               | 7.19                              |
| NADP*              | 0.72 ± 0.10| 125.9 ± 3.55              | 9.87                | 13.71                             |

Predicted protein structures of BdCAD3 and BdCAD5

The high similarity between the predicted structures of BdCAD3 and BdCAD5 is presented in Fig. 6. The structures of the two proteins were in general very similar to the overall fold of AtCAD5, PotSAD, and other ADHs with known structures.

Fig. 6. In silico models of the homo-dimers of BdCAD3 (left) and BdCAD5 (right). Structures were modelled using Modeller (Sali and Blundell, 1993). The template for BdCAD3 was PotSAD (1YQD) and for BdCAD5 was AtCAD5 (2CF5). Zinc ions are shown as grey spheres and cofactor NADP$^+$ as green sticks.
Table 3. Comparison of the proposed key amino-acid residues found in the predicted active sites of AtCAD5, PotSAD, BdCAD3, and BdCAD5

| AtCAD position | AtCAD5 | BdCAD5 | PotSAD | BdCAD3 | Notes |
|----------------|--------|--------|--------|--------|-------|
| 49             | T      | T      | S      | T      | NADP-binding/ H-shuttle |
| 52             | H      | H      | H      | H      | NADP-binding |
| 53             | Q      | Q      | S      | I      | Substrate-binding |
| 57             | D      | H      | D      | E      | H-shuttle |
| 58             | L      | L      | W      | W      | Substrate-binding |
| 60             | M      | A      | F      | N      | Substrate-binding |
| 70             | E      | E      | E      | E      | Zn-binding |
| 95             | C      | V      | C      | Y      | Substrate-binding |
| 119            | W      | W      | L      | L      | Substrate-binding |
| 192            | V      | V      | L      | L      | NADP-binding |
| 211            | S      | S      | S      | S      | NADP-binding |
| 212            | S      | S      | T      | S      | NADP-binding |
| 213            | S      | S      | S      | S      | NADP-binding |
| 216            | K      | K      | K      | K      | NADP-binding |
| 276            | V      | V      | A      | A      | Substrate-binding |
| 286            | P      | P      | F      | Y      | Substrate-binding |
| 289            | M      | M      | I      | I      | Substrate-binding |
| 290            | L      | L      | A      | T      | Substrate-binding |
| 299            | F      | F      | G      | G      | Substrate-binding |
| 300            | I      | I      | I      | V      | Substrate-binding |
| Ratio to AtCAD5| 20/20  | 17/20  | 7/20   | 7/20   |

Residues were selected based on the published structures of AtCAD5 and PotSAD.

contained six β-sheets and five α-helices. The amino acids in the catalytic and the structural Zn sites, respectively, were all fully conserved in BdCAD3 and BdCAD5 as seven Cys and one Glu. According to the alignment (Fig. 4), the residues forming the binding pocket in BdCAD3 were similar to the residues found in PotSAD (Table 3). The substrate-binding cavity of BdCAD3 was similar to PotSAD whereas BdCAD5 was similar to AtCAD5. The two Zn-binding sites both showed the general patterns seen in other CAD enzymes: GHE(X)2G(X)5V for the catalytic Zn-binding site and GD(X)10C(X)2C(X)7C for the structural Zn-binding site. Furthermore the NADP(H)-binding site pattern G(X)2G(X)2GLGG(X)GH(X)2VK(X)2K(X)2G-(X)2VT(X)S(X)S(X)2K was found in all 86 sequences.

Discussion

The aim of the present study was to identify, isolate, and characterize genes coding for the CAD enzyme in B. distachyon, the model for temperate grasses. In silico screening of the Brachypodium database using known CAD genes as queries yielded seven CAD-like genes coding for putative CAD enzymes. The amino-acid identity between the BdCAD enzymes ranged from 41 to 67%, compared with approximately 90% for the bona fide CADs across plant species. BdCAD6 was excluded from further analysis because its protein sequence lacked vital amino acids responsible for the coordination of the Zn metal ions.

Identification of CAD genes in Brachypodium

Of the identified Brachypodium CADs, BdCAD5 showed high sequence similarity to group I (Fig. 3), which have been biochemically characterized as bona fide CADs (Goffner et al., 1992; Kim et al., 2004). BdCAD5 is thus the most likely candidate gene for lignin biosynthesis in Brachypodium. BdCAD5 has been shown to be expressed in all tissues tested (Fig. 2) but most highly in roots and stems. Similar expression patterns have been seen by other group I CAD, such as rice OsCAD2 (Tobias and Chow, 2005), and L. perenne LpCAD3 (Lynch et al., 2002) and these two genes also exhibited the highest expression in roots and stem. Wheat TaCAD1 is also high expressed in the stem, while has only minor expression in roots (Ma, 2010). The similarity of expression patterns between different plant species could indicate similar functions for these genes.

Localization of CAD enzymes

Six of the seven CAD genes were devoid of N-terminal signal peptides and they were likely cytosolic, which is consistent with the immunolocalization studies of these genes in sugarcane and maize (Ruelland et al., 2003). The signal peptide in BdCAD4 was found to be 59 residues but showed varying length in HvCAD8B, OsCAD8B and -C, ShCAD4-4 and ShCAD8-1, ZmMTD, and PoprCAD9. These organellar CAD-like enzymes were placed in the large group III in the phylogenetic tree (Fig. 3), which did not provide much information about their function besides a general need for ADH activity in the plastids. BdCAD4 was located next to BdCAD3 on chromosome 4, which could be a result of a duplication event, where the number of and localization of the introns were conserved (Fig. 1). However only BdCAD4, and not BdCAD3, had a signal peptide, and the identity between the two mature proteins was only 76%. Furthermore, BdCAD4 was only weakly expressed and in stem tissue only, whereas BdCAD3 was highly expressed in the stem and less in the spike. These differences in expression patterns were supported by low similarities between the promoter sequences (data not shown). The position of BdCAD3 and BdCAD4 on chromosome 4 is syntenic to rice chromosome 9 and sorghum chromosome 2 (Vogel et al., 2010). In both rice and sorghum, two homologous genes were found, of which one has the signal peptide. It is likely that duplication was followed by diversification of the two genes including signal-peptide and promoter sequences and that they were inherited as a unit in order to preserve functionality.

Characterization of expressed BdCAD enzymes

After purification, the molecular mass of BdCAD5 was estimated to be 40 and 97 kDa by size exclusion chromatography. The functional unit of AtCAD5 has been shown to be a homo-dimer tightly associated through two two-fold related β-strands (Youn et al., 2006). This taken into account indicated that BdCAD3 and BdCAD5 are likely to maintain the dimer structure on the basis of the intact β-sheets β13–β17 (βD–βF in AtCAD5). Previously
characterized CADs have been found to be catalytically active in either a dimeric form as AtCAD5 (Youn et al., 2006) and EgCAD2 (Goffner et al., 1992) or in a monomeric form as isolated from Phaseolus vulgaris (Grima-Pettenati et al., 1994; Hawkins and Boudet, 1994) and from the bacterium Helicobacter pylori (Mee et al., 2005). E. gunnii CAD1 (EgCAD1) has been isolated in an active monomeric form (Goffner et al., 1992). However, as discussed in the introduction, EgCAD1 has been shown to be significantly different from other CAD-like enzymes such as AtCAD5 and BdCAD5. The current data indicate that a monomer–homo-dimer equilibrium exists for BdCAD5.

The molar absorbance coefficient, ε, of coniferyl aldehyde, was determined as a function of pH (Table 1 and Supplementary Fig. S2) and was higher compared with the values reported by Wyrambik and Grisebach (1975) as 15.800 M–1 cm–1 at 340 nm, pH 6.5 and 21,000 M–1 cm–1 at 400 nm, pH 8.8 and the values reported by Larroy et al. (2002). The reason for this difference could be the purity of the coniferyl aldehyde used. The estimation of the pK_a for the coniferyl aldehyde is found to be in good agreement with the value of 7.94 reported by Ragnar et al. (2000).

The conversion of coniferyl aldehyde to coniferyl alcohol by the expressed and purified BdCAD3 and BdCAD5 was shown to be strongly pH dependent (Fig. 5) with optima at pH 5.2 and 6.2, respectively. The observed difference in optimal pH between BdCAD3 and BdCAD5 could partly be explained by the amino-acid differences at position 57–58 (Fig. 4, AtCAD5 numbering). BdCAD3 had an EW motif whereas BdCAD5 had an HL motif. In AtCAD5 the residue was found to be an Asp which could be one of the key residues in the catalytic mechanism (Youn et al., 2006; Saathoff et al., 2010). CADs associated with cell-wall lignification and displaying significant catalytic activity towards monolignols have the HL or DL dipeptide motif (Saathoff et al., 2010). However, as wheat CAD1 contains a DL motif and displays a pH optimum of 7 (Ma, 2010) the two residues are not likely to be the sole key residues determining the regulation of the catalytic pH optimum.

Even though BdCAD3 and BdCAD5 contained a large number of residue changes compared with AtCAD5 and PotSAD (Table 3), they both remained similar in the overall structure (Fig. 6). According to the modelled structures of BdCAD3 and BdCAD5, the overall fold is very similar to other CADs, both bona fide and other types. Of the 20 amino acids located in the active site of AtCAD5, 17 were conserved in BdCAD5 and only seven in BdCAD3, none of which related to the binding of the monolignol moiety (Table 3). The efficiency (k_cat/K_m) of the reduction of coniferyl aldehyde by BdCAD3 and BdCAD5 displayed a difference by a factor of 50 in favour of BdCAD5. The k_cat/K_m of BdCAD5 is comparable to TaCAD1 (Ma, 2010), AtCAD5 (Kim et al., 2004), and PviCAD1 and -2 (Saathoff et al., 2010), which have all been classified as bona fide CADs. The efficiency (k_cat/K_m) and K_m of BdCAD5 in the conversion of coniferyl aldehyde and sinapyl aldehyde was comparable and the same was observed for BdCAD3, but with significantly lower values compared with BdCAD5 (Table 2). The difference in the K_m values for BdCAD3 and BdCAD5 with coniferyl aldehyde could indicate a difference in substrate preferences even though BdCAD3 seemed to have a low K_m value for sinapyl aldehyde but still a slow conversion (low k_cat) and low effectiveness compared with BdCAD5 (much lower k_cat/K_m). The value of K_m for BdCAD5 was lower than observed for AtCAD5/AtCAD4 and TaCAD1 and comparable to PviCAD2, indicating an even-more dedicated enzyme in the conversion of monolignols.

When comparing the substrate-binding pocket of BdCAD3 to that of AtCAD5, substitutions in the involved amino acids were generally going from larger to smaller residues. Such substitutions would probably result in a change in size, conformation, and hydrophobicity of the binding pocket. The roles of BdCAD3 and this group of CAD enzymes have not yet been elucidated. The substrate-binding site in BdCAD3 was equivalent to the substrate cavity of PotSAD when compared with AtCAD5 (Table 3). However, the catalytic activity of BdCAD3 against sinapyl aldehyde was lower than against coniferyl aldehyde, even though a slight lower K_m was observed. The findings that it is not possible to determine the preferred substrate solely from the composition of the binding pocket (AtCAD5 sequence vs. PotSAD sequence) are in agreement with Anterola and Lewis (2002) who reinterpreted the data of Botami and Noel (2005). The CAD and CAD-like enzymes have only a subsidiary role and do not entirely determine the final composition of H-, G-, and S-lignin.

PotSAD has been associated with lignin biosynthesis (Li et al., 2001) and AtCAD7 (ELI3-1) and AtCAD8 (ELI3-2) have been associated with plant defence compounds (Trezzini et al., 1993; Somssich et al., 1996). AtCAD7 has been characterized as a benzyl alcohol dehydrogenase with low catalytic activity against monolignol compounds (Somssich et al., 1996). Other CADs than bona fide CADs have additional roles such as plant defence or catalytic activity under specific conditions (Barakat et al., 2009; Saathoff et al., 2010). SbCAD4 has shown poor activity against monolignol substrates (Sattler et al., 2009), similarly to BdCAD3, indicating that these types of CADs have other functions in planta. The substrate-binding site is not the only determinant for understanding the different catalytic potentials and the roles in planta of the different CAD and CAD-like enzymes in Brachypodium.

In conclusion, biochemical and phylogenetic studies have identified targets CAD genes for gene modification. A TILLING strategy will be employed for identifying Brachypodium lines with mutations in these genes, in order to determine their in planta function. In addition, the bioinformatics lead to the suggestion of a chloroplastic location of a single CAD-like enzyme across plant species, with an as-yet unclear function, to be elucidated in the near future with biochemical analysis.

**Supplementary material**

Supplementary data are available at JXB online.

- Supplementary Fig. S1. Alignment of the seven Brachypodium protein sequences.
- Supplementary Fig. S2. pH-dependent UV-visible spectra of coniferyl alcohol.
- Supplementary Fig. S3. SDS-PAGE analysis of the purified BdCAD3 and BdCAD5.
- Supplementary Table S1. Primers used in the isolation of CAD genes and semi-quantitative reverse-transcription PCR.
Supplementary Table S2. List of plant proteins used in CAD protein phylogenetic analyses.
Supplementary Table S3. Putative CAD genes in Brachypodium distachyon.

Acknowledgements

Parts the work was supported by Fuel-for-Life and Bio4Bio from Danish Strategic Research Council. The authors thank Dr. Scott Sattler and Dr. Aaron Saathoff for providing the plasmid harbouring the sorghum gene CAD4 (SbCAD4).

References

Anisimova M, Gascuel O. 2006. Approximate likelihood-ratio test for branches: a fast, accurate, and powerful method. Systematic Biology 55, 539–552.

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

Barakat A, Bagniewska-Zadworna A, Choi A, Piakkat U, DiLoreto DS, Yellanki P, Carlson JE. 2009. The cinnamyl alcohol dehydrogenase gene family in Populus: phylogeny, organization, and expression. BMC Plant Biology 9, 26.

Barakat A, Bagniewska-Zadworna A, Frost CJ, Carlson JE. 2010. Phylogeny and expression profiling of CAD and CAD-like genes in hybrid Populus (P. deltoides × P. nigra): evidence from herbivore damage for subfunctionalization and functional divergence. BMC Plant Biology 10, 100.

Barriere Y, Ralph J, Mechin V, Guillamme S, Grabber JH, Argillier O, Chabbert B, Lapierre C. 2004. Genetic and molecular basis of grass cell wall biosynthesis and degradability. II. Lessons from brown-midrib mutants. Comptes Rendus Biologies 327, 847–860.

Bedon F, Levasseur C, Grima-Pettenati J, Seguin A, MacKay J. 2009. Sequence analysis and functional characterization of the promoter of the Picea glauca Cinnamyl Alcohol Dehydrogenase gene in transgenic white spruce plants. Plant Cell Reports 28, 787–800.

Bendtsen JD, Nielsen H, von Heijne G, Brunak S. 2004. Improved prediction of signal peptides: SignalP 3.0. Journal of Molecular Biology 340, 783–795.

Boerjan W, Ralph J, Baucher M. 2003. Lignin biosynthesis. Annual Review of Plant Biology 54, 519–546.

Bomati EK, Noel JP. 2005. Structural and kinetic basis for substrate selectivity in Populus tremuloides sinapyl alcohol dehydrogenase. The Plant Cell 17, 1598–1611.

Bonawitz ND, Chapple C. 2010. The genetics of lignin biosynthesis: connecting genotype to phenotype. Annual Review of Genetics 44, 337–363.

Bowie JU, Lüthy R, Eisenberg D. 1991. A method to identify protein sequences that fold into a known three-dimensional structure. Science 253, 164–170.

Campbell MM, Sederoff RR. 1996. Variation in lignin content and composition (mechanisms of control and implications for the genetic improvement of plants). Plant Physiology 110, 3–13.

Campbell MM, Sederoff RR. 1996. Variation in lignin content and composition (mechanisms of control and implications for the genetic improvement of plants). Plant Physiology 110, 3–13.

Chabannes M, Barakate A, Lapiere C, Marita JM, Ralph J, Pean M, Danoun S, Halpin C, Grima-Pettenati J, Boudet AM. 2001. Strong decrease in lignin content without significant alteration of plant development is induced by simultaneous down-regulation of cinnamoyl CoA reductase (CCR) and cinnamyl alcohol dehydrogenase (CAD) in tobacco plants. The Plant Journal 28, 257–270.

Chen F, Dixon RA. 2007. Lignin modification improves fermentable sugar yields for biofuel production. Nature Biotechnology 25, 759–761.

Chevenet F, Brun C, Banuls AL, Jacq B, Christen R. 2006. TreeDyn: towards dynamic graphics and annotations for analyses of trees. BMC Bioinformatics 7, 439.

Damiani I, Morreel K, Danoun S, et al. 2005. Metabolite profiling reveals a role for atypical cinnamyl alcohol dehydrogenase CAD1 in the synthesis of coniferyl alcohol in tobacco xylem. Plant Molecular Biology 59, 753–769.

Dereeper A, Guignon V, Blanc G, et al. 2008. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Research 36, W465-W469.

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

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32, 1792–1797.

Emanuelsson O, Brunak S, von Heijne G, Nielsen H. 2007. Locating proteins in the cell using TargetP, SignalP and related tools. Nature Protocols 2, 953–971.

Fu CX, Xiao XR, Xi YJ, Ge YX, Chen F, Bouton J, Dixon RA, Wang ZY. 2011. Downregulation of cinnamyl alcohol dehydrogenase (CAD) leads to improved saccharification efficiency in switchgrass. BioEnergy Research 4, 153–164.

Goffner D, Joffroy I, Grimalsenati J, Halpin C, Knight ME, Schuch W, Boudet AM. 1992. Purification and characterization of isoforms of cinnamyl alcohol dehydrogenase from Eucalyptus xylem. Planta 188, 48–53.

Goffner D, Van Doorslaere J, Yahiaoui N, Samaj J, Grima-Pettenati J, Boudet AM. 1998. A novel aromatic alcohol dehydrogenase in higher plants: molecular cloning and expression. Plant Molecular Biology 36, 755–765.

Gouet P, Courcelle E, Stuart DI, Metoz F. 1999. ESPript: analysis of multiple sequence alignments in PostScript. Bioinformatics 15, 305–308.

Grima-Pettenati J, Campargue C, Boudet A, Boudet AM. 1994. Purification and characterization of cinnamyl alcohol-dehydrogenase isoforms from Phaseolus vulgaris. Phytochemistry 37, 941–947.

Guillaumie S, Pichon M, Martinant JP, Bosio M, Goffner D, Barriere Y. 2007. Differential expression of phenylpropanoid and related genes in brown-midrib bm1, bm2, bm3, and bm4 young near-isogenic maize plants. Planta 226, 235–250.
Guo DM, Ran JH, Wang XQ. 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. Journal of Molecular Evolution 71, 202–218.

Halpin C, Holt K, Chojdecki J, Oliver D, Chabbert B, Monties B, Edwards K, Barakate A, Foxon GA. 1998. Brown-midrib maize (bm 1) – a mutation affecting the cinnamyl alcohol dehydrogenase gene. Plant Journal 14, 545–553.

Halpin C, Knight ME, Foxon GA, Campbell MM, Boudet AM, Boon JJ, Chabbert B, Tollier MT, Schuch W. 1994. Manipulation of lignin quality by downregulation of cinnamyl alcohol dehydrogenase. The Plant Journal 6, 339–689.

Hawkins SW, Boudet AM. 1994. Purification and characterization of cinnamyl alcohol-dehydrogenase isozymes from the periderm of Eucalyptus gunnii hook. Plant Physiology 104, 75–84.

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

Jorgenson LR. 1931. Brown midrib in maize and its linkage relations. Journal of the American Society of Agronomy 23, 549–1106.

Kim H, Ralph J, Lu FC, Pilate G, Leple JC, Pollet B, Lapierre C. 2002. Identification of the structure and origin of thioacidolysis marker compounds for cinnamyl alcohol dehydrogenase deficiency in angiosperms. Journal of Biological Chemistry 277, 47412–47419.

Kim SJ, Kim KW, Cho MH, Franceschi VR, Davin LB, Lewis NG. 2007. Expression of cinnamyl alcohol dehydrogenases and their putative homologues during Arabidopsis thaliana growth and development: lessons for database annotations? Phytochemistry 68, 1957–1974.

Kim SJ, Kim MR, Bedgar DL, Moinuddin SGA, Cardenas CL, Davin LB, Kang CL, Lewis NG. 2004. Functional reclassification of the putative cinnamyl alcohol dehydrogenase multigene family in Arabidopsis. Proceedings of the National Academy of Sciences, USA 101, 1455–1460.

Kim Y-H, Bae J, Huh G-H. 2010. Transcriptional regulation of the cinnamyl alcohol dehydrogenase gene from sweetpotato in response to plant developmental stage and environmental stress. Plant Cell Reports 29, 779–791.

Lapierre C, Pilate G, Pollet B, Mila I, Leple JC, Jouanin L, Kim H, Ralph J. 2004. Signatures of cinnamyl alcohol dehydrogenase deficiency in poplar lignins. Phytochemistry 65, 313–321.

Lapierre C, Pollet B, MacKay JJ, Sederoff RR. 2000. Lignin structure in a mutant pine deficient in cinnamyl alcohol dehydrogenase. Journal of Agricultural and Food Chemistry 48, 2326–2331.

Larkin MA, Blackshields G, Brown NP, et al. 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947–2948.

Larroy C, Fernandez MR, Gonzalez E, Pares X, Biosca JA. 2002. Characterization of the Saccharomyces cerevisiae YMR318C (ADH6) gene product as a broad specificity NADP+ dependent alcohol dehydrogenase: relevance in aldehyde reduction. The Biochemical Journal 361, 163–172.

Li LG, Cheng XF, Leshkevich J, Umezawa T, Harding SA, Chiang VL. 2001. The last step of syringyl monolignol biosynthesis in angiosperms is regulated by a novel gene encoding sinapyl alcohol dehydrogenase. The Plant Cell 13, 1567–1585.

Li X, Weng J-K, Chapelle C. 2008. Improvement of biomass through lignin modification. The Plant Journal 54, 569–581.

Li XJ, Yang Y, Yao JL, Chen GX, Li XH, Zhang QF, Wu CY. 2009. FLEXIBLE CULM 1 encoding a cinnamyl-alcohol dehydrogenase controls culm mechanical strength in rice. Plant Molecular Biology 69, 685–697.

Lynch D, Lidgett A, Mclnnes R, Huxley H, Jones E, Mahoney N, Spangenberg G. 2002. Isolation and characterization of three cinnamyl alcohol dehydrogenase homologues from perennial ryegrass (Lolium perenne L.). Journal of Plant Physiology 159, 653–660.

Lüthy R, Bowie JU, Eisenberg D. 1992. Assessment of protein models with three-dimensional profiles. Nature 356, 83–85.

Ma QH. 2010. Functional analysis of a cinnamyl alcohol dehydrogenase involved in lignin biosynthesis in wheat. Journal of Experimental Botany 61, 2735–2744.

MacKay JJ, Liu W, Whetten R, Sederoff RR, O’Malley DM. 1995. Genetic analysis of cinnamyl alcohol dehydrogenase in loblolly pine: single gene inheritance, molecular characterization and evolution. Molecular and General Genetics 247, 537–545.

Marita JM, Vermerris W, Ralph J, Hatfield RD. 2003. Variations in the cell wall composition of maize brown midrib mutants. Journal of Agricultural and Food Chemistry 51, 1313–1321.

Mee B, Kelleher D, Frias J, Malone R, Tipton KE, Henenan GTM, Windle HJ. 2005. Characterization of cinnamyl alcohol dehydrogenase of Helicobacter pylori – an aldehyde dismutating enzyme. FEBS Journal 272, 1255–1264.

Nüss D, Goettig P, Magler I, Denk U, Breitenbach M, Schneider PB, Brandstetter H, Simon-Nobbe B. 2010. Crystal structure of the NADP-dependent mannitol dehydrogenase from Cladosporium herbarum: implications for oligomerisation and catalysis. Biochimie 92, 985–993.

R Development Core Team. 2011. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Ragnar M, Lindgren CT, Nilvebrant N-O. 2000. pKa-values of guaiacyl and syringyl phenols related to lignin. Journal of Wood Chemistry and Technology 20, 277–305.

Ritz C, Streibig JC. 2008. Nonlinear regression With R. New York: Springer.

Rossmann MG, Moras D, Olsen KW. 1974. Chemical and biological evolution of a nucleotide-binding protein. Nature 250, 194–199.

Ruelland E, Campalans A, Selman-Housein G, Puigdomenech P, Rigau J. 2003. Cellular and subcellular localization of the lignin biosynthetic enzymes caffeic acid-O-methyltransferase, cinnamyl alcohol dehydrogenase and cinnamyl-coenzyme A reductase in two monocots, sugarcane and maize. Physiologia Plantarum 117, 93–99.

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, 783–795.
Sali A, Blundell TL. 1993. Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology* **234**, 779–815.

Sattler SE, Funnell-Harris DL, Pedersen JF. 2010. Brown midrib mutations and their importance to the utilization of maize, sorghum, and pearl millet lignocellulosic tissues. *Plant Science* **178**, 229–238.

Sattler SE, Saathoff AJ, Haas EJ, Palmer NA, Funnell-Harris DL, Sarath G, Pedersen JF. 2009. A nonsense mutation in a cinnamyl alcohol dehydrogenase gene is responsible for the sorghum brown midrib phenotype. *Plant Physiology* **150**, 584–595.

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

Saathoff AJ, Sattler S, Haas E, Twigg P, Sarath G. 2010. Switchgrass contains two cinnamyl alcohol dehydrogenases involved in lignin formation. *BioEnergy Research* **4**, 1–14.

Saathoff AJ, Sarath G, Chow EK, Dien BS, Tobias CM. 2011. Downregulation of cinnamyl-alcohol dehydrogenase in switchgrass by RNA silencing results in enhanced glucose release after cellulase treatment. *PLoS One* **6**, e16416.

Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology* **56**, 564–577.

Thompson JD, Higgins DG, Gibson TJ. 1994. Clustal W – improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**, 4673–4680.

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**, 678–688.

Vanholme R, Morreel K, Ralph J, Boerjan W. 2008. Lignin engineering. *Current Opinion in Plant Biology* **11**, 278–285.

Vogel JP, Garvin DF, Mockler TC, et al.; International Brachypodium Initiative. 2010. Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* **463**, 763–768.

Wyrambik D, Grisebach H. 1975. Purification and properties of isoenzymes of cinnamyl-alcohol dehydrogenase from soybean-cell-suspension cultures. *European Journal of Biochemistry* **59**, 9–15.

Youn B, Camacho R, Moinuddin SGA, Lee C, Davin LB, Lewis NG, Kang CH. 2006. Crystal structures and catalytic mechanism of the Arabidopsis cinnamyl alcohol dehydrogenases AtCAD5 and AtCAD4. *Organic and Biomolecular Chemistry* **4**, 1687–1697.