Plant Mol Biol (2016) 92:1–23
DOI 10.1007/s11103-016-0492-5

Transcriptome-wide identification and characterization of CAD isoforms specific for podophyllotoxin biosynthesis from *Podophyllum hexandrum*

Dipto Bhattacharyya[1] · Saptarshi Hazra[2] · Anindyajit Banerjee[3] · Riddhi Datta[2] · Deepak Kumar[2] · Saikat Chakrabarti[3] · Sharmila Chattopadhyay[2]

Received: 19 October 2015 / Accepted: 14 May 2016 / Published online: 7 July 2016

Abstract Podophyllotoxin (ptox) is a therapeutically important lignan derived from *Podophyllum hexandrum* and is used as a precursor for the synthesis of anticancer drugs etoposide, teniposide and etopophose. In spite of its enormous economic significance, genomic information on this endangered medicinal herb is scarce. We have performed de novo transcriptome analysis of methyl jasmonate (MeJA)-treated *P. hexandrum* cell cultures exhibiting enhanced ptox accumulation. The results revealed the maximum up-regulation of several isoforms of cinnamyl alcohol dehydrogenase (CAD). CAD catalyzes the synthesis of coniferyl alcohol and sinapyl alcohol from coniferaldehyde (CAld) and sinapaldehyde respectively. Coniferyl alcohol can produce both lignin and lignan while sinapyl alcohol produces only lignin. To isolate the CAD isoforms favoring ptox, we deduced full length cDNA sequences of four CAD isoforms: *PhCAD1, PhCAD2, PhCAD3* and *PhCAD4* from the contigs of the transcriptome data. In vitro enzyme assays indicated a higher affinity for CAld over sinapaldehyde for each isoform. In silico molecular docking analyses also suggested that PhCAD3 has a higher binding preference with CAld over sinapaldehyde, followed by PhCAD4, PhCAD2, and PhCAD1, respectively. The transgenic cell cultures overexpressing these isoforms independently revealed that *PhCAD3* favored the maximum accumulation of ptox as compared to lignin followed by *PhCAD4* and *PhCAD2*, whereas, *PhCAD1* favored both equally. Together, our study reveals transcriptome-wide identification and characterization of ptox specific CAD isoforms from *P. hexandrum*. It provides a useful resource for future research not only on the ptox biosynthetic pathway but on overall *P. hexandrum*, an endangered medicinal herb with immense therapeutic importance.

Keywords Cinnamyl alcohol dehydrogenase · Molecular docking · Podophyllotoxin · *Podophyllum hexandrum* · Transcriptome · Transgenic

Introduction

Lignans represent an abundant class of ubiquitous natural products of vascular plants with their important roles in human health protection, pharmacological applications, as well as in plant defense (Lewis and Davin 1999). Lignan contains two phenylpropane units which are connected by C8 Carbon, situated at the side chain of each unit (Umezawa 2003). The biosynthesis of lignans involves enantioselective dimerization of monolignols like coniferyl alcohol. The conversion of coniferyl alcohol to ptox involves hydroxylation, methylations, and methylenedioxy bridge formation.

Dipto Bhattacharyya and Saptarshi Hazra contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s11103-016-0492-5) contains supplementary material, which is available to authorized users.

✉ Sharmila Chattopadhyay
sharmila@iicb.res.in; chattopadhyay62@gmail.com

[1] Present address: Division of Biotechnology, Chonbuk National University, 79 Gobong-ro, Iksan-si, Jeollabuk-do 570–752, Republic of Korea

[2] Plant Biology Lab., Organic and Medicinal Chemistry Division, CSIR-Indian Institute Chemical Biology, 4, Raja S. C. Mullick Road, Kolkata 700 032, India

[3] Structural Biology and Bioinformatics Division, CSIR-Indian Institute Chemical Biology, 4, Raja S. C. Mullick Road, Kolkata 700 032, India

Ptox, an aryltetralin lignan, along with other related lignans, is present in significant amount in the roots and rhizomes of the medicinal herb *Podophyllum hexandrum* (Royle), commonly referred as the Himalayan Mayapple. At present, ptox is being used as the starting compound for the production of semi-synthetic drugs etoposide (VP-16-213), teniposide (VM-26) and ethophos, which are used in the treatment of lung and testicular cancers (Stahelin and von Wartburg 1991), leukemia and rheumatoid arthritis (Lerndal and Svensson 2000).

Currently, ptox is being extracted from the wild rhizomes of the genus *Podophyllum*. As noted earlier, the chemical synthesis of ptox is possible, but not economically feasible (Smollny et al. 1998). To attain the ever-increasing demand, cell culture-based production of ptox has been investigated as well (Kadkade 1981; van Uden et al. 1989; Heyenga et al. 1990; Wichers et al. 1991; Seidel et al. 2002; Chattopadhyay et al. 2003). However, production of ptox using cell culture systems may not be sufficient towards its production at commercial scale (Fuss 2003). Other than these, elicitation of cell culture is another approach that may overcome the limitations of the in vitro culture system.

Jasmonic acid (JA) and its more active derivative methyl jasmonate (MeJA), collectively called jasmonates (JAs), has been considered as key signal compounds in the elicitation process which leads to the de novo transcription and translation and, ultimately, to the biosynthesis of secondary metabolites in plant cell cultures (Gundlach et al. 1992). Cell suspension cultures of *Linum* spp. has also been reported to have higher yield of ptox after using MeJA and salicylic acid (SA) as elicitors (van Furden et al. 2005; Yousefzadi et al. 2010). Earlier we have also shown that the protein profile is significantly altered in the MeJA treated cell suspension culture of *P. hexandrum* having an enhanced ptox accumulation (Bhattacharyya et al. 2012). MeJA is known to reprogram cellular metabolism and cell cycle progression with upregulation of both monolignol biosynthetic pathway gene expression as well as monolignol and oligolignol production at the metabolite level (Pauwels et al. 2008). Elicitors induced rapid stimulation of the monolignol pathway has also been noted in flax (*Linum usitatissimum*) cell suspensions (Hano et al. 2006). The final step of monolignol biosynthesis is catalyzed by CAD (CAD; EC 1.1.1.95), which uses various phenylpropenyl aldehyde derivatives as substrates to ensure the diversity of lignin. CAD belongs to a family of NADPH-dependent enzymes first discovered by Gross et al. (1973). Based on the expression in lignified tissues, different CAD isoforms have been reported, as either lignin or lignan specific. Recent studies have suggested that CADs are encoded by a multigene family, and its homologs have been detected in a wide variety of bacteria and eukaryotes (Guo et al. 2010). Till date, several CAD/CAD-like genes have been reported from various plant species viz. *Arabidopsis* (Kim et al. 2004, 2007; Raes et al. 2003; Sibout et al. 2003, 2005), rice (Li et al. 2009; Tobias and Chow 2005), wheat (Ma 2010), *Brachypodium distachyon* (Bukh et al. 2012), tea (Deng et al. 2013), *Populus tomentosa* (Chao et al. 2014) etc.

Informations regarding the complete sequence of ptox biosynthetic pathway gene/s, especially the later steps, is limited. The complete CDS of dirigent protein oxidase [Gen Bank: DQ414685], pinoresinol-lariciresinol reductase [Gen Bank: EU855792.1] and [Gen Bank: EU240218], secoisolariciresinol dehydrogenase [Gen Bank: GU324975] and CAD from *P. hexandrum* [Gen Bank: HQ268590] has been reported in the National Centre for Biotechnology Information (NCBI). More recently, our group and others have focused on the next generation high throughput transcriptome sequencing of *P. hexandrum* and *P. peltatum* in order to fully characterize the ptox pathway (Bhattacharyya et al. 2013; Marques et al. 2013). In a recent study, six pathway enzymes have been identified, including an oxoglutarate-dependent dioxygenase that closes the core cyclohexane ring of the aryltetralin scaffold and the pathway has been reconstituted up to (–)-4′-desmethylepipodophyllotoxin, an etoposide aglycone (Lau and Sattely 2015).

The transcriptome is a comprehensive set of transcribed regions of the genome. It provides us important insights into the functional elements of the genome, their expression patterns and tissue-specific regulations under different conditions (Kyndt et al. 2012). When no genome sequence is available, transcriptome sequencing is an effective way to identify transcripts involved in specific biological processes. Moreover, generation of large-scale sequence data obtained through 454 GS-FLX titanium pyrosequencing not only enables gene discovery, molecular marker development, etc., but is also considered as one of the most effective tools when conservation of non-model organisms are concerned.

Here, we have used next generation sequencing technology to identify the major transcripts of ptox biosynthesis from MeJA treated *P. hexandrum* cell suspension culture exhibiting enhanced ptox content. CAD isoforms, derived from the contigs of transcriptome data, were characterized and noted to exhibit their specificity towards ptox. In addition, several candidate phenylpropanoid pathway genes identified here may expand our understanding of the molecular mechanism of ptox biosynthesis in this endangered medicinal plant.

Materials and methods

Plant growth conditions and MeJA treatment

Callus was induced from mature leaves of *P. hexandrum* in MS medium (Murashige and Skoog 1962) supplemented

with 2.68 µM alpha-Naphthaleneacetic acid (NAA) and 8.88 µM 6-Benzylaminopurine (BAP) (Chakraborty et al. 2010). Callus was subcultured after every 2 weeks in the above-mentioned medium. Cell suspension cultures were initiated from fresh subcultured green calli of *P. hexandrum* in modified liquid MS medium (Bhattacharyya et al. 2012) containing 60 mM total N_2 content, 1.25 mM potassium dihydrogen phosphate, 6 % glucose and 11.41 µM 3-Indoleacetic acid (IAA). An inoculum of 5 g cells was used in 50 ml of cell suspension culture medium. 6 flasks containing cell suspension cultures were shaken at 110 rpm for 12 days. 3 days old cell suspension cultures were treated with 100 µM MeJA. Cell suspension cultures, mock treated with ethanol (without MeJA), was used as the controls. Sampling was done in 0, 3rd (day of treatment with MeJA), 6th, 9th and 12th day. The cells were collected by centrifugation at 1000*g* for 5 min, and immediately put in liquid nitrogen and preceded to RNA isolation.

Extraction of ptox and HPLC analysis

Extraction of ptox of samples was done as described previously (Bhattacharyya et al. 2012). The media was extracted two times with ethyl acetate, evaporated and dissolved in methanol and combined with sample extracts for analysis and combined with sample extracts. Analysis of ptox was done by HPLC (Kartal et al. 2004). Column and instrumentation for analysis were used as standardized before (Bhattacharyya et al. 2012).

RNA extraction and 454 pyrosequencing

Total RNA was isolated from 12 days cell suspension cultures derived from green *P. hexandrum* callus samples using Purelink miRNA isolation kit (Invitrogen, Carlsbad, CA, USA). The total RNA was quantified on nanodrop, checked in 1 % denaturing agarose gel as well as on bioanalyzer. Removal of rRNA from total RNA was performed using the RiboMinus Plant kit for RNA seq, by Invitrogen as per standard procedure and then concentrated by RiboMinus concentration module (Ambion, Austin, TX, USA) as per standard procedure. Library preparation was preceded further as per protocol (cDNA Rapid Library Preparation Method Manual-GS FLX Titanium Series Roche, Mannheim, Germany). For transcriptome sequencing, 1 µg of Ribo-minus total RNA from each sample was used for fragmentation using $ZnCl_2$ solution followed by ds cDNA synthesis using a standard cDNA synthesis kit (Roche). This ds cDNA was processed for fragment end repair, followed by an adaptor ligation process using Rapid Library Prep kit (Roche). Small fragments were removed using a sizing solution supplied by Roche. Each final library was eluted in 53 µl of TE buffer. The DNA libraries were quantified on TBS 380 Fluorometer CA, USA and a RL curve were prepared to generate a standard curve of fluorescence readings and calculate the library sample concentrations for the preparation of emulsion PCR. Quality assessment of adaptor-ligated cDNA libraries were performed on high sensitivity DNA chip (HS chip) on Agilent 2100 bioanalyzer USA. The size distribution of the cDNA library of the sample on the HS chip was found about 500–2000 bp. The emulsion-based clonal amplification (emPCR amplification) of cDNA library was performed by following Roche emPCR Method Manual protocol. Clonally amplified cDNA library beads obtained from emPCR amplification reaction was deposited on a Pico Titre Plate (PTP) for sequencing using pyrosequencing chemistry. The next generation sequencing run for whole transcriptome analysis was performed on Roche 454 GS FLX.

De novo assembly

Raw reads obtained from 454 pyrosequencing were preprocessed by removing low-quality reads, and adapter/primer sequences using PRINSEQ (Schmieder and Edwards 2011). PRINSEQ also generates summary statistics of sequence and quality data. The preprocessed sequences were then assembled using assembly programs. Among the various programs available, we used GS De novo Assembler (Version 2.5.3) USA default parameters and optimized parameters. All high quality (HQ) reads were deposited in the NCBI and are accessible in Short Read Archive (SRA) under the accession numbers SRX180871 and SRX180389.

Functional annotation, GO mapping, pathway analysis, FPKM value determination

The annotation of transcripts contigs and singlets, were carried out using green plants of non-redundant (nr) protein database in the NCBI using BLASTX. GO mapping was carried out with BLAST 2GO (which assigns Gene Ontology GO; http://www.geneontology.org annotation) in order to retrieve GO terms for all the BLASTX functionally annotated transcript contigs. The GO mapping uses the following criteria to retrieve GO terms for annotated transcript contigs: (1) BLASTX result accession IDs are used to retrieve gene names or symbols, identified gene names or symbols are then searched in the species specific entries of the gene-product tables of GO database; (2) BLASTX result accession IDs are used to retrieve UniProt IDs making use of PIR which includes PSD, UniProt, SwissProt, TrEMBL, RefSeq, GenPept and PDB databases; (3) accession IDs are searched directly in the gene product table of GO database. KEGG maps (Kyoto Encyclopedia of Genes and Genomes, KASS) and an enzyme classification number (EC number) using a combination of similarity searches and

statistical analysis were built for pathway analysis. FPKM values for the transcripts have been determined using the formula, FPKM = (number of reads mapped × 10^9)/(length of transcript × total number of reads).

Protein domains and transcription factor identification in *P. hexandrum*

40,380 transcripts generated from Newbler Optimized parameter assembly and 3372 contigs generated from Newbler default parameter assembly have been searched in conserved domain database (CDD v3.07) with an E-value cutoff of 0.01 for probable PCBER (phenylcoumaran benzylic ether reductase), SDR (Secoisolariciresinol-DH-like-SDR-c), NADB-Rossmann superfamily, dirigent (dirigent superfamily), methyl transferases, monooxygenase, laccase, peroxidase, dioxygenase, oxidoreductase, and CAD domains. For the identification of transcription factor families represented in *P. hexandrum* cell culture transcriptome, the transcript contigs were searched against all the transcription factor protein sequences at plant transcription factor database (http://plntfdb.bio.uni-potsdam.de) using BLASTX with an E-value cutoff $1E^{-06}$.

Semi-quantitative RT-PCR (RT-qPCR) analysis

Total RNA was isolated from treated and control cells of *P. hexandrum* with Trizol (Invitrogen) following a standardized protocol (Ghanta et al. 2011) with three biological replicates and was used further. 2 µg RNA was reverse transcribed with oligo (dT) primers using RevertAid H Minus First Strand cDNA Synthesis kit (Thermo Scientific, Waltham, MA, USA) following manufacturer's instructions. Transcripts were detected by semiquantitative RT-PCR using gene specific primers and *Actin* as a loading control (Supplemental Table S1).

Cloning, sequence analysis, and functional expression of *Podophyllum* CAD, protein identification and immune detection and enzyme assay

Full lengths of four isoforms were obtained from contigs, resulted from 454 pyrosequencing of MeJA treated cell culture of *P. hexandrum*, by 5′-RACE and 3′-RACE strategy using SMARTer™ RACE cDNA Amplification Kit (Clonetech, Mountain View, CA, USA). CAD full length cDNA was deduced by overlapping 5′-RACE and 3′-RACE fragments. Gene-specific primers (Supplemental Table S1) for each isoform were used to amplify the full length cDNA. PCR products were ligated into pGEM-T vector (Promega, Madison, WI, USA) for sequencing (Xcelris Labs Ltd. Ahmedabad, India).

For recombinant protein expression *PhCAD*1, *PhCAD*2 and *PhCAD*3 were cloned in pET-15b between restriction sites of *Xho*I and *Bam*HI and *PhCAD*4 was cloned into pET-28a vector between restriction sites of *Bam*HI and *Hin*dIII. pET-15b vectors containing *PhCAD*1, *PhCAD*2, *PhCAD*3 and pET-28a containing *PhCAD4* were transformed into BL21 (DE3) strain for their expression. Culture of transformed BL21 (DE3) of each isoform was induced with 0.8 mM IPTG when OD_{600} reached at 0.4–0.6. Culture was further grown for 19 h at 22 °C in the dark. Cells were harvested by centrifugation, resuspended in 15 ml of lysis buffer (50 mM Tris–HCl, pH 8.0, 300 mM NaCl, 10 mM imidazole and 10 % glycerol) containing 1 mg/ml lysozyme and protease inhibitor cocktail. Cells were incubated for 30 min on ice and lysed by sonication with 10 pulse-rest cycles (5 s pulse at 25 W with a 1 min interval after each pulse). Purification of recombinant protein was performed using Ni-NTA column (Qiagen, CA, USA) by the method of native purification. The Ni-NTA purified protein was subjected to extensive dialysis using dialysis buffer using a Protein refolding kit (Novagen, Madison, WI, USA) and stored in 20 % glycerol. From the total cell protein analysis, proteins were extracted and then run into denaturing 12 % (w/v) SDS-PAGE. The gels were stained with Coomassie blue G-250.

Immunodetection of recombinant PhCADs from BL21 (DE3) was performed with Penta His antibody (Qiagen) as the primary antibody according to the manufacturer's instructions. PhCAD1, PhCAD2, PhCAD3, PhCAD4 were incubated with CAld and sinapaldehyde for 3 min at 30 °C with 104 µM NADPH, 150 µM Tris–HCl buffer pH 7.5 and 64 µg of protein. The products were separated by HPLC according to Santos et al. (2006). CAD activity was expressed as µM CAld or sinapaldehyde consumed per second per milligram protein. Maximum velocity (Vmax) and Km were determined by Michaelis–Menten equation using GraphPad Prism version 4.00 for Windows (GraphPad, Software, La Jolla, CA, USA). Enzyme activity assay for different pH and temperature of each isoform were performed as mentioned above.

Fold prediction and homology modelling of CAD from *P. hexandrum*

Four CAD protein sequences (PhCAD1, PhCAD2, PhCAD3, and PhCAD4) from *P. hexandrum* were subjected to HHPRED (Hildebrand et al. 2009) server for the prediction of secondary and tertiary structure. Fold prediction results suggest a strong structural similarity of each PhCAD sequence with the three-dimensional (3D) structure of sinapyl alcohol dehydrogenase (SAD) from *Populus tremuloides* (PDB code: 1YQD). The 3D coordinates of the individual PhCAD dimer was generated by MODELLER 9.8 (Eswar et al. 2008) package using chain A and chain B of 1YQD as reference structures. Suitable 3D models were generated using the MODELLER v9.8 package and filtered based on the best energy parameters (MOLPDF

and DOPE scores) and were further validated using PROCHECK (Laskowski et al. 1993) and Verify 3D (Eisenberg et al. 1997) structure validation tools. Structural superposition of each CAD and SAD structures was performed using the MUSTANG (Konagurthu et al. 2006) program while multiple sequence alignment of CAD sequences were done by the MAFFT (Katoh et al. 2005) program.

Molecular docking analysis

Structural cavity analysis was performed using the CASTp web server (Dundas et al. 2006). Molecular docking of different aldehyde ligands onto each PhCAD model was performed using the GOLD v5.2 (Genetic Optimization for Ligand Docking) package from the Cambridge Crystallographic Data Centre (Jones et al. 1997). The 3D coordinates of CAld and sinapaldehydes were collected from the PubChem (Bolton et al. 2008) database. GOLD software optimizes the fitness score of many possible docking solutions using a genetic algorithm. Following parameters were used in the docking cycles: population size (100), selection pressure (1.100000), number of operations (100,000), number of islands (5), niche size (2), crossover weight (95), mutate weight (95) and migrate weight (10). 100 docking calculations were run for each ligand and the best docking solutions were identified based on critical manual inspection satisfying favorable interactions between the ligand and the protein molecule. GOLDscore (GOLD fitness score) was used to further identify the lowest energy-docking mode among the manually selected docking solutions.

Substrate preference for each PhCAD was estimated by the ratio of (Kcat/Km) of CAld over (Kcat/Km) of sinapaldehyde using the following formula,

$$Experimental\ \ fold\ change = \frac{\left(\frac{\text{Kcat}}{\text{Km}}\right)\text{of CAld}}{\left(\frac{\text{Kcat}}{\text{Km}}\right)\text{of sinapaldehyde}}$$

Similarly, computational fold changes were calculated, by the ratio of CAld docking score with respect to sinapaldehyde docking score for each PhCAD

$$Computational\ \ fold\ change = \frac{\text{Docking score of CAld}}{\text{Docking score of sinapaldehyde}}$$

Selected docked solutions were further refined and restored using the FireDock docking refinement server (Mashiach et al. 2008). Both the GOLD fitness scores and the global energy score obtained from FireDock have been considered for computational fold change calculation.

Approximate ΔG_{com} of binding of the selected docked complexes were calculated using the MOPAC program package. Similarly, experimentally derived binding constants (Km) were converted to ΔG_{exp} using the following formula.

$$\Delta G = -RT\ln K_m$$

where, R = 0.0019872041 cal K^{-1} mol^{-1} and T = 297 K.

Difference in ΔGs (ΔΔG) for CAld and sinapaldehyde binding for each PhCAD was calculated using the following formula

$$\Delta\Delta G = \Delta G_{\text{CAld}} - \Delta G_{sinapaldehyde}.$$

Phylogenetic tree analysis

P. hexandrum CAD sequences are subjected to homology search using BLASTp (Altschul et al. 1990) against NR (Pruitt et al. 2007) database with threshold criteria of E-value: $1e^{-05}$, sequence identity of 40 % and query or subject coverage of 70 %. Plant specific CAD sequences are collected and they are subject to CD-hit (Li and Godzik 2006). All the collected sequences further they are aligned with MAFFT (Katoh et al. 2002). The resulting alignment was utilized to create a combined tree by RAXML (Stamatakis 2014) program using default parameter settings. Tree image is generated using Fig Tree v1.4.2 program. The figure shows a phylogenetic tree of *Sinopodophyllum*CADs and all other related CAD sequences from different plant species.

Construction of transgenic and HPLC analysis

The *PhCAD*1, *PhCAD*2, *PhCAD*3, *PhCAD*4 genes were cloned into pGEM-T Easy vector (Promega). *PhCAD*1, *PhCAD*3 then subcloned into the binary vector *pBI*121 between *Xba*I and *Sac*I sites and *PhCAD2* and *PhCAD4* between *Bam*HI and *Sac*I sites to give rise to the final construct 35S::*PhCAD*1, 35S::*PhCAD*2, 35S::*PhCAD*3, 35S::*PhCAD*4. Transformation of *P. hexandrum* cell culture and callus were carried out by the standard *Agrobacterium tumefaciens*-mediated method with modifications (Kim et al. 2009). Briefly, transformed *A. tumefaciens* (GV3101) was grown up to OD_{600}, centrifuged and the pellet resuspended in half strength MS medium containing 3 % sucrose. 2 ml of this *A. tumefaciens* suspension was added to cell culture and incubated in the dark for 2 days without agitation. The suspension cells were then washed four times with fresh liquid medium. To obtain stable transformants, cells were maintained in cell culture with 50 mg ml^{-1} kanamycin and 100 mg ml^{-1} cefotaxime. Transformants that were

PCR positive both for marker gene, i.e. *npt*II and the gene of interest were selected. The overexpression of genes was confirmed by semi RT-qPCR after 8 days of transformation. Analysis of ptox and lignin were done after 12 days of subculturing by HPLC and acetyl bromide methods respectively. Construction of independent transgenic lines of *Nicotiana tabacum* overexpressing four isoforms under *CaMV35S* promoter was done according to Ghanta et al. 2011. Transformation efficiency was checked by GUS assay of control transgenic callus, *N. tabacum* overexpressing *GUS* gene under *CaMV35S* promoter of *pBI*121, according to the previously standardized method in our lab (Datta et al. 2015). Further experiments with transgenic *N. tabacum* and callus of *P. hexandrum* were done with 4 weeks old plant sample after successful selection of transgenic plant material.

Determination of lignin content

Lignin content of cells was analyzed using the acetyl bromide procedure (Dence 1992). Briefly, 0.1 ml acetyl bromide (25 % in acetic acid) and 4 µl 70 % Perchloric acid were added to acetone washed and oven dried 1 mg sample and the mixture was incubated at 70 °C for 40 min. 200 µl 2 M NaOH and 0.5 ml acetic acid were then added to the test tube after incubation. After centrifugation at 15,000*g* at room temperature for 15 min, the supernatant was separated from the pellet. The pellet was further washed by adding 500 µl acetic acid (vortex for 5 s), followed by centrifugation at 15,000*g* at room temperature for 15 min. The second supernatant was combined with the first. Acetic acid was added to the combined supernatants to a volume of 2 ml, and absorption was measured at 280 nm using a UV-visible spectrophotometer. An extinction coefficient for lignin at 280 nm of 23.35 L g^{-1} cm^{-1} was used (Chang et al. 2008).

Mass spectroscopy of ptox

Ptox was extracted and purified by HPLC as above mentioned method from the control, treated and four transgenic cell lines of *P. hexandrum*. The purified ptox was further placed for mass spectroscopy to compare with a mass spectrum of standard ptox. Mass spectroscopic analysis was done by JEOLJMS700 FAB Mass spectrometer using glycerol as a matrix.

Results

MeJa treatment and estimation of ptox content

In the current investigation, we studied the time-dependent ptox accumulation in MeJA treated cell cultures of *P. hexandrum* till 12 days. Initially, ptox increased steadily till 6th day followed by an abrupt enhancement from 6th to 9th day and thereafter a steady increase of ptox up to 12th day resulted in a final 8–10 fold enhancement of its content in comparison to '0' day control cell cultures (Fig. 1).

454 pyrosequencing of the *P. hexandrum* cell suspension culture transcriptome in response to MeJA and denovo assembly using optimized Newbler parameters

1,601,729 unique HQ reads with an average read length of 138 bp and the total length of 220,593,962 bp was obtained from 2,645,631 raw reads (Table 1). 525,953 high-quality sequencing reads were unique and mapped to Rfam, non-coding RNA database. Further analysis was preceded by, approximately, 50 % of the filtered reads.

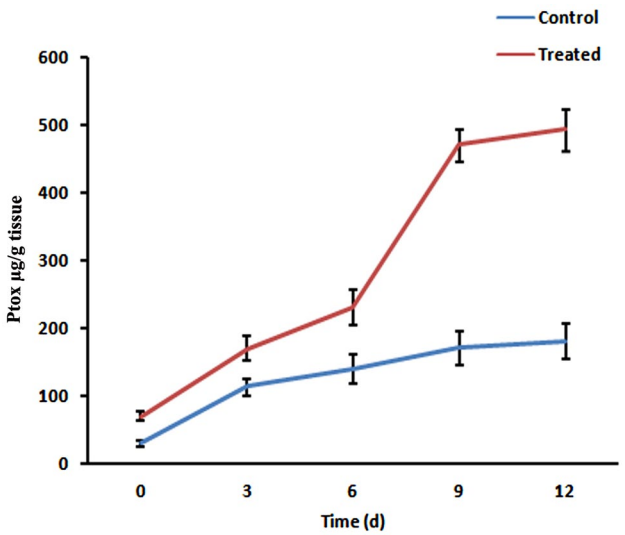

Fig. 1 Ptox accumulation measured by HPLC from control and MeJA treated cell cultures. Data are the mean ± SD of three independent experiments

Table 1 Overview of sequencing reads and reads after preprocessing

Statistics	454 pyrosequencing of *P. hexandrum* cell culture sample treated with MeJA
Sequencing reads before preprocessing	2,645,631
Number of unique HQ	1,601,729
Average length of HQ read (bp)	138 bp
Total length (bp)	220,593,962 bp
Number of reads used for assembly	525,953

Comparison between Newbler default and optimized parameters

We checked the options "Use duplicate reads", "extend low depth overlaps", "Reads limited to one contig" and "Single Ace file options" and thereby optimized the parameters for Newbler assembly to increase the number of reads and contigs. Our observations suggest that the number of non-assembled singlets was around threefold higher than Newbler default parameters and assembled contigs were comparatively lower in number (Table 2). 80.2% of the transcripts generated from Newbler optimized assembly (7487 contigs and 22,407 singlets), were 200–300 bp in length, 14% were 300–500 bp in length and 5.75% were greater than 500 bp in length (Supplemental Fig. S1). From 4231 contigs assembled by Newbler default assembly, 66.9% transcript contigs were of 200–300 bp length, 25.4% between 300–500 bp and 7.7% were greater than 500 bp in length.

Functionally annotated transcripts and differential expression of phenylpropanoid pathway genes

Annotation of transcripts was carried out using green plants of non-redundant (nr) protein database at the NCBI by BLASTX (Altschul et al. 1997). BLASTX resulted in the annotation of 29,894 contigs/singlets out of 7487 contigs and 22,407 singlets using Newbler optimized parameter and 4231 transcript contigs using Newbler default parameters (Supplemental Table S2 and S3). 62,725 singletons obtained using Newbler default parameters were not annotated since 12,852,082 bases were used and mean singlet length was only 204.89 bp. Top-hit species distribution of contigs/singlets obtained from default parameters showed the highest similarity towards *Populus trichocarpa*. In contrast, contigs/singlets from optimized parameters showed the highest similarity towards *Sorghum bicolour* (Supplemental Fig. S2). Both E-value score and alignment length of contigs/singlets were better in the case of the optimized parameter (Supplemental Fig. S3 and S4). GO (Conesa et al. 2005) was used to classify the functions of the predicted transcript contigs. From the transcripts obtained using Newbler optimized parameters, 3901 transcripts were classified as molecular functions, 3204 transcripts were classified as biological processes, and 2812 transcripts could be annotated to cellular components. Using default parameters, maximum numbers of transcript contigs were allocated to molecular function (1208), followed by biological processes (1026) and cellular component (892) (Fig. 2; Supplemental Table S4, S5). Digital expression profile was expressed as fragments per kilobase of exon per million mapped (FPKM values) of each transcript, obtained from both default and optimized Newbler assembly (Supplemental Table S6 and S7). Out of these 452 differentially expressed (p-value < 0.05) transcripts obtained from default parameters, 89 are down-regulated while 363 transcript contigs were found to be significantly up-regulated (Supplemental Table S8). Among 773 annotated transcripts identified as differentially expressed and overlapping between control and treated samples by Newbler optimized parameter assembly, 591 transcripts were upregulated in treated samples, while 182 transcripts were downregulated as noted from log of fold change from their FPKM values (Supplemental Table S9). A heat map clustering of 50 upregulated and 50 downregulated genes for both default and optimized parameter assembly has been generated (Fig. 3).

29,894 annotated sequences were mapped by KEGG (Kanehisa et al. 2004). The KEGG biochemical mapping of pathways for Newbler default parameter assembly resulted in 537 unique non-redundant sequences and 66 unique sequences could be assigned to secondary metabolism. Again assembly using Newbler optimized parameters and subsequent annotation resulted in 1395 unique sequences and 104 unique sequences could be classified to secondary metabolism (Supplemental Table S10 and Table S11).

Table 2 Assembly results of 454 data using various programs

Description	Newbler (optimized parameter)		Newbler (default parameter)	
	Contigs	Singlets	Contigs	Singlets
Number of contigs/singlets	7487	22,407	4231	62,725
Total bases of contigs/singlets (bp)	3,016,258	5,055,301	1,327,609	12,852,082
Mean contig/singlet length	402	226	313	205
Contig/singlet N50	443	216	294	193
Max contig/singlet size	2935	399	2295	584

Additional insight into probable downstream ptox pathway genes by domain searching and comparison with control cell cultures

Important functional domains related to ptox biosynthesis pathway were identified from default and optimized parameters (Table 3; Supplemental Table S12 and S13). Important pathway genes were identified from control and treated data and generated by optimized parameters (Supplemental Table S14). Multiple transcripts were annotated as secoisolariciresinol dehydrogenase (SDR) from the optimized parameter (Supplemental Table S15). Comparison of treated and control data according to their

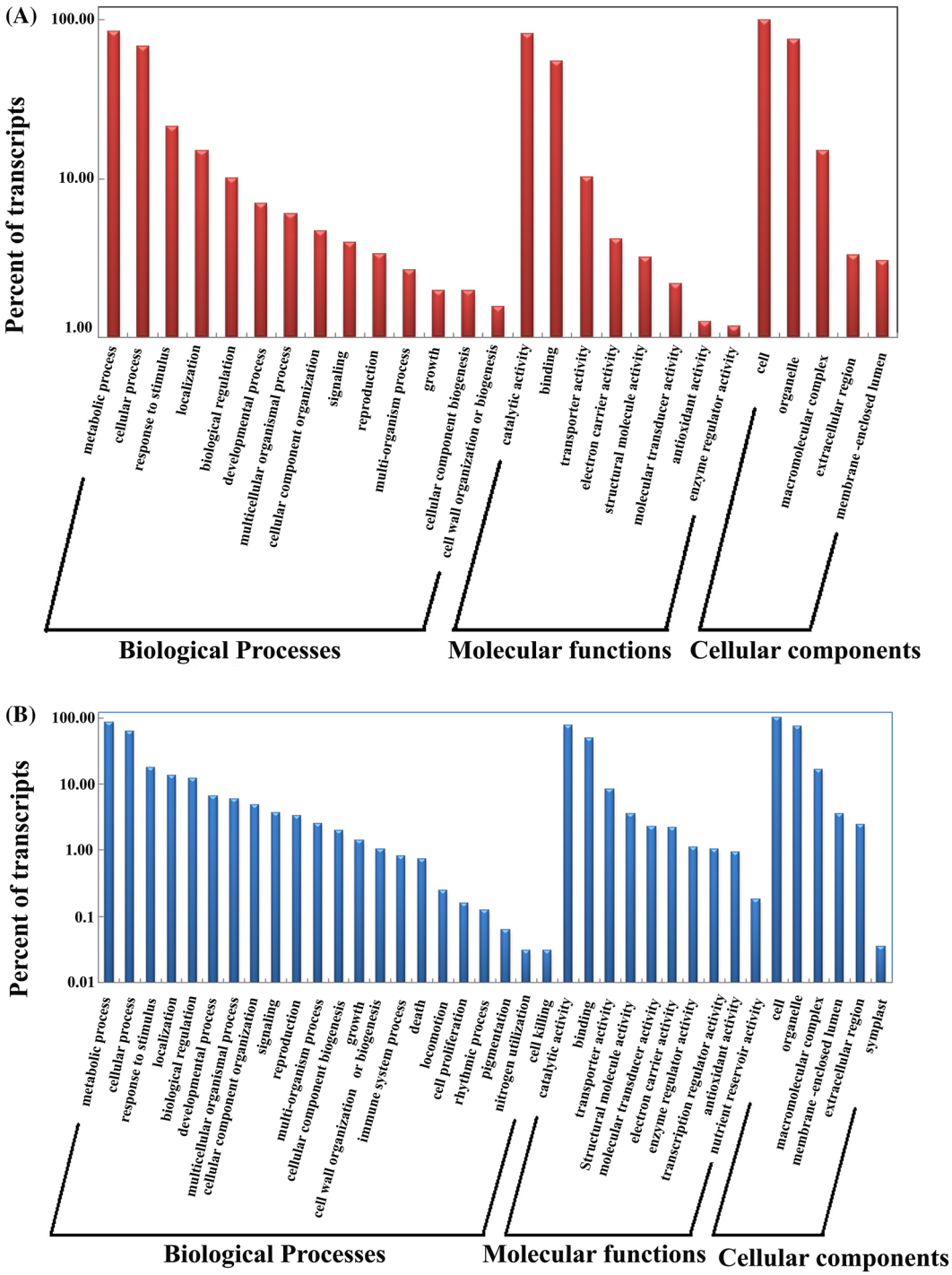

Fig. 2 Gene ontology-level sequence distribution. GO-level sequence distribution for Biological processes, molecular functions, and cellular components of transcript contig/singlets generated by Newbler **a** default and **b** optimized parameter assembly

FPKM value interestingly revealed that phenylalanine ammonia lyase (*PAL*) remained unaffected after treatment. In contrast, other pathway genes like *HCT* (hydroxyl cinnamoyl transferase), *C4H* (cinnamate-4-hydroxylase), *4CL* (4-Coumarate Ligase), *CCR* (cinnamoyl reductase), *CAD, cytochrome P450* and *SAM-dependent methyltransferase* were upregulated in the treated culture (Supplemental Fig. S5). Among those genes *CAD* and *4CL* were most significantly upregulated, whereas *C4H* and *HCT* were only noted in the treated cultures.

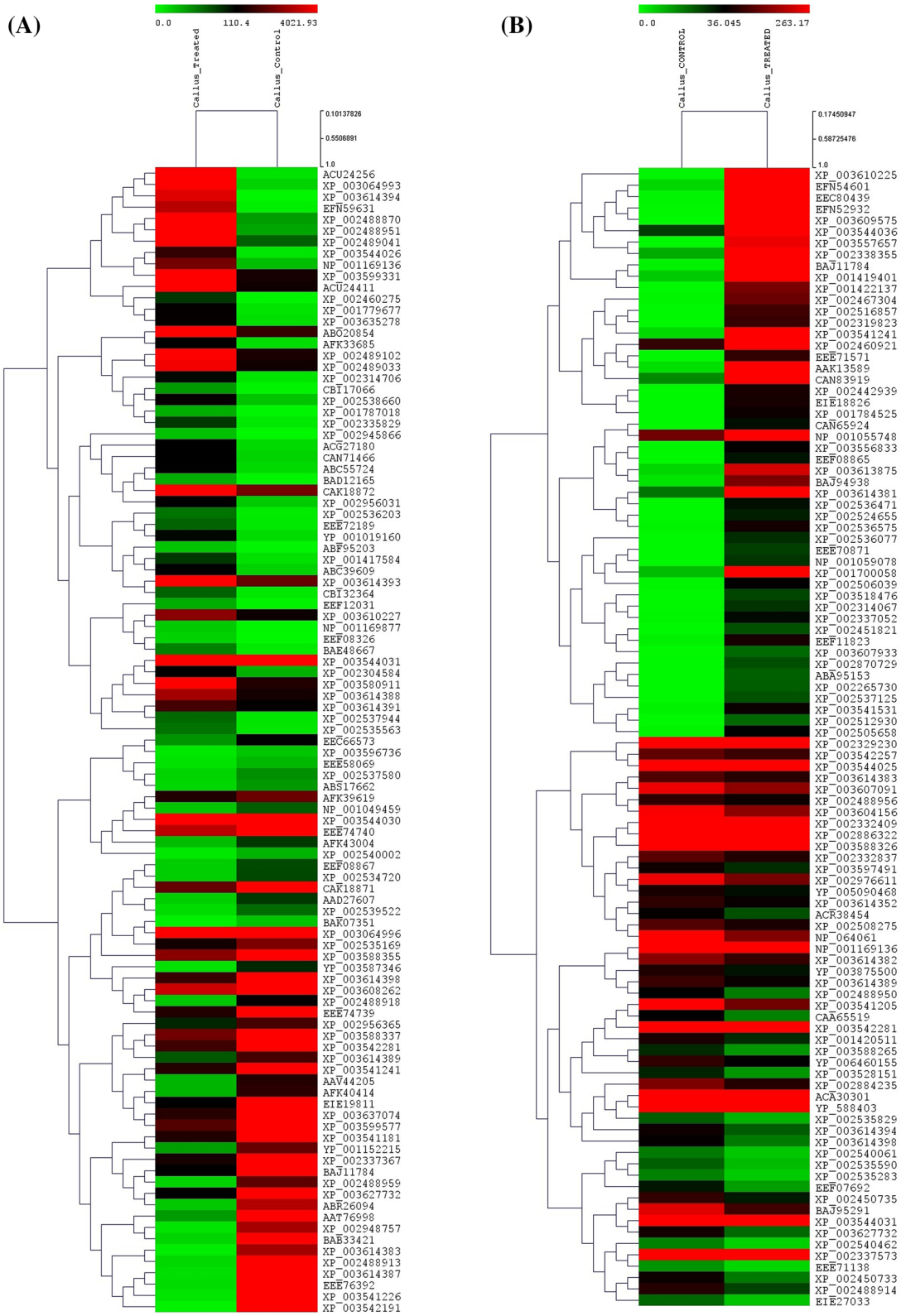

Fig. 3 Heat map cluster of 50 upregulated and 50 downregulated genes for **a** default and **b** optimized parameter assembly

Table 3 Important domains that may regulate ptox biosynthesis

CDD identifier	Identifier domain name	Number of cDNAs
cl09931	NADB_Rossmann superfamily	69
cl00480	Methyltransf_6 superfamily	3
cl07892	Flavin_utilizing_monoxygenases superfamily	2
cl00196	plant_peroxidase_like superfamily	4
cl14733	DIOX_N superfamily	5
cl17060	Oxidored_q1 superfamily	7
cl08484	GHMP_kinases_C superfamily	1
cl17037	NBD_sugar-kinase_HSP70_actin superfamily	6
cl09155	Pyruvate_Kinase superfamily	4
cl00192	ribokinase_pfkB_like superfamily	6
cl00780	Kinase-PPPase superfamily	1
cl12078	CypX superfamily	15
cl00464	URO-D_CIMS_like superfamily	1
cl00210	Isoprenoid_Biosyn_C1 superfamily	5
cl16911	AdoMet_MTases superfamily	10

Transcription factors related to secondary metabolism and comparison with the control culture

We identified 474 transcripts with an optimized Newbler assembly that may code for probable TFs. Among them, 16 transcripts were *AP2-EREBP* TFs, 27 were *NAC* TFs, 12 transcripts were *bHLH* TF, 24 transcripts coded for *MYB* or *MYB* related TFs, 12 transcripts coded for *mTERF* TFs, 9 transcripts coded for *WRKY* TFs, and 2 transcripts coded for *C2C2*-like TF. Different TFs identified from transcripts generated by Newbler default and optimized parameters assembly were represented according to the number of transcripts annotated for respective TFs (Fig. 4a, b; Supplemental Table S16). Comparison of TFs obtained from treated and control data according to their FPKM values revealed that the TFs related to phenylpropanoid pathway like *AP2-EREBP, NAC, bHLH, MYB, mTERF* and *WRKY* were upregulated in the treated cultures. Among the upregulated TFs *bHLH* and *NAC* were most significantly upregulated. In contrast, *C2C2* was slightly downregulated in treated culture (Supplemental Fig. S6).

Validation of pyrosequencing results

Semi RT-qPCR analysis of 18 randomly selected diferentially expressed genes (9 upregulated and 9 downregulated transcripts each) was conducted to confirm the expression patterns in the treated and control samples (Supplemental Fig. S7).

Expression profile of selected phenylpropanoid pathway genes in control and treated cell culture

Transcripts annotated as CAD in the transcriptome analysis, shared similarity to *Arabidopsis AtCAD*s and were referred to as *CAD*1, *CAD*4, *CAD*5, *CAD*7, *CAD*8, and *CAD*9 according to the *Arabidopsis CAD* isoforms (Supplemental Table S14). *CAD*1 and *CAD*8 were upregulated in treated samples with significant upregulation of *CAD*8 (Supplemental Fig. S8a). *CAD*4 and *CAD5* were downregulated. *CAD9* expression did not vary in control or in the treated samples. Comparison of relative gene expression levels between control and treated samples revealed upregulation of almost all the phenylpropanoid pathway genes with marked upregulation of *PAL, CCR* and *4CL* in treated cultures (Supplemental Fig. S8b).

Raising of full length of *CAD* isoforms, cloning and expression in *E. coli*

Full lengths of four CAD isoforms, namely *PhCAD1, PhCAD2, PhCAD3*, and *PhCAD4* were raised by 5′ and 3′ RACE from the contigs which were obtained from transcriptome analysis of MeJA treated cell suspension cultures of *P. hexandrum. PhCAD1, PhCAD2, PhCAD3* and *PhCAD4* have open reading frames of 1086, 1155, 1077 and 1092 bp respectively. *PhCAD1, PhCAD2* and *PhCAD3* were very closely related and showed more than 74% of similarity between them, while *PhCAD*4 was distantly related to the other three isoforms, showing less than 69% of similarity (Supplemental Fig. S9). For each of these four isoforms, the recombinant proteins, after purification, were noted with a major band at the expected molecular weight, which were further confirmed by western blot analysis using anti-His antibody (Fig. 5a, b). Theoretically, PhCAD1, PhCAD2, PhCAD3 and PhCAD4 have a relative molecular mass of 38.8 kD, 42.1 kD, 38.7 kD and 38.9 kD respectively, with theoretical pI values of 6.5, 6.2, 6.0 and 5.9 respectively.

Fold prediction and homology modelling of CAD from *P. hexandrum* and molecular docking analysis

Multiple sequence alignment (Fig. 6) of *P. hexandrum CAD*s (PhCAD1-4) and a sinapyl alcohol dehydrogenase (SAD) (PDB code: 1YQD) showed that in PhCAD1 the G302 from the SAD is mutated to C302 and A293 is mutated to M293. Whereas, in PhCAD2, these are mutated to A300 and D290, respectively. In PhCAD3 and PhCAD4 the G302 remains conserved, but the A293 is mutated to T290 and M296, respectively. However, other variations within the active site residues are also evident.

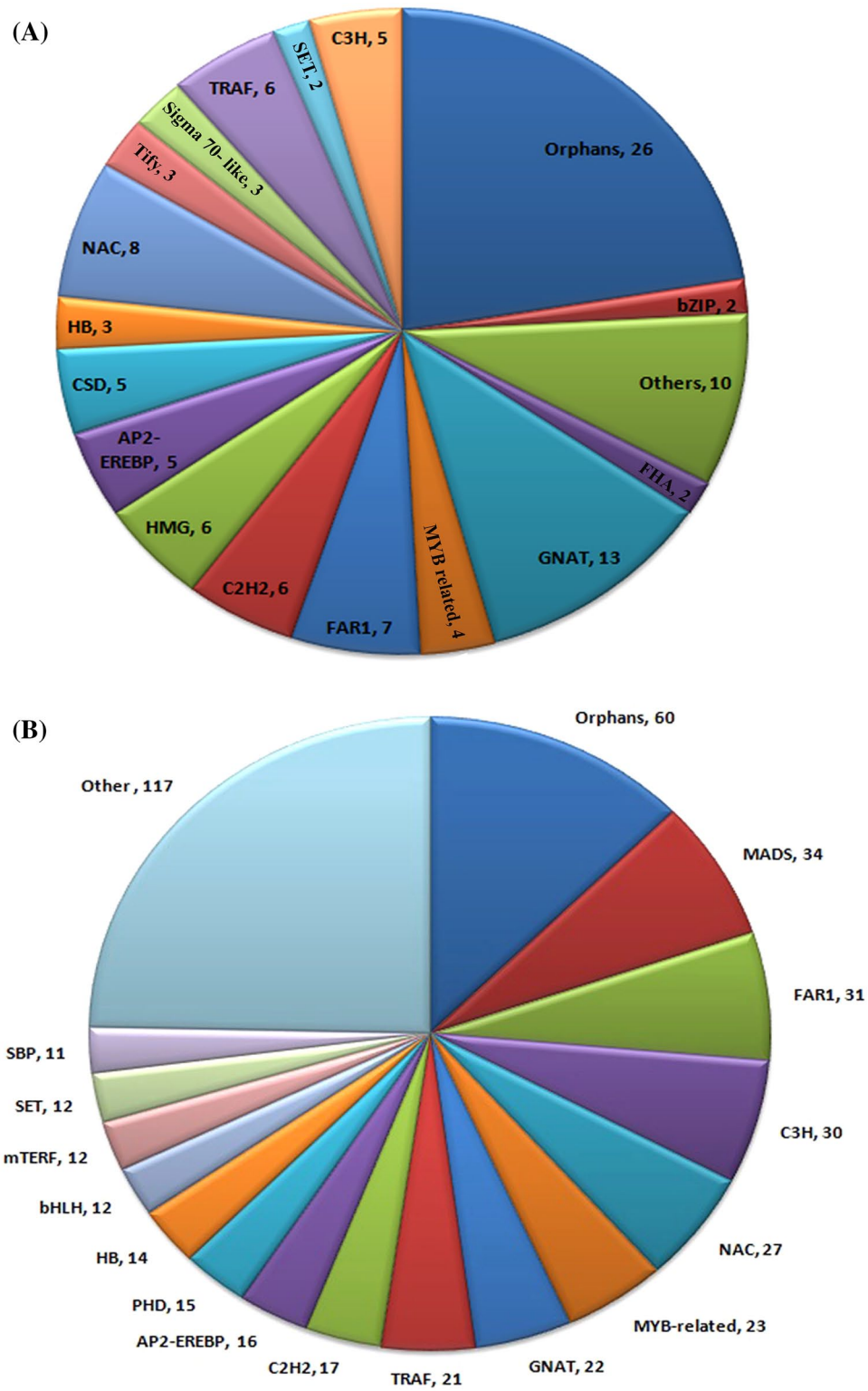

Fig. 4 Different transcription factors identified from transcripts generated by Newbler **a** default and **b** optimized parameter assembly

Despite the significantly high sequence similarities among the PhCAD sequences structural variations in terms of stability and cavity size were observed within the 3D models of PhCADs (Supplemental Fig. S9, S10). Phylogenetic analysis of PhCAD1, PhCAD2, PhCAD3 and PhCAD4 with all other related CADs from different plant species showed that PhCAD1, PhCAD2 and PhCAD3 are clustered together whereas PhCAD4 is positioned very distant from the other three *P. hexandrum* CADs and clustered with the CAD of *Medicago truncatula* (Fig. 7; Supplemental table S17). To understand the possible modes of ligand/substrate-binding, we further simulated the binding modes

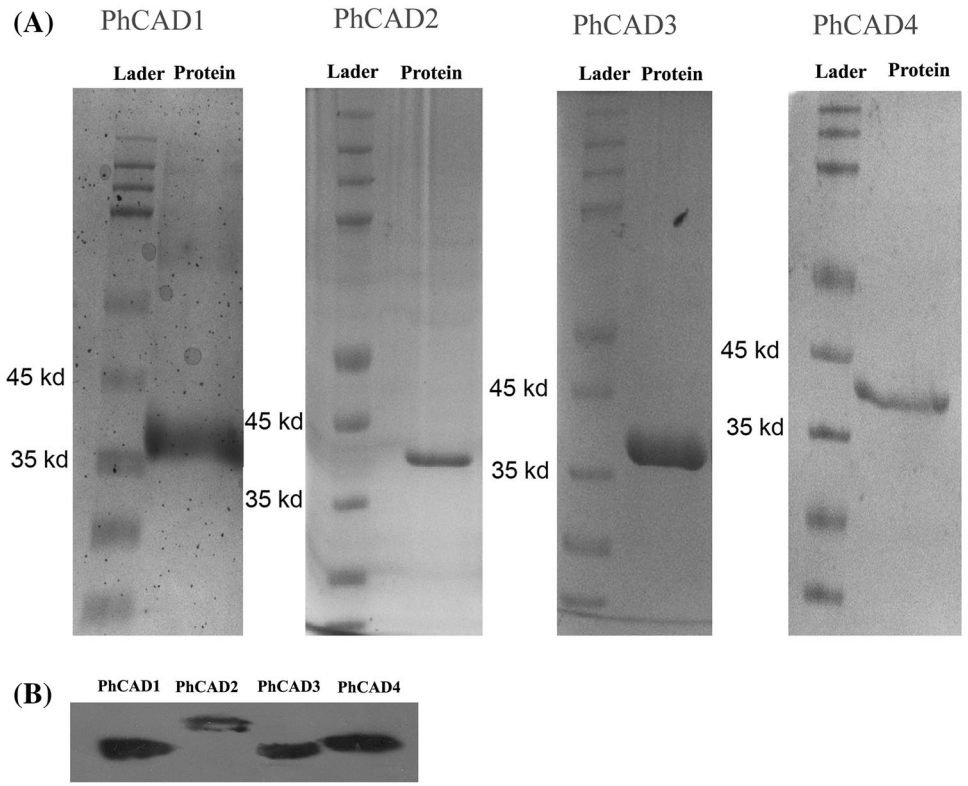

Fig. 5 **a** SDS PAGE electrophoresis His tagged niNTA purified proteins. **b** Western blot analysis of four recombinant His tagged isoforms namely PhCAD1, PhCAD2, PhCAD3, PhCAD4 by Penta His antibody

of substrates (CAld and sinapaldehyde) with the PhCAD proteins using the GOLD software package version 5.2 (Jones et al. 1997). Figure 8a shows the changes in the fold ratio where the computational docking score of each PhCAD is matched with the experimentally derived binding/affinity value.

Both experimentally and computationally derived scores suggest highest preference of CAld binding over sinapaldehyde for PhCAD3 (Fig. 8a). Figure 8b showing the correlation between the experimental and computational ΔΔG, clearly suggests that the PhCAD3 has comparatively higher and preferential binding with CAld over sinapaldehyde, followed by PhCAD4, PhCAD2, and PhCAD1.

Figure 9a–d shows the docking poses of CAld while Fig. 9e–h provides best sinapaldehyde docking poses with respect to PhCAD1, PhCAD2, PhCAD3, and PhCAD4, respectively. In PhCAD1 the cinnamyl ring and the methoxy substituents of coniferaldehyde are stabilized quite well within the active site, sandwiched between the hydrophobic cleft formed by different parts of chain A (I303, N115, and Y116) and chain B (F289, I292, and M293). The phenolic hydroxyl moiety is oriented toward the backbone carbonyl of I292 and M293 of monomer B. The cinnamyl ring is also positioned on top of C302 residue forming the base of the active site. In PhCAD2, the cinnamyl ring and the methoxy substituents are also sandwiched between the hydrophobic cleft formed by different chain A (I301, N113, and Y114) and chain B (L287, D290, and S291) residues. However, the cinnamyl ring is positioned on top of the A300 residue forming the base of the active site.

In PhCAD3 and PhCAD4 the cinnamyl ring and the methoxy substituents of CAld are stabilized within the active site, sandwiched between the hydrophobic cleft formed by different parts of chain A (I300, N112 and Y113), (V306, N115 and Y116) and chain B (F285, L289 and T290), (F292, I295 and M296), respectively. The phenolic hydroxyl moiety of PhCAD3 and PhCAD4 is oriented toward the backbone carbonyl of L289, T290 (in PhCAD3) and I295, M296 (in PhCAD4). Interestingly, the cinnamyl rings are positioned on top of G299 (in PhCAD3) and G305 (in PhCAD4) while an equivalent glycine (G302, see Fig. 9) in SAD was suggested to be crucial for substrate specificity (Bomati and Noel 2005).

In vitro enzymatic characterization of four PhCAD isoforms

In vitro enzyme assay showed that four enzymes have more Kenz (Kcat/Km) value for CAld than sinapaldehyde (Table 4; Fig. 10). PhCAD1 resulted in highest Kenz value for both the substrates of CAld and sinapaldehyde, followed by PhCAD3, PhCAD4, and PhCAD2. Kenz values for CAld was 1.23 fold higher than sinapaldehyde in the case of PhCAD1, whereas in the case of PhCAD3 Kenz value for CAld was 16.36 fold higher than sinapaldehyde. This suggests that PhCAD1 has almost equal preference for both

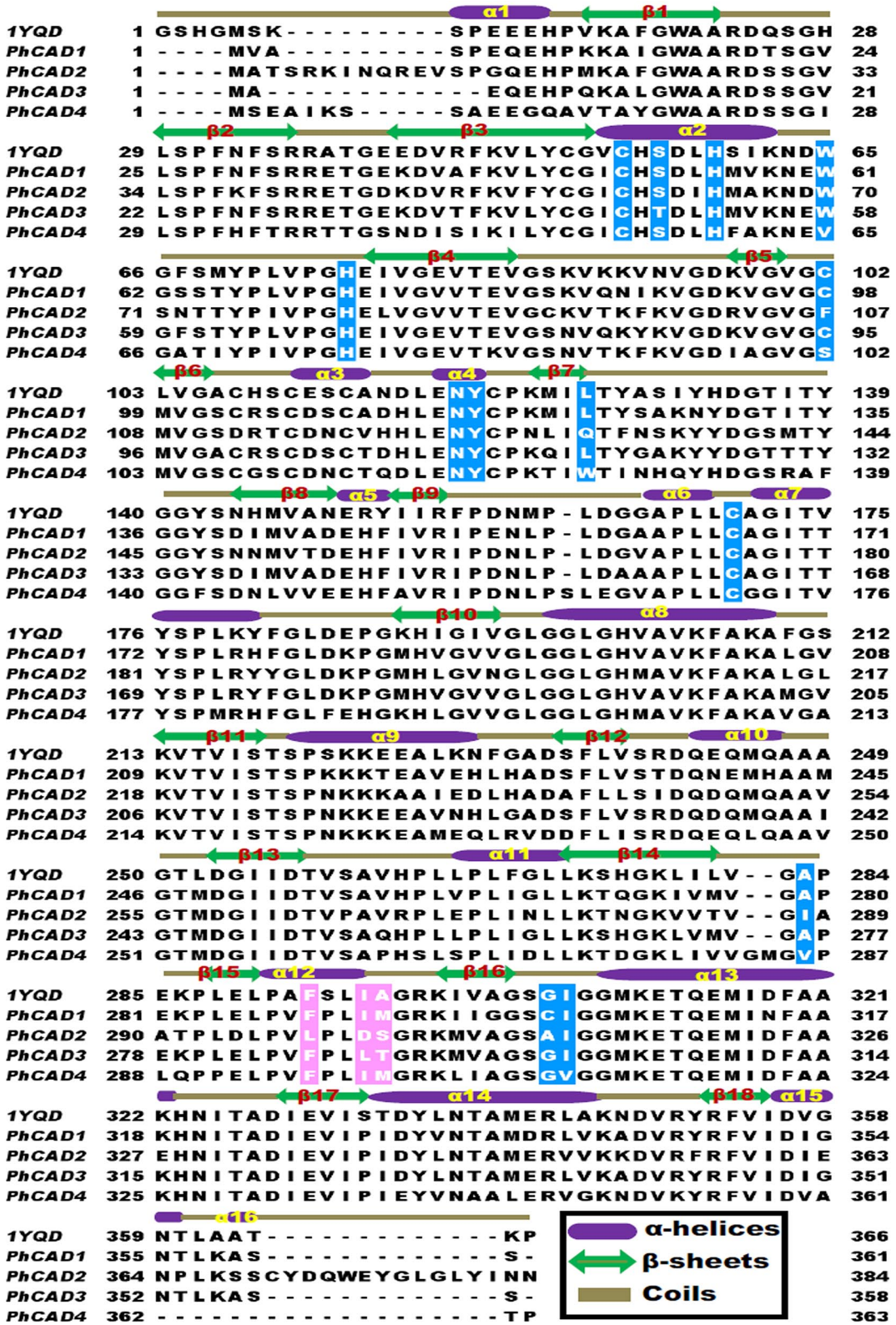

Fig. 6 Multiple sequence alignment of PhCADs and 1YQD sequences were done by the MAFFT (6) program. The secondary structure information of 1YQD is projected over the alignment. Active site residues from chain A (colored *blue*) and chain B (colored *pink*) are marked

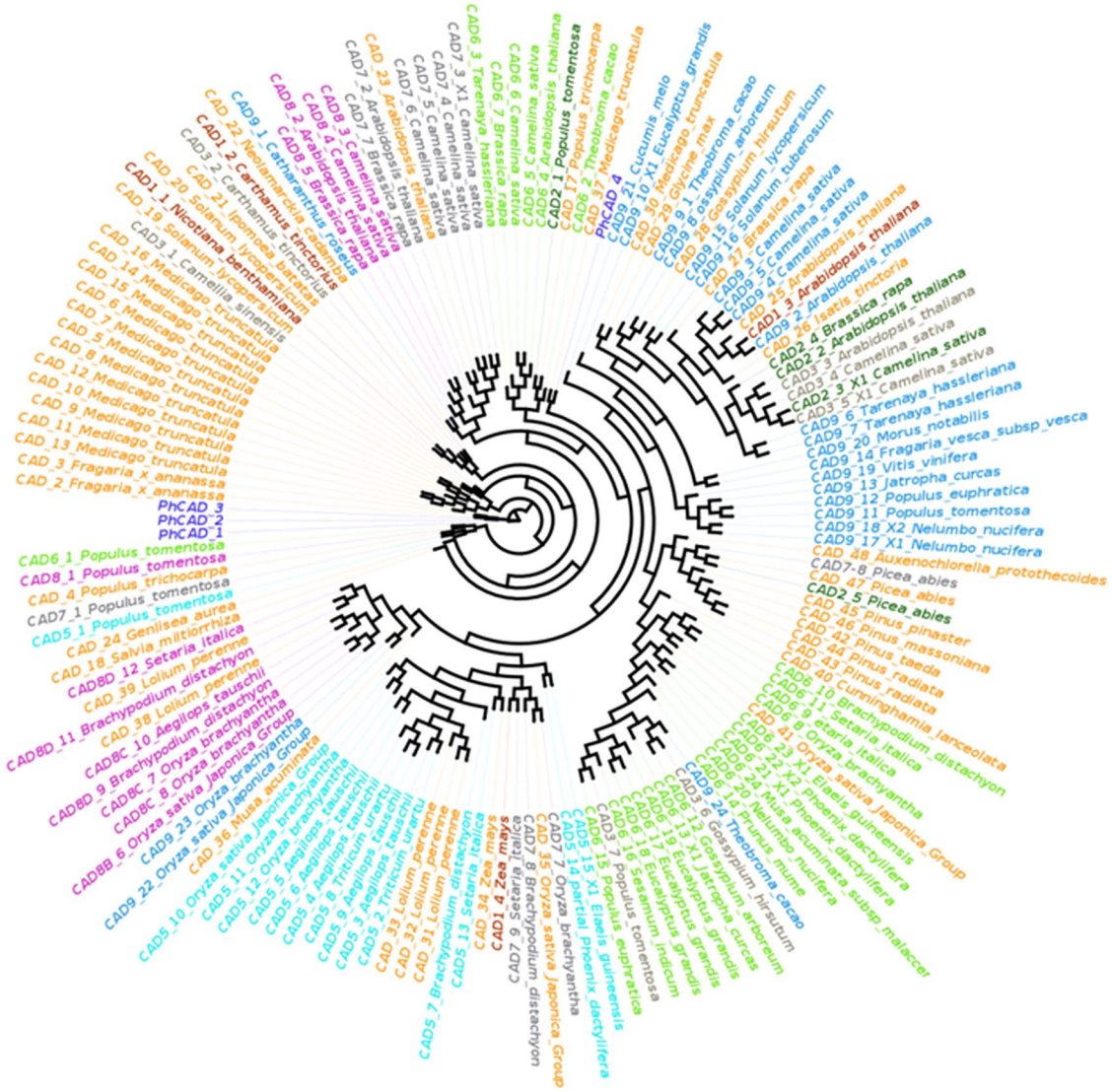

Fig. 7 Phylogenetic tree of *Sinopodophyllum*CADs and other related CAD sequences

the substrates while PhCAD3 has higher preferential affinity towards CAld than sinapaldehyde followed by PhCAD4 and PhCAD2. HPLC chromatogram of enzyme assays of the four isoforms for both of the substrate have been performed (Fig. 11). Optimum pH and temperature for each of four isoforms were 7.5 and 30 °C respectively (Fig. 12).

Accumulation of ptox and lignin in four PhCAD isoforms overexpressing transgenic lines

Transformed lines of *P. hexandrum* were selected against kanamycin selection and were further confirmed by the expression of *npt*II gene by PCR analysis (Fig. 13a). Overexpression of the four isoforms in respective lines were confirmed by semi RT-qPCR analysis using gene specific primers independently (Fig. 13b). For each of the lines, fold change of ptox was higher than lignin in comparison to control. The line overexpressing PhCAD1 showed maximum fold change for both ptox and lignin accumulation than other three lines, though fold change of ptox accumulation only 1.04 times higher than lignin up accumulation. In contrast fold change of ptox accumulation was 5.5 times higher than lignin accumulation in the case of PhCAD3 followed by PhCAD4. PhCAD2 overexpressing line showed very less fold change for both the ptox and lignin accumulation (Fig. 14a). Comparative chromatogram for ptox accumulation of each of the overexpressing lines is represented in Fig. 14b. To validate our observation further, we raised four transgenic lines of *P. hexandrum* callus and *N. tabacum* plantlets (Supplemental Fig. S11A). GUS assay of *P. hexandrum* callus and *N. tabacum* was done to confirm the transformation efficiency (Supplemental Fig. S11B). The fold change ratio of lignin and ptox in four transgenic calli support our previous observation of cell culture

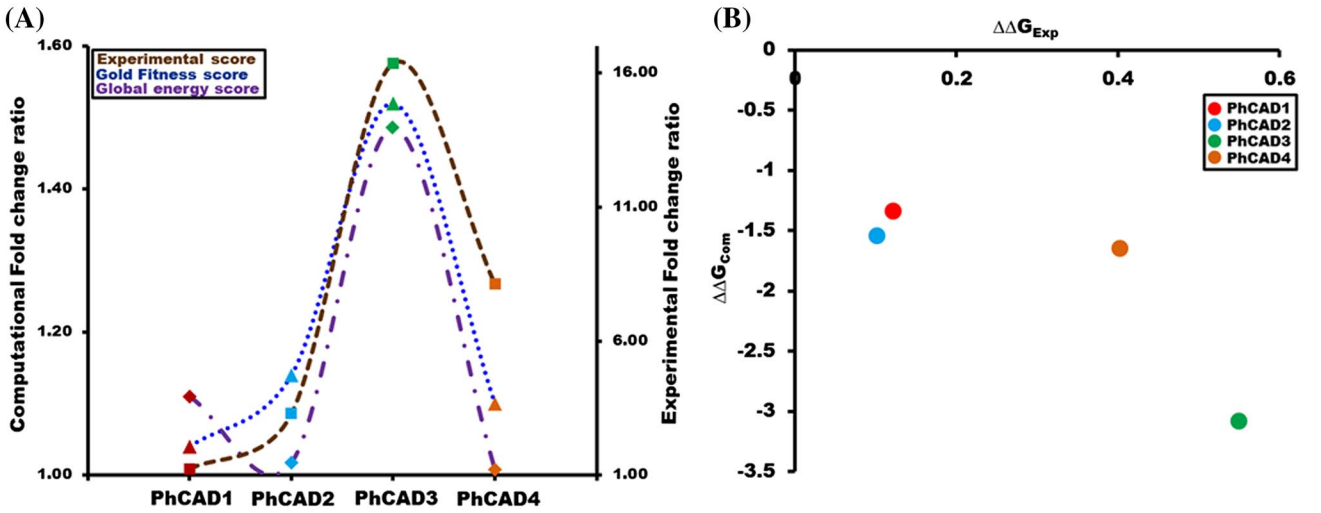

Fig. 8 Experimental and computational fold change analysis of *P. hexandrum* CAD enzymes. To understand the possible modes of ligand/substrate-binding, PhCAD1, PhCAD2, PhCAD3 and PhCAD4 are subjected to dock with CAld and sinapaldehyde using the GOLD v5.2 docking package. Each docking solution is considered for rescoring using the FireDock docking software. **a** The comparison between computational fold change of the Gold fitness score and the global energy score of CAld over sinapaldehyde obtained from GOLD and FireDock program and the experimental fold change when the Kcat/Km of CAld is compared over sinapaldehyde. **b** The correlation between computational and experimental ΔΔG. ΔΔG is calculated when the $\Delta G_{sinapaldehyde}$ is subtracted from $\Delta G_{coniferaldehyde}$

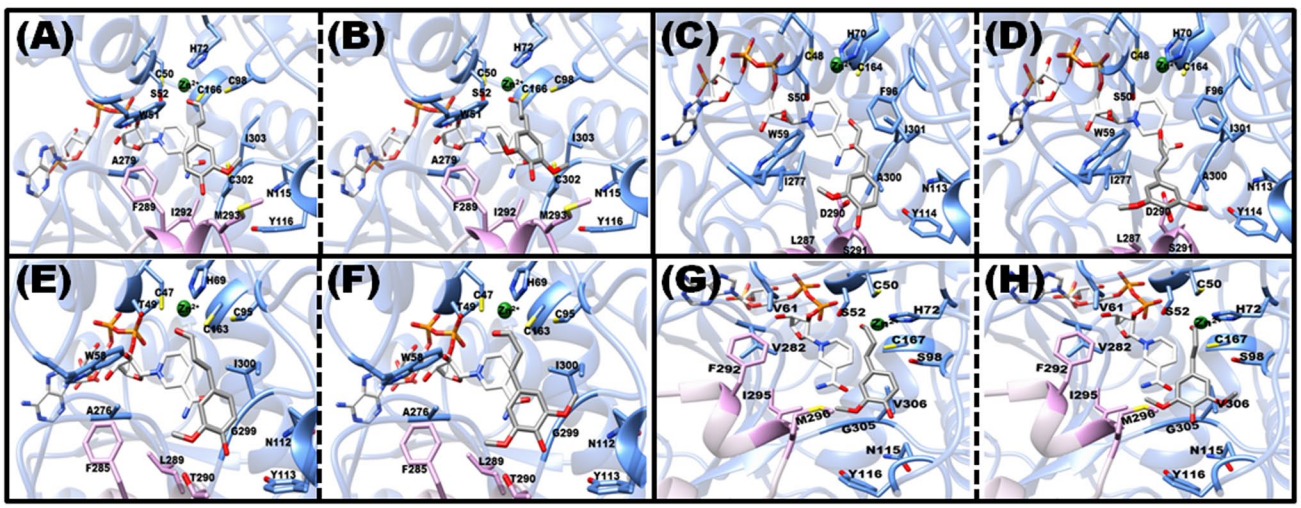

Fig. 9 Substrate docking in the active sites of *P. hexandrum* CADs enzyme. Cartoon representations of the representative docking pose of coniferaldehyde (**a**–**d**) and sinapaldehyde (**e**–**f**) onto the 3D model structures of PhCAD1, CAD2, CAD3, CAD4 proteins, respectively. The active site of each PhCAD is formed by residues from chain A (colored *blue*) and chain B (colored *pink*). The NADP molecule and docked substrate are shown in white and gray stick model, respectively

based experiments. Transgenic *N. tabacum* overexpressing PhCAD1 showed best up accumulation of lignin. In contrast, PhCAD3 overexpressing line showed less up accumulation followed by PhCAD4. Interestingly PhCAD2 did not show any significant up accumulation of lignin (Supplemental Fig. S11 E, F) Ptox and lignin content in control cell culture and callus have been represented (Supplemental Table S18). Mass spectra of control, MeJA treated and four transgenic *P. hexandrum* calli has been done (Supplemental Figure S12).

Discussion

In the present investigation, we report de novo assembly and an in-depth analysis of the MeJA treated transcriptome of 12 days old *P. hexandrum* cell cultures which exhibited enhanced accumulation of ptox. A comparison between present transcriptome data of MeJA treated cell culture and previously published transcriptome data of control cell culture from our lab (Bhattacharyya et al. 2013), revealed

Table 4 The kinetic parameters of the purified recombinant PhCAD1, PhCAD2, PhCAD3, PhCAD4 with CAld and sinapaldehyde as substrates

	CAld				Sinapaldehyde			
	Km (μM)	Vmax (μM s^{-1} mg^{-1})	Kcat (S^{-1})	Kenz Kact/Km (s^{-1} μM^{-1})	Km (μM)	Vmax (μM s^{-1} mg^{-1})	Kcat (S^{-1})	Kenz Kcat/Km (s^{-1} μM^{-1})
PhCAD1	36.9 ± 5.11	208.1 ± 7.55	7.9	0.21	47.92 ± 10.00	218.5 ± 13.28	8.3	0.170
PhCAD2	113.6 ± 13.23	214.6 ± 10.22	9.0	0.07	137.8 ± 24.9	80.9 ± 6.44	3.4	0.024
PhCAD3	67.6 ± 9.00	318.4 ± 14.15	12.0	0.18	208.2 ± 19.47	63.0 ± 3.02	2.4	0.011
PhCAD4	79.7 ± 12.11	286.7 ± 15.52	10.8	0.13	169.9 ± 22.21	72.6 ± 4.52	2.76	0.016

Values represent means of three independent replicates ± standard deviations

that ptox biosynthetic pathway gene like *PAL* was minutely affected by the treatment, whereas *CCR, HCT, 4CL, CAD, cytochrome P450* and *SAM-dependent methyltransferase* were significantly upregulated in treated condition. As CAD was most significantly upregulated, our further studies were performed by exploring the nature of different CAD isoforms from *P. hexandrum*.

Controlled transcription of biosynthetic pathway genes is an important mechanism regulating secondary metabolite production in plant cells (Endt et al. 2002). The known TFs involved in regulation of secondary metabolism are the basic helix-loop-helix (bHLH) proteins like CrMYC2, AP2/ERF family proteins, R2R3-MYB, WRKY, NAC, DOF, HD-ZIP, and TFIIIA zinc finger TFs. A recent report has shown that MYC2a and MYC2b TFs in tobacco are involved in the regulation of JA inducible nicotine biosynthesis (Zhang et al. 2012). JA-responsive AP2/ERF TFs, AaERF1, and AaERF2 have been shown to positively regulate artemisinin biosynthesis (Yu et al. 2012). TGA TFs are important regulators of JA signaling (Stotz et al. 2013). In our study, comparison of treated and control transcriptomes revealed that *AP2-EREBP, NAC, bHLH, MYB, mTERF* and *WRKY* were significantly up-regulated, whereas *C2C2* was slightly down-regulated in treated cultures.

The de novo assembly was validated by semi RT-qPCR analysis. Results revealed that the enhanced accumulation of ptox at 12th day MeJA treated cell cultures corresponded to an upregulation of phenylpropanoid pathway genes, viz. a marked upregulation of the *Podophyllum* CADs, similar to *Arabidopsis* CAD isoform CAD8 (Brill et al. 1999), whereas, downregulation of true lignifying CADs, viz. CAD4 and CAD5 is a significant point to be noted (Sibout et al. 2003, 2005; Kim et al. 2004; Eudes et al. 2006). This transcriptome data provides a crucial starting point for an in-depth analysis of this economically important herb and will enhance the progress of gene discovery, characterization of phenylpropanoid pathway as a whole and facilitate future whole-genome sequence assembly and annotation of other related non-model herbs.

Nine putative *CAD* genes were identified from *the Arabidopsis genome database* in previous studies (Raes et al. 2003; Goujon et al. 2003; Sibout et al. 2003; Kim et al. 2004). True lignifying CADs are those which have been implicated in lignification of the vascular tissue and includes *Arabidopsis* CAD5, CAD4 and CAD1 (Sibout et al. 2003, 2005; Kim et al. 2004; Eudes et al. 2006). AtCAD7 and AtCAD8, being very divergent from "true" CADs, represents a group of multi-substrate alcohol dehydrogenases (Brill et al. 1999). Expression levels measured using Affymetrix's GeneChip microarray technology (Schmid et al. 2005) noted AtCAD7 transcripts to be mainly found in leaves and flowers, whereas AtCAD8 was only detected in the flowers. AtCAD2, 3, 7 and 8 have been shown to possess

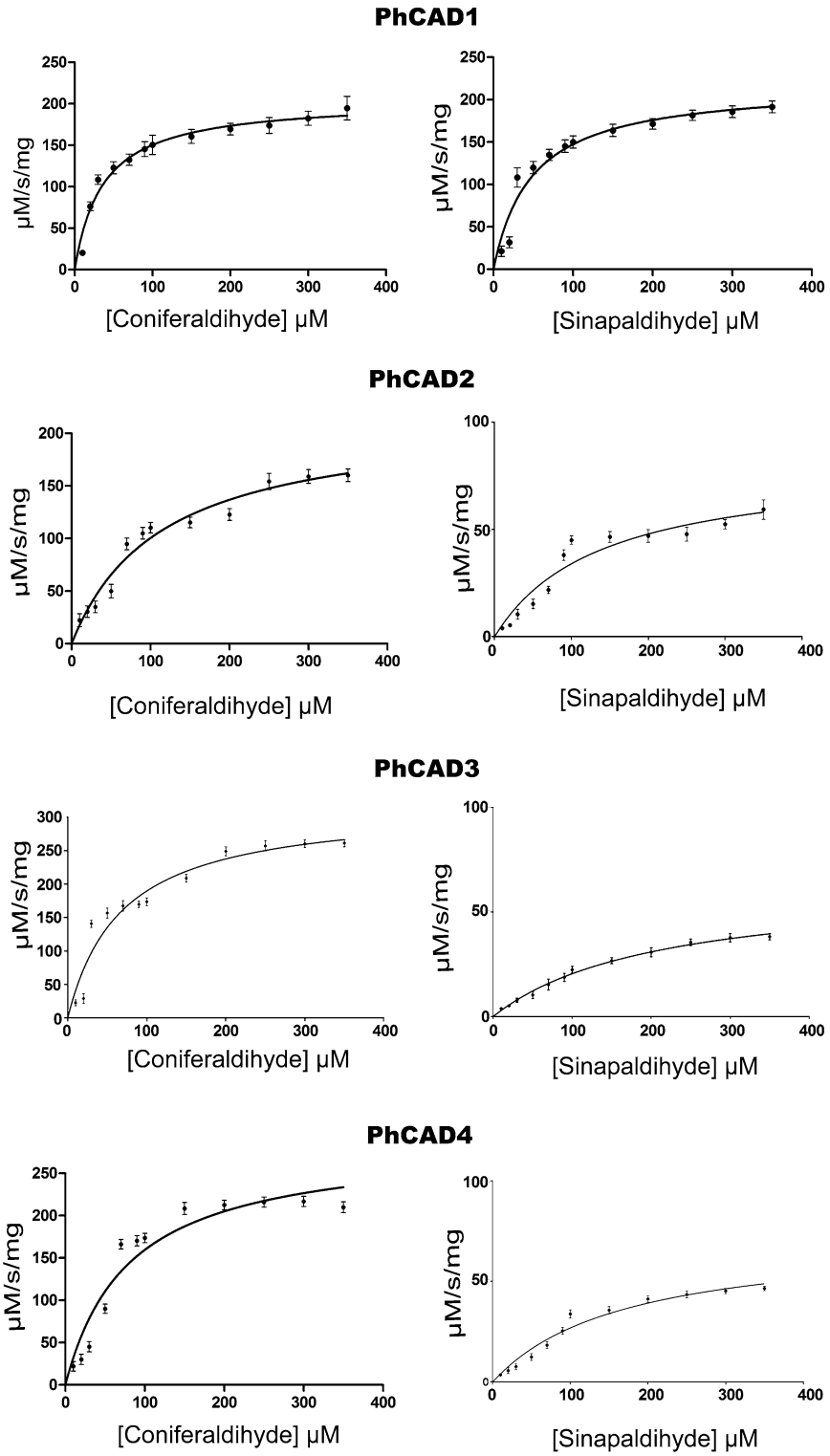

Fig. 10 The enzyme kinetics plots of initial reaction velocity (*V*) versus substrate concentration (*S*) of four recombinant isoforms namely PhCAD1, PhCAD2, PhCAD3, PhCAD4 for the substrates CAld and sinapaldehyde. Data are the mean ± SD of three independent experiments

much lower levels of CAD activity, with AtCAD1, 6 and 9 being inactive (Kim et al. 2004).

In earlier studies expression of nine CAD isoforms from *A. thaliana*, namely AtCAD1–9 was studied in different developmental stages and it was found that some AtCADs like AtCAD 2 and 3 were expressed in non-lignifying tissue. Whereas AtCAD7 and AtCAD8 displayed gene expression patterns largely resembling those of AtCAD4/5, which perhaps indicates a quite minor role in monolignol/lignin formation and may have some other function On the other hand, AtCAD1 and AtCAD9 are truly involved in lignifications (Kim et al. 2007). So in this study to

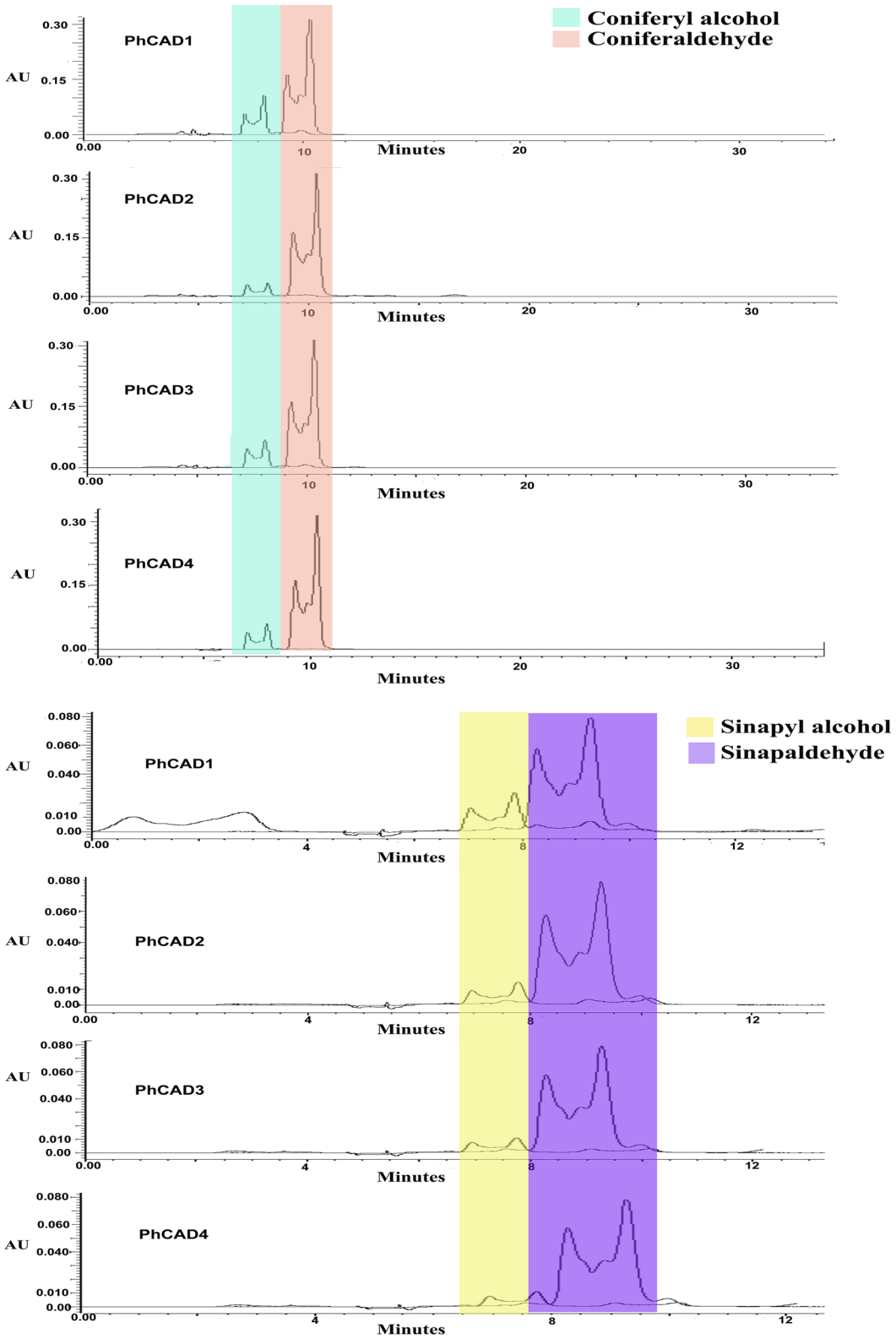

Fig. 11 HPLC elution profiles of enzyme assay products catalyzed by four recombinant isoforms, PhCAD1, PhCAD2, PhCAD3, PhCAD4 for the substrates of coniferaldehyde and sinapaldehyde respectively

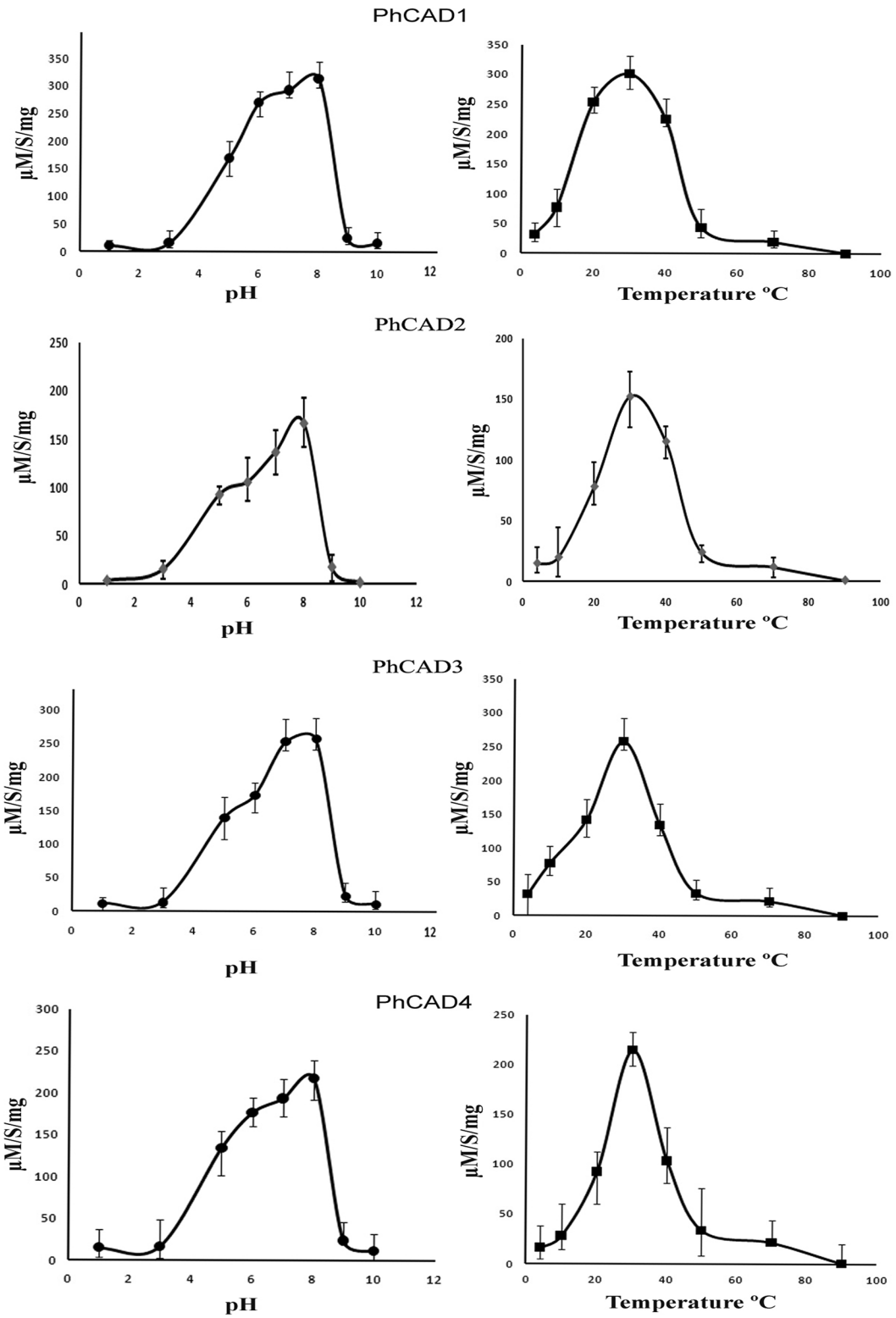

Fig. 12 pH and temperature dependent activity of four recombinant isoforms namely PhCAD1, PhCAD2, PhCAD3, PhCAD4. Data are the mean ± SD of three independent experiments

investigate how PhCAD isoforms favour lignin and lignan production, we isolated four isoforms of PhCAD namely PhCAD1, PhCAD2, PhCAD3and PhCAD4. Among these isoforms, PhCAD1, PhCAD2 and PhCAD3 were very closely related to each other, sharing a similarity of 74% and above whereas PhCAD4 was distantly related to these

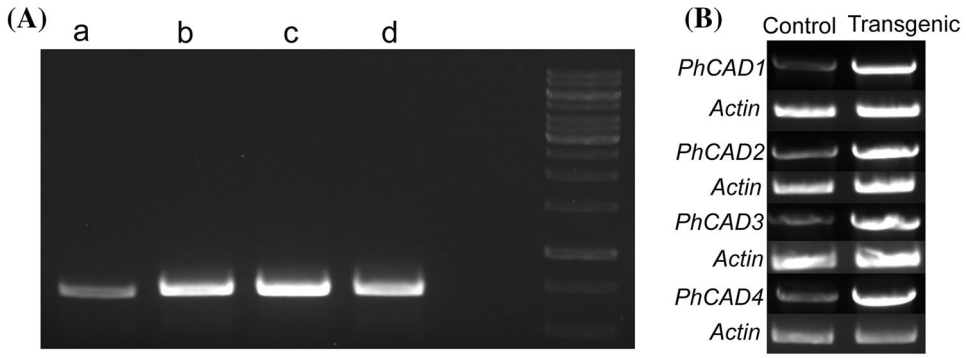

Fig. 13 **a** PCR analysis using *npt*II forward and reverse primer from genomic DNA of four transgenic lines over-expressing four isoforms a. *PhCAD1*, b. *PhCAD2*, c. *PhCAD3*, d. *PhCAD4* to confirm the transgenic line. **b** Semi RT-qPCR of four isoforms *PhCAD1, PhCAD2, PhCAD3, PhCAD4* from four transgenic lines overexpressing *PhCAD1, PhCAD2, PhCAD3, PhCAD4* compared to control (wild type) to confirm their overexpression

three isoforms sharing 69 % and less similarity between them (Supplementary Fig. S9). Phylogenetic analysis of PhCADs with other related CADs revealed that PhCAD1, PhCAD2, and PhCAD3 made a cluster, which is different from other related CADs. However, PhCAD4 is distantly related to these three PhCADs and clustered with *Medicago truncatula*.

We observed reasonable structural variations among the 3D models of PhCADs in terms of stability and cavity size despite the presence of higher sequence similarities among them. We further investigated the differential binding modes and/or affinities for substrates (coniferyl and sinapyl aldehydes) among the PhCAD isoforms via rigorous molecular docking analyses. Both computational and experimentally derived binding scores suggest preferential binding of coniferaldehyde over sinapaldehyde for PhCAD3 than other PhCADs as illustrated by the correlation study between the experimental and computationally derived ΔΔG of binding. Previous studies have emphasized the importance of these residues in interaction with respective ligands of SAD (Bomati and Noel 2005) and our docking solutions generally fit well with these previous findings. Earlier study (Bomati and Noel 2005) suggested that the correct orientation of the aldehyde substrate would be facilitated by Zn^{2+} coordination to the aldehyde carbonyl oxygen and hydrogen bond formation with the S52 hydroxyl side chain. Docking solutions with CAld and sinapaldehyde match well with this hypothesis forming suitable Zn^{2+} coordination and hydrogen bond with Serine in PhCAD1, PhCAD2 and PhCAD3 and threonine in PhCAD4. Hence, the substrates are suitably tethered at the aldehyde head, via serine/threonine hydrogen bonding and Zn^{2+} coordination, whereas at the phenolic tail, via hydrophobic interactions and probable hydrogen bonding with the methionine in PhCAD1 and PhCAD4, serine in PhCAD2, threonine in PhCAD3.

Interestingly, in vitro enzyme assay showed that PhCAD1 has a very high and almost equal affinity towards two substrates CAld and sinapaldehyde. In contrast, PhCAD3 showed a moderate affinity towards CAld, but very less affinity towards sinapaldehyde which was followed by PhCAD4 and PhCAD2. As a result, the ratio of Kenz for CAld:kenz for sinapaldehyde (Kenz denotes the affinity towards substrate) was found to be highest for PhCAD3 followed by PhCAD4, PhCAD2, and PhCAD1. However, PhCAD2 showed very less affinity towards both of the substrates. This observation was completely supported by in silico docking study (Fig. 8). To demonstrate this observation in vivo, we generated four transgenic cell lines overexpressing four isoforms and observed the ratio of up accumulation of ptox and lignin for each of the lines compared to control (wild type). Because CAD produces coniferyl alcohol from coniferadehyde and sinapyl alcohol from sinapaldehyde, which are responsible for both lignin and lignan and only lignin production respectively. This overexpression study revealed that PhCAD3 showed the highest ratio of up accumulation of ptox:lignin which is followed by PhCAD4, PhCAD2, and PhCAD1. However, PhCAD2 was noted with comparatively less accumulation for both ptox and lignin. To study the functions of these PhCADs in another plant we made four transgenic lines of *N. tabacum* and it revealed that PhCAD1 favored the lignin formation more whereas PhCAD3 showed very little up accumulation of lignin followed by PhCAD4. PhCAD2 did not show any significant function in lignin formation. In vitro, in silico, and in vivo studies revealed that *PhCAD1* favors ptox and lignin production almost equally, in contrast, *PhCAD3* favors ptox production more specifically than lignin which followed by *PhCAD4* and *PhCAD2*.

In conclusion, our transcriptome data revealed how MeJA affects various pathway genes and transcription factors related to ptox biosynthesis. Furthermore, a comparative

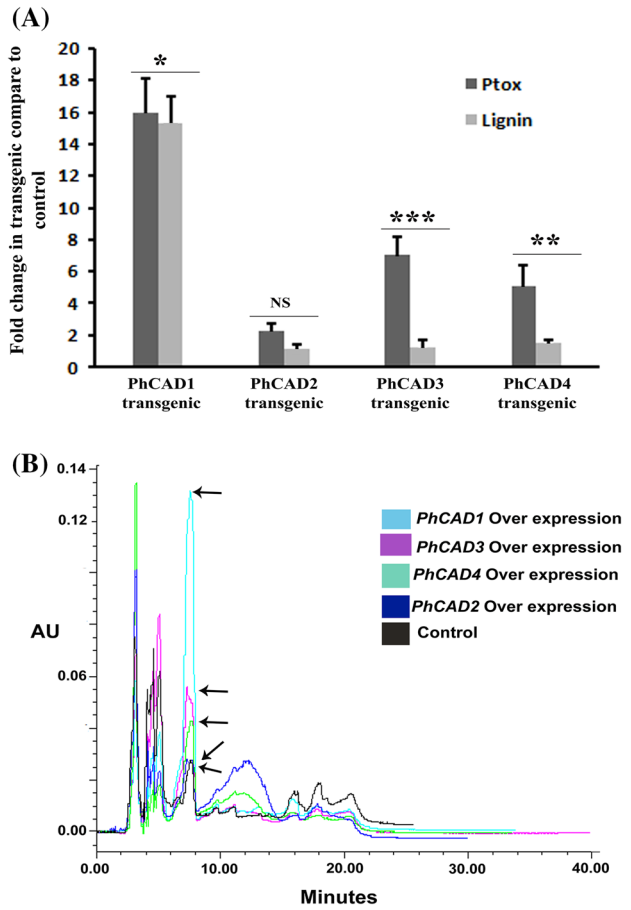

Fig. 14 **a** Fold change to check the accumulation of ptox and lignin in four transgenic cell lines overexpressing four isoforms compared to wild type cell culture. Data are the mean ± SD of three independent experiments. The statistical analysis of the difference between fold changes was done using GraphPad Instat-3 software. Comparison between groups was done using one-way analysis of variance (ANOVA) followed by Student–Newman–Keuls test. Data were fitted using Sigma plot represented as means ± SD. $p<0.05$ was accepted as level of significance; ***highly significant $p<0.001$; **significant $p<0.01$; *less significant $p<0.05$; NS not significant for $p>0.05$. **b** HPLC profile of accumulation of ptox from four transgenic lines over-expressing four isoforms *PhCAD1, PhCAD2, PhCAD3, PhCAD4* including control (wild type). *Arrows* indicate the peak of ptox in HPLC chromatograph

study of our previously reported control and present MeJA treated transcriptome datasets showed a significant up-regulation of CAD, a regulating enzyme of the lignan/lignin pathway, over other phenylpropanoid/ptox pathway genes. The present study also characterized ptox specific CAD isoforms from *P. hexandrum* which represents a valuable resource for future investigation on this high-value medicinal herb and also provides new insights into the ptox biosynthetic pathway.

Acknowledgments Authors thank Dr. P. S. Ahuja, Ex-Director, CSIR-IHBT, Palampur, & Ex-Director-General, CSIR, New Delhi, for the generous gift of fresh *P. hexandrum*. We are grateful to Prof. S. Roy, Ex-Director, CSIR-IICB for his constant guidance and encouragement. The authors would like to express their gratitude to the Director, CSIR-IICB, Kolkata. This work received financial support from the Council of Scientific and Industrial Research (CSIR), New Delhi. DB, AB and DK acknowledge the CSIR, SH to University Grants Commission, RD to Indian Council of Medical Research for their fellowships. SC[b] acknowledged the Department of Biotechnology, India for the Ramalingaswami Fellowship and CSIR network project GENESIS (code: BSC0121) for financial support.

Author contributions Dipto Bhattacharyya and Saptarshi Hazra carried out the experimental work, analyzed the data and framed out the manuscript. Anindyajit Banerjee and Saikat Chakrabarti carried out the docking and related studies. Riddhi Datta developed the transgenic lines. Deepak Kumar helped the seniors to analyze the data. Saikat Chakrabarti helps to draft the manuscript. Sharmila Chattopadhyay designed the experiments and prepared the final manuscript.

Compliance with ethical standards

Conflicts of interest The authors declare that they have no competing interests.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Bhattacharyya D, Sinha R, Ghanta S, Chakraborty A, Hazra S, Chattopadhyay S (2012) Proteins differentially expressed in elicited cell suspension culture of *Podophyllum hexandrum* with enhanced podophyllotoxin content. Proteome Sci 10:34

Bhattacharyya D, Sinha R, Hazra S, Datta R, Chattopadhyay S (2013) De novo transcriptome analysis using 454 pyrosequencing of the Himalayan Mayapple, *Podophyllum hexandrum*. BMC Genom 14:748

Bolton E, Wang Y, Thiessen PA, Bryant SH (2008) Integrated platform of small molecules and biological activities. In: Wheeler RA, Spellmeyer DC (eds) Annual reports in computational chemistry. Elsevier, Oxford, pp 217–241

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

Brill EM, Abrahams S, Hayes CM, Jenkins CL, Watson JM (1999) Molecular characterisation and expression of a wound-inducible cDNA encoding a novel cinnamylalcohol dehydrogenase enzyme in lucerne (*Medicago sativa* L.) Plant Mol Biol 41:279–291

Bukh C, Nord-Larsen PH, Rasmussen SK (2012) Phylogeny and structure of the cinnamyl alcohol dehydrogenase gene family in *Brachypodium distachyon*. J Exp Bot 63:6223–6236

Chakraborty A, Bhattacharya D, Ghanta S, Chattopadhyay S (2010) An efficient protocol for in vitro regeneration of *Podophyllum hexandrum*, a critically endangered medicinal plant. Ind. J Biotechnol 9:217–220

Chang XF, Chandra R, Berleth T, Beatson RP (2008) Rapid, microscale, acetyl bromide-based method for high-throughput determination of lignin content in *Arabidopsis thaliana*. J Agric Food Chem 56:6825–6834

Chao N, Liu SX, Liu BM, Li N, Jiang XN, Gai Y (2014) Molecular cloning and functional analysis of nine cinnamyl alcohol

dehydrogenase family members in *Populus tomentosa*. Planta 240:1097–1112

Chattopadhyay S, Mehra RS, Srivastava AK, Bhojwani SS, Bisaria VS (2003) Effect of major nutrients on podophyllotoxin production in *Podophyllum hexandrum* suspension cultures. Appl Microbiol Biotechnol 60:541–546

Conesa A, Gotz S, Garcia Gomez JM, Terol J, Talon M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674–3676

Datta R, Kumar D, Sultana A, Hazra S, Bhattacharyya D, Chattopadhyay S (2015) Glutathione regulates 1-aminocyclopropane-1-carboxylate synthase transcription via WRKY33 and 1-aminocyclopropane-1-carboxylate oxidase by modulating messenger RNA stability to induce ethylene synthesis during stress. Plant Physiol 169:2963–2981

Dence CW (1992) The determination of lignin. In: Lin SY, Dence CW (eds) Methods in lignin chemistry. Springer, Berlin, pp 33–61

Deng WW, Zhang M, Wu JQ, Jiang ZZ, Tang L, Li YY, Wei CL, Jiang CJ, Wan XC (2013) 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

Dundas J, Ouyang Z, Tseng J, Binkowski A, Turpaz Y, Liang J (2006) CASTp: computed atlas of surface topography ofproteins with structural and topographical mapping of functionally annotated residues. Nucleic Acids Res 34:116–118

Eisenberg D, Lüthy R, Bowie JU (1997) VERIFY3D: assessment of protein models with three-dimensional profiles. Methods Enzymol 277:396–404

Endt DV, Kijne JW, Memelink J (2002) Transcription factors controlling plant secondary metabolism: what regulates the regulators? Phytochemistry 61:107–114

Eswar N, Eramian D, Webb B, Shen MY, Sali A (2008) Protein structure modeling with MODELLER. Methods Mol Biol 426:145–159

Eudes B, Pollet A, Sibout R, Do CT, Séguin A, Lapierre C, Jouanin L (2006) Evidence for a role of AtCAD1 in lignification of elongating stems of *Arabidopsis thaliana*. Planta 225:23–39

Fuss E (2003) Lignan in plant cell and organ cultures: an overview. Phytochem Rev 2:307–320

Ghanta S, Bhattacharyya D, Sinha R, Banerjee A, Chattopadhyay S (2011) *Nicotiana tabacum* overexpressing γ-ECS exhibits biotic stress tolerance likely through NPR1-dependent salicylic acid-mediated pathway. Planta 233:895–910

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

Gross G, Sto¨ckigt J, Mansell R, Zenk M (1973) Three novel enzymes involved in the reduction of ferulic acid to coniferyl alcohol in higher plants: ferulate-Co a ligase, feruloyl-Co a reductase and coniferyl alcohol oxidoreductase. FEBS Lett 31:283–286

Gundlach H, Muller MJ, Kutchan TM, Zenk MH (1992) Jasmonic acid is a signal transducer in elicitor-induced plant cell cultures. Proc Natl Acad Sci USA 89:2389–2393

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 bonafide CAD. J Mol Evol 71:202–218

Hano C, Addi M, Bensaddek L, Crônier D et al (2006) Differential accumulation of monolignol-derived compounds in elicited flax (*Linum usitatissimum*) cell suspension cultures. Planta 223:975–989

Heyenga AG, Lucas JA, Dewick PM (1990) Production of tumour inhibitory lignans in callus cultures of *Podophyllum hexandrum*. Plant Cell Rep 9:382–385

Hildebrand A, Remmert M, Biegert A, Soding J (2009) Fast and accurate automatic structure prediction with HHpred. Proteins 77:128–132

Jones G, Willett P, Glen RC, Leach ARL, Taylor R (1997) Development and validation of a genetic algorithm for flexible docking. J Mol Biol 267:727–748

Kadkade PG (1981) Formation of podophyllotoxins by *Podophyllum peltatum* tissue cultures. Naturwissenschaften 68:481–482

Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M (2004) The KEGG resource for deciphering the genome. Nucleic Acids Res 32:277–280

Kartal M, Konuklugil B, Indrayanto G, Alfermann AW (2004) Comparison of different extraction methods for the determination of podophyllotoxin and 6-methoxypodophyllotoxin in *Linum* species. J Pharmaceut Biomed Anal 35:441–447

Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30:3059–3066

Katoh K, Kuma K, Miyata T, Toh H (2005) Improvement in the accuracy of multiple sequence alignment program MAFFT. Genome Inform 16:22–33

Kim SJ, Kim MR, Bedgar DL, Moinuddin SG, Cardenas CL, Davin LB, Kang C, Lewis NG (2004) Functional reclassification of the putative cinnamyl alcohol dehydrogenase multigene family in *Arabidopsis*. Proc Natl Acad Sci USA 101:1455–1460

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 HJ, Ono E, Morimoto K, Yamagaki T, Okazawa A, Kobayashi A, Satake H (2009) Metabolic engineering of lignan biosynthesis in *Forsythia* cell culture. Plant Cell Physiol 50:2200–2209

Konagurthu AS, Whisstock JC, Stuckey PJ, Lesk AM (2006) MUSTANG: a multiple structural alignment algorithm. Proteins 64:559–574

Kyndt T, Denil S, Haegeman A, Trooskens G, De MT, Van CW, Gheysen G (2012) Transcriptome analysis of rice mature root tissue and root tips in early development by massive parallel sequencing. J Exp Bot 63:2141–2157

Laskowski RA, Mac AMW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 26:283–291

Lau W, Sattely ES (2015) Six enzymes from mayapple that complete the biosynthetic pathway to the etoposideaglycone. Science 349:1124–1128

Lerndal T, Svensson B (2000) A clinical study of CPH82 vs methotrexate in early rheumatoid arthritis. Rheumatology 39:316–320

Lewis NG, Davin LB (1999) Lignans: biosynthesis and function. In: Barton DHR, Nakanishi K, Meth-Cohn O (eds) Comprehensive natural products chemistry, Elsevier, London, pp 639–712

Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22:1658–1659

Li XYY, Yao J, Chen G, Zhang Q, Wu C (2009) FLEXIBLE CULM 1 encoding a cinnamyl alcohol dehydrogenase controls culm mechanical strength in rice. Plant Mol Biol 69:685–697

Liang J, Edelsbrunner H, Woodward C (1998) Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. Protein Sci 7:1884–1897

Ma QH (2010) Functional analysis of a cinnamyl alcohol dehydrogenase involved in lignin biosynthesis in wheat. J Exp Bot 61:2735–2744

Marques JV, Kim KW, Lee C, Costa MA, May GD, Crow JA, Davin LB, Lewis NG (2013) Next generation sequencing in predicting gene function in podophyllotoxin biosynthesis. J Biol Chem 288:466–479

Mashiach E, Schneidman DD, Andrusier N, Nussinov R, Wolfson HJ (2008) FireDock: a web server for fast interaction refinement in molecular docking. Nucleic Acids Res 36:229–232

Murashige T, Skoog F (1962) A revised medium for rapid growth and bioassays with tobacco tissue cultures. Plant Physiol 15:473–497

Pauwels L, Morreel K, Witte D, Lammertyn E, Van Montagu F, Boerjan M, Inzé W, Goossens D (2008) Mapping methyl jasmonate-mediated transcriptional reprogramming of metabolism and cell cycle progression in cultured *Arabidopsis* cells. Proc Natl Acad Sci USA 105:1380–1385

Pruitt KD, Tatusova T, Maglott DR (2006) NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res 35:61–65

Raes J, Rohde A, Christensen JH, Van de PY, Boerjan W (2003) Genome-wide characterization of the lignification toolbox in *Arabidopsis*. Plant Physiol 133:1051–1071

Santos WD, Ferrarese ML, Ferrarese FO (2006) High performance liquid chromatography method for the determination of cinnamyl alcohol dehydrogenase activity in soybean roots. Plant Physiol Biochem 44:511–515

Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Schölkopf B, Weigel D, Lohmann JU (2005) A gene expression map of *Arabidopsis thaliana* development. Nat Genet 37:501–506

Schmieder R, Edwards R (2011) Quality control and preprocessing of metagenomic datasets. Bioinformatics 27:863–864

Seidel V, Windhövel J, Eaton G, Alfermann AW, Arroo RR, Medarde M, Petersen M, Woolley JG (2002) Biosynthesis of podophyllotoxin in *Linum album* cell cultures. Planta 215:1031–1039

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

Sibout R, Eudes A, Mouille G, Pollet B, Lapierre C, Jouanin L, Séguin A (2005) Routing sinapyl and coniferyl alcohol pathway though the last reduction step in *Arabidopsis thaliana* using a double CAD mutant. Plant Cell 17:2059–2076

Smollny T, Wichers H, Kalenberg S, Shahsavari A, Petersen M, Alfermann AW (1998) Accumulation of podophyllotoxin and related lignans in cell suspension cultures of *Linum album*. Phytochemistry 37:864–868

Stahelin HF, von Wartburg A (1991) The chemical and biological route from podophyllotoxin-glucoside etoposide. Cancer Res 51:5–15

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313

Stotz HU, Mueller S, Zoeller M, Mueller MJ, Berger S (2013) TGA transcription factors and jasmonate-independent COI1 signalling regulate specific plant responses to reactive oxylipins. J Exp Bot 64:963–975

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

Umezawa T (2003) Diversity in lignan biosynthesis. Phytochem Rev 2:371–390

van Uden W, Pras N, Visser JF, Malingré TM (1989) Detection and identification of Podophyllotoxin produced by cell cultures derived from *Podophyllum hexandrumroyle*. Plant Cell Rep 8:165–168

van Furden B, Humburg A, Fuss E (2005) Influence of methyl jasmonate on podophyllotoxin and 6-methoxypodophyllotoxin accumulation in *Linum album* cell suspension cultures. Plant Cell Rep 24:312–317

Wichers HJ, Versluis-De Haan GG, Marsman JW, Harkes MP (1991) Podophyllotoxins in plants and cell cultures of *Linum lavum*. Phytochemistry 30:3601–3604

Yousefzadi M, Sharifi M, Behmanesh M, Ghasempour A, Moyano E, Palazon J (2010) Salicylic acid improves podophyllotoxin production in cell cultures of *Linum album* by increasing the expression of genes related with its biosynthesis. Biotechnol Lett 32:1739–1743

Yu ZX, Li JX, Yang CQ, Hu WL, Wang LJ, Chen XY (2012) The jasmonate-responsive AP2/ERF transcription factors AaERF1 and AaERF2 positively regulate artemisinin biosynthesis in *Artemisia annua* L. Mol Plant 5:353–365

Zhang HB, Bokowieca MT, Rushtona PJ, Hana SC, Timkoa MP (2012) Tobacco transcription factors NtMYC2a and NtMYC2b form nuclear complexes with the NtJAZ1 repressor and regulate multiple jasmonate-inducible steps in nicotine biosynthesis. Mol Plant 5:73–84