Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

Available online at www.sciencedirect.com
ScienceDirect
Chinese Journal of Natural Medicines 2016, **14**(11): 0801−0812
Chinese Journal of Natural Medicines

·Research articles·

The phenylalanine ammonia-lyase gene family in *Isatis indigotica* Fort.: molecular cloning, characterization, and expression analysis

MA Rui-Fang [1], LIU Qian-Zi [2], XIAO Ying [3], ZHANG Lei [4], LI Qing [2], YIN Jun [1*], CHEN Wan-Sheng [3*]

[1] *Development and Utilization Key Laboratory of Northeast Plant Materials, School of Traditional Chinese Materia Medica, Shenyang Pharmaceutical University, Shenyang 110016, China;*
[2] *College of Pharmacy, Jiamusi University, Jiamusi 154007, China;*
[3] *Department of Pharmacy, Changzheng Hospital, Second Military Medical University, Shanghai 200003, China;*
[4] *Department of Pharmaceutical Botany, School of Pharmacy, Second Military Medical University, Shanghai 200433, China*

Available online 20 Nov., 2016

[ABSTRACT] Phenolic compounds, metabolites of the phenylpropanoid pathway, play an important role in the growth and environmental adaptation of many plants. Phenylalanine ammonia-lyase (PAL) is the first key enzyme of the phenylpropanoid pathway. The present study was designed to investigate whether there is a multi-gene family in *I. Indigotic* and, if so, to characterize their properties. We conducted a comprehensive survey on the transcription profiling database by using tBLASTn analysis. Several bioinformatics methods were employed to perform the prediction of composition and physicochemical characters. The expression levels of *IiPAL* genes in various tissues of *I. indigotica* with stress treatment were examined by quantitative real-time PCR. Protoplast transient transformation was used to observe the locations of *IiPAL*s. *IiPAL*s were functionally characterized by expression with pET-32a vector in *Escherichia colis* strain BL21 (DE3). Integration of transcripts and metabolite accumulations was used to reveal the relation between *IiPAL*s and target compounds. An new gene (*IiPAL2*) was identified and both *IiPAL*s had the conserved enzymatic active site Ala-Ser-Gly and were classified as members of dicotyledon. *IiPAL1* and *IiPAL2* were expressed in roots, stems, leaves, and flowers, with the highest expression levels of *IiPAL1* and *IiPAL2* being observed in stems and roots, respectively. The two genes responded to the exogenous elicitor in different manners. Subcellular localization experiment showed that both *IiPAL*s were localized in the cytosol. The recombinant proteins were shown to catalyze the conversion of L-Phe to *trans*-cinnamic acid. Correlation analysis indicated that *IiPAL1* was more close to the biosynthesis of secondary metabolites than *IiPAL2*. In conclusion, the present study provides a basis for the elucidation of the role of *IiPAL*s genes in the biosynthesis of phenolic compounds, which will help further metabolic engineering to improve the accumulation of bioactive components in *I. indigotica*.

KEY WORDS] Phenylalanine ammonia-lyase; *Isatis indigotica* Fort.; Phenolic compounds; Lignan; Correlation analysis

CLC Number] R282.5 **[Document code]** A **[Article ID]** 2095-6975(2016)11-0801-12

Introduction

Isatis indigotica Fort. (*I. intinctoria*) is a biennial herbaceous plant, belonging to the family of Brassicaceae [1]. Its dried roots (Banlangen, *Isatis* root) and leaves (Daqingye, *Isatis* leaf) show notable anti-inflammatory, antibacterial, and antiviral activities [2]. Additionally, Banlangen has been demonstrated to have potentials to treat SARS (severe acute respiratory syndromes) [3] and H1N1-influenza [4].

In previous experiments, lariciresinol and larch lignan glycosides isolated from *Isatis indigotica* have been proven to possess a number of biological activities, such as anti-influenza A1 virus [5], anti-inflammation [6], and anti-fungal effects [7] as well as reducing the risk of cardiovascular

[Received on] 25-Feb.-2016
[Research funding] This work was supported by the Natural Science Foundation of China (Nos. 31100221 and 81325024).
[*Corresponding authors] Tel: 86-24-23986491, E-mail: yinjun2002@yahoo.com (YIN Jun); Tel: 86-21-81886181, Fax: 86-21-33100038, E-mail: chenwanshengsmmu@aliyun.com (CHEN Wan-Sheng).
These authors have no conflict of interest to declare.
Published by Elsevier B.V. All rights reserved

diseases [8]. However, the contents of lariciresinol and larch lignan glycosides in the roots are very low, only 47.14 and 84.67 μg·g^{-1}, respectively [5]. Knowledge of the biosynthetic enzymes and their corresponding genes would enable a much higher production of the valuables in engineered plant or microbial cells [9].

As one of guaiacy lignins, lariciresinol is derived from phenylpropanoid pathway with many enzymes involved [10] (Fig. 1). As the first key enzyme in the phenylpropanoid biosynthesis, PAL links primary and secondary metabolism by catalyzing the conversion of L-phenylalanine to cinnamic acid and is also a rate-limiting step of the phenylpropanoid metabolism [11]. Since the first *PAL* was discovered from barley by Koukol and Conn in 1961 [12], more and more PAL genes have been cloned in many higher plants, such as *Salvia Miltiorrhiza* [13], *Dendrobium* [14], *Salix viminalis* [14], and *Picrorhiza kurrooa* [16], and it also have been found in some liverworts [17] and fungi [18].

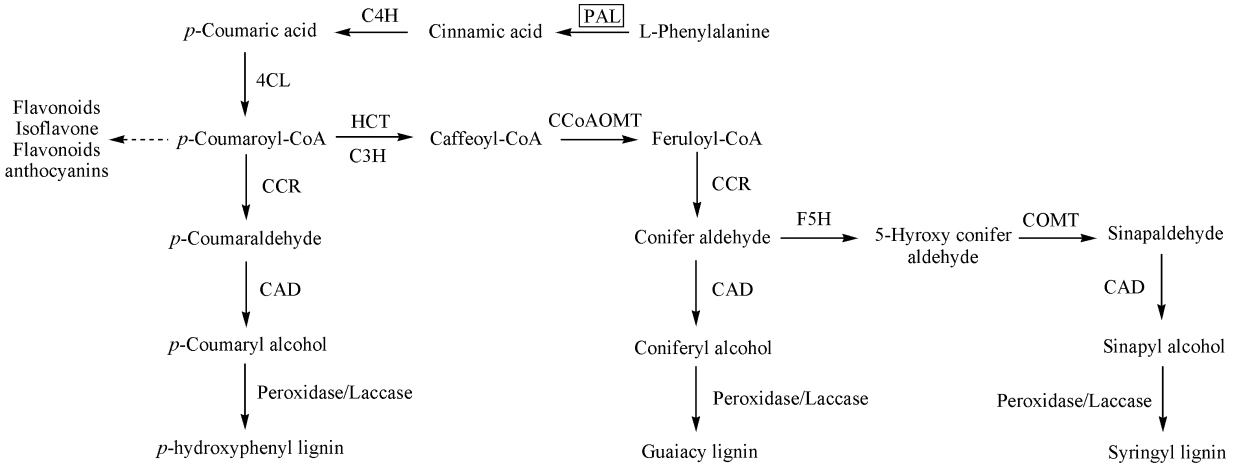

Fig. 1 Lignin biosynthetic pathway and the involved genes. *PAL* catalyzes the first step in the conversion of L-phenylalanine (L-Phe) to *trans-cinnamic* acid. Abbreviations: C4H, cinnamate 4-hydroxylase; 4CL, 4-coumarate: CoA ligase; HCT, *p*-hydroxycinnamoyl-CoA: quinate shikimate *p*-hydroxycinnamoyltransferase; C3H, *p*-coumarate 3-hydroxylase; CCoAOMT, caffeoyl-CoA *O*-methyltransferase; CCR, cinnamoyl-CoA reductase; F5H, ferulate 5-hydroxylase; COMT, caffeic acid *O*-methyltransferase and CAD, cinnamyl alcohol dehydrogenase. The solid line represents one step reaction; the dotted line represents a multi-step reaction

In a number of plants, the PAL proteins are encoded by a multi-gene family. The number of PAL genes is three in *S. Miltiorrhiza* [19] and *Coffea canephora* [20], four in *S. viminalis* [15], five in *Populus trichocarpa* [21], six in *Oryza sativa* [22], and twelve in *Citrullus lanatus* [23]. The individual PAL may respond differentially to biotic or abiotic stress, and its different expressions in tissues may be involved in the production of different products under specific conditions. In *C. canephora* [20], *CcPAL2* transcripts appear predominantly in flower, fruit pericarp, roots, and branches, whereas *CcPAL1* and *CcPAL3* are highly expressed in immature fruits. What's more, *CcPAL1* and *CcPAL3* are associated with the accumulation of cholorogenic acids (CGA), whereas *CcPAL2* may contribute more significantly to flavonoid accumulation. In *S. Miltiorrhiza* [19], all three *SmPAL*s are regulated by drought and MeJA treatments, although the time and degree of reactions differ one from another.

In a previous study, we have cloned a new plant PAL gene (designated as *IiPAL1*) from *I. indigotica* [24]. The open reading frame (ORF) of *IiPAL1* is 2178-bp and it encodes a polypeptide of 725 amino acid residues. *IiPAL1* is constitutively expressed in roots, stems, and leaves, with the highest expression being found in stems, and it responds to gibberellin (GA3), abscisic acid (ABA), methyl jasmonate (MeJA), and cold treatments. However, the characteristics of *IiPAL1* need to be further investigated and a systematic analysis of different PAL genes in *I. indigotica* also is needed. Moreover, the relations between *IiPAL*s and synthesis of secondary metabolites remain to be explored.

Under the umbrella of a transcription profiling of *I. indigotica* [25], one additional gene (*IiPAL2*) was identified in the present study. This was the first time to report the existence of a small multi-gene family in *I. indigotica.* The relations between *IiPAL*s and secondary metabolites were also evaluated in the present study. The results from the present study would enable us to further understand the role of *IiPAL*s in the synthesis of phenylpropanoid compounds in *I. indigotica* at the molecular level, which might be helpful to overcome the low rate of production of secondary metabolites, such as lariciersinol, in the future.

Materials and Methods

Plant materials

The seeds of *I. indigotica* were collected from the School

of Pharmacy, Second Military Medical University, Shanghai, China, and authenticated by Professor ZHANG Han-Ming (Department of Pharmaceutical Botany, School of Pharmacy, Second Military Medical University). The seeds were pretreated with 75% alcohol for 5 min, washed thrice with distilled water, treated with 0.1% $HgCl_2$ for 10 min, and then washed with sterile distilled water four times. The sterilized seeds were incubated between several layers of sterilized wet filter paper and then cultured in MS basal medium for germination. The seedlings were grown at 25 °C under 16-h light/8-h dark photoperiod cycles for 2 months until treatments, for RNA and DNA isolation.

Hairy root culture and various treatments

The *I. indigotica* hairy root cultures were derived after the infection of plantlets with Ri T-DNA-bearing *A. rhizogenes* bacterium (C58C1). Hairy roots developed at cut edges 2−3 weeks after co-cultivation were excised and cultured in solid, hormone-free, half-strength MS medium. After bacteria were eliminated, the hairy roots (0.1 g fresh weight) were cultivated in a 250-mL shake flask containing 200 mL of the liquid basal medium on an orbital shaker at 110 $r \cdot min^{-1}$ at 25 °C in the dark.

After 4 weeks of shaking culture, the hairy roots at the exponential growth phase were prepared for induction. The hairy roots were treated with 100 $\mu mol \cdot L^{-1}$ of MeJA and harvested at various times (0, 2, 4, 6, 8, 12, and 24 h) after treatment. For UV-B treatment, the hairy roots were exposed to 1 500 $J \cdot m^{-2}$ UV-B light and sampled at 0, 5, 10, and 30 min, respectively. Then UV-B light was turned off and samples were taken at 30, 60, and 120 min post-treatment.

RNA and DNA isolation

The roots, stems, leaves, and flowers of *I. indigotica* as well as hairy root samples collected at various times after various treatments were used for RNA isolation. Total RNA was extracted using RNA prep pure plant kit (Tiangen Biotech Co., Beijing, China), according to the manufacturer's protocol. The genomic DNA was isolated using the modified CTAB method [26]. The quality and concentration of RNA and DNA samples were examined by EB-stained agarose gel electrophoresis and spectrophotometer analysis on a Helios Gamma ultraviolet spectrophotometer (Thermo electron corporation, Waltham, Massachusetts, USA.

Discovery of IiPALs from the transcription profiling database

In order to obtain PAL genes, we searched the *I. indigotica* transcription profiling database [25] through tBLASTn analysis using protein sequences, nucleotide sequences, and expressed sequence tag (EST) records of target genes of other plants from the National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov/). An e-value cut-off of 10^{-5} was applied to the homologue recognition.

The Pfam database (http://pfam.janelia.org/) [27] was used to screen the above putative sequences and identify the conserved protein domains using default parameters. As a final quality check, the simple modular architecture research tool (SMART, http://smart.embl-heidelberg.de/) [28] was used to find the PAL domain.

Molecular cloning of the IiPALs full-length cDNA

Total RNA isolated from *I. indigotica* was reversely transcribed using TransScript First-Strand cDNA Synthesis Super Mix (TransGen Biotech Co., Beijing, China, The full length of *IiPAL1* was cloned based on the sequence obtained from Lu [24] with primers for *IiPAL1*-F and *IiPAL1*-R through the PCR reaction under the following conditions: denatured at 94 °C for 2 min, followed by 35 cycles of amplification (94 °C for 35 s, 56 °C for 35 s, and 72 °C for 3 min), and 72 °C for 10 min.

Molecular cloning of *IiPAL2* from *I. indigotica* was based on the sequencing result from transcription profiling. The full length cDNA sequence was obtained by using the first-strand cDNA as the template under the following PCR conditions: 1 min at 95 °C, 35 cycles of amplification (20 s at 95 °C, 20 s at 60 °C, and 75 s at 72 °C), and a final extension of 5 min at 72 °C. The resulting amplified full length ORF was purified and cloned into PMD18-T vector and then sequenced.

Bioinformatics analysis

Sequence alignments and molecular mass calculation of the predicted protein were carried out on Vector NTI Advance 11. ORF translation and Genbank Blast were done on NCBI (http://www.ncbi.nlm.nih.gov). Phylogenetic analysis of *IiPAL2* and other known *PALs* from other plant species retrieved from GenBank were aligned uisng ClustalX software (version 1.80) and a phylogenetic tree was subsequently constructed using the neighbor joining (NJ) method [29] (1 000 bootstrap replicates) with the MEGA 5.0 software. Protein secondary structures were predicated using NetSurfP (http://www.cbs.dtu.dk/services/NetSurfP/) and SOMPA (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html) [30]. A homology model was generated from the native crystal structure of *Petroselinum crispum PAL* using Discovery studio 2.5 [31].

Expression profiling of IiPALs in different tissues under different stresses

The responses of *IiPAL1* and *IiPAL2* in hairy roots exposed to exogenous MeJA and UV-B and the expression profiles in different tissues were characterized using quantitative real-time PCR (Q-PCR). According to the corresponding sequences of *IiPAL1* and *IiPAL2,* gene specific primers *IiPAL1*-qRT-F, *IiPAL1*-qRT-R, *IiPAL2*-qRT-F and *IiPAL2*- qRT-R were designed. Partial polyubiquitin gene was amplified with primers *Actin*-F and *Actin*-R as a control. The Q-PCR assay was carried out in an assay mixture (final volume of 25 μL) containing 12.5 μL of 2 × SYBR Green Real Time PCR Master Mix (TaKaRa, Osaka, Japan), 0.5 μm of each primer, and 2 μL of cDNA. The program for all the Q-PCR reactions was as follows: 10 s pre-denaturation at 95 °C, 1 cycle; 5 s denaturation at 95 °C, 30 s annealing at 60 °C, 40 cycles; and separation reaction (15 s at 95 °C, 30 s

at 60 °C, 15 s at 95 °C). Quantification of the gene expression was done with comparative CT method. All the PCR reactions consisted of three technical replicates.

Subcellular localization of IiPALs

The full-length coding regions of *IiPAL1* and *IiPAL2* were amplified by PCR with sticky *Bsp*HI and *Spe*I ends inserted into the vector *pCAMBIA*1301-GFP under the control of cauliflower mosaic virus 35S promoter. The expression plasmids were transferred into the rice protoplast cells and were observed under a confocal microscope (Nikon, Tokyo, Japan) with argon laser excitation at 488 nm and a 505−530- nm emission filter set. The red autofluorescence of chlorophylls was imaged at emission wavelength longer than 650 nm [32].

Expression and characterization of IiPALs in Escherichia coli.

The full-length *IiPAL1* and *IiPAL2* cDNAs were cloned into plasmid pET32a(+) (Novagan, Copenhagen, Denmark) using *Eco*RV/*Eco*RI restriction sites to generate *IiPAL1*-pET-32a and *Ii*PAL2-pET-32a constructs. The gene-specific primers for *IiPAL1-Eco*RV-F, *IiPAL1-Eco*RI-R, *IiPAL2-Eco*RV-F, and *IiPAL2-Eco*RI-R were listed in Supplementary Table S1. After sequencing confirmation, the *IiPAL1*-pET-32a and *IiPAL2*-pET-32a constructs were transfected into *Escherichia coli* BL21(DE3) cells using the heat shock method. The *E. coli* BL21(DE3) cells harboring *IiPAL1*-pET-32a or *Ii*PAL2-pET- 32a in a single colony were inoculated at 37 °C into Luria-Bertani (LB) medium containing ampicillin (100 mg·L^{-1}) and grown with shaking (200 r·min^{-1}) at 37 °C until the optical density (OD600) reached about 0.6. The protein expression was induced for 4 h by an addition of isopropyl-*β*-D-thiogalactoside (IPTG, Bio-Rad, Berkeley, California, USA) at a final concentration of 1 mmol·L^{-1}. Protein purification was performed on BioLogic DuoFlow using Bio-Scale™ Mini chromatographic column, following the manufacturer's instructions (Bio-Rad). The purity of the His-tag-fused *Ii*PALs (ht-*IiPAL1* and ht-*IiPAL2*) was assessed by analyzing the total protein on 12% SDS-PAGE, followed by Coomassie Brilliant Blue R250 (Beyotime Biotech Co., Shanghai, China) staining and the protein concentration were determined by the Bradford method [33] using bovine serum albumin as the standard.

Enzyme activity assay for IiPAL1 and IiPAL2

The enzyme activities of *IiPAL*s were measured using the method of Yan [34] with minor modifications. The enzyme extract (0.5 mL) was incubated with 50 μL of 0.1 mmol·L^{-1} L-phenylalanine and 450 μL of 0.01 mmol·L^{-1} Tris-HCL, pH 7.5, at 37 °C for 60 min. The enzyme reaction was terminated by an addition of 50 μL of 5 mol·L^{-1} HCL after 1 h. After centrifugation at 25 °C (12 000 r·min^{-1}, 15 min), the compound determination was performed on an Agilent 1260 series liquid chromatography (Agilent, Santa Clara, California, USA) equipped with a quaternary solvent delivery system, an autosampler and a photodiode array detector (DAD). A Diamonsil C_{18} column (4.6 mm × 250 mm, 5 μm, Dikma Beijing, China) was used for analysis, using a mobile phase consisting of 48% acetonitrile (HPLC grade, Merck KGaA, Darmstadt, Germany) and 52% formic acid (0.1%, HPLC grade, Merck). The flow rate was set at 1.0 mL·min^{-1} and the injection volume was 10 μL. Elution of the compounds was monitored at 290 nm. Standards of L-phenylalanine and *trans*-Cinnamic acid were obtained from Sigma-Aldrich Co. (St. Louis, MO, USA).

The effects of the reaction time and protein concentration on the enzyme activity were examined. To determine the optimum reaction time, the assays were performed at 37 °C. The assay mixture consisted of 450 μL of 100 mmol·L^{-1} Tris-HCL (pH 7.5), 500 μL of purified *IiPAL1* or *IiPAL2* protein and 50 μL of 100 mmol·L^{-1} L-phenylalanine. 50 μL of 5 mol·L^{-1} HCL was used to stop reaction at different times (10, 20, 30, 40, and 50 s, and 1, 2, 3, 4, and 5 min). For optimal protein concentration determination, the reaction was carried out with different protein concentrations (10, 20, 40, 60, 80, 90, 100, 200, and 300 mmol·L^{-1}) at the same conditions as above. The *IiPAL1* and *IiPAL2* activities were determined by measuring absorbance of the reaction solution at 290 nm.

Transcript abundance of IiPALs in I. indigotica hairy roots treated with MeJA

The Illumia RNA-Seq data obtained in previous research [25] were utilized to get an insight into the *IiPAL*s' transcript abundance induced with MeJA in *I. Indigotica*. The RNA-Seq expression profile data were generated using the Illumina HiSeq™ 2000 platform, after the hairy roots of *I. indigotic*a were treated with MeJA at different times (0, 1, 3, 6, 12, and 24 h). The expression levels at different times were normalized to the level of the control (0 h).

Metabolite analysis

MeJA-treated hairy roots of *I. indigotica* (100 mg) were dried at 45 °C, ground into fine powder, and extracted twice with 10 mL of methanol under sonication for 30 min. After centrifuged at 4 000 r·min^{-1} for 5 min, the supernatant was diluted with methanol to 10 mL in total volume. The final solution was filtered through a 0.22-μm organic membrane filter prior to analysis.

The LC-MS/MS analysis was conducted on an Agilent 1200 series coupled with an Agilent 6410 triple Quadrupole mass spectrometer and an electrospray ionization source (Agilent). The data were processed with MassHunter Workstation Software. Chromatographic separation was achieved on an Agilent ZORBAX SB-C_{18} (3.5 μm, 100 mm × 2.1 mm i.d.) at column temperature of 35 °C. The mobile phase consisted of acetonitrile (eluent A) and 5 mmol·L^{-1} of ammonium acetate solution (eluent B, HPLC grade, Merck KGaA, Darmstadt, Germany), eluted at a rate of 0.3 mL·min^{-1} with a gradient program as follows: 0−4.00 min, 14% A; 4.00−4.50 min, 50% A, 4.50−8.50 min, 85% A and the run time was 8.5 min. The injection volume was 5 μL. Quantification was achieved in multiple reaction monitoring mode (MRM), and the selected transitions of *m/z* were

401→179 for coniferin, 359→329 for laricinesinol, 361→164 for secoisolariciresinol, and 357→151 for pinoresinol. All standards were purchased from Sigma-Aldrich Co..

Correlation analysis of genes and metabolites

The correlations between two *IiPALs* and four lignans were calculated using the Pearson correlation coefficient by canonical correlation analysis [33]. Gene-to-metabolite network was visualized to identify probable relation between *IiPALs* and lignan biosyntheses.

Results

Molecular cloning of the IiPALs full-length cDNA

The decoding of the *I. indigotica* transcriptome enabled us to identify a novel PAL gene (*IiPAL2*) that had not been reported before. The full-length cDNA sequence of *IiPAL2* was verified by PCR amplification.

The open reading frame of *IiPAL2* was slightly shorter than that of *IiPAL1* (2 115 bp *vs.* 2 530 bp), but *IiPAL1* and *IiPAL2* only shared a 70.06% of sequence identity. The deduced amino acid sequences of *IiPAL1* and *IiPAL2* included 725 and 705 amino acids (Fig. 2). The predicted molecular masses of *IiPAL1* and *IiPAL2* proteins were 78.65 kDa and 76.89 kDa respectively, and their theoretical pI were 5.96 and 5.63, respectively.

Protein-protein BLAST showed that *IiPAL1 and IiPAL2* had a high degree of similarity (77%−93%) to *PALs* from other plant species at the amino acid level. A detailed sequence alignment of the *IiPALs* proteins is shown in Fig. 3. The results showed that *IiPALs* contained the conserved Ala-Ser-Gly (216−218) catalytic triad. In addition, sequence alignment showed that *IiPALs* contained conserved deamination sites (*i.e.*, L-220, V-221, L-269, and A-270) and catalytic active sites (*i.e.*, N-272, G-273, H-350, NDN [396−398 aa], and HNQDV [500−505 aa]), just as *Dendrobium candidum* [14].

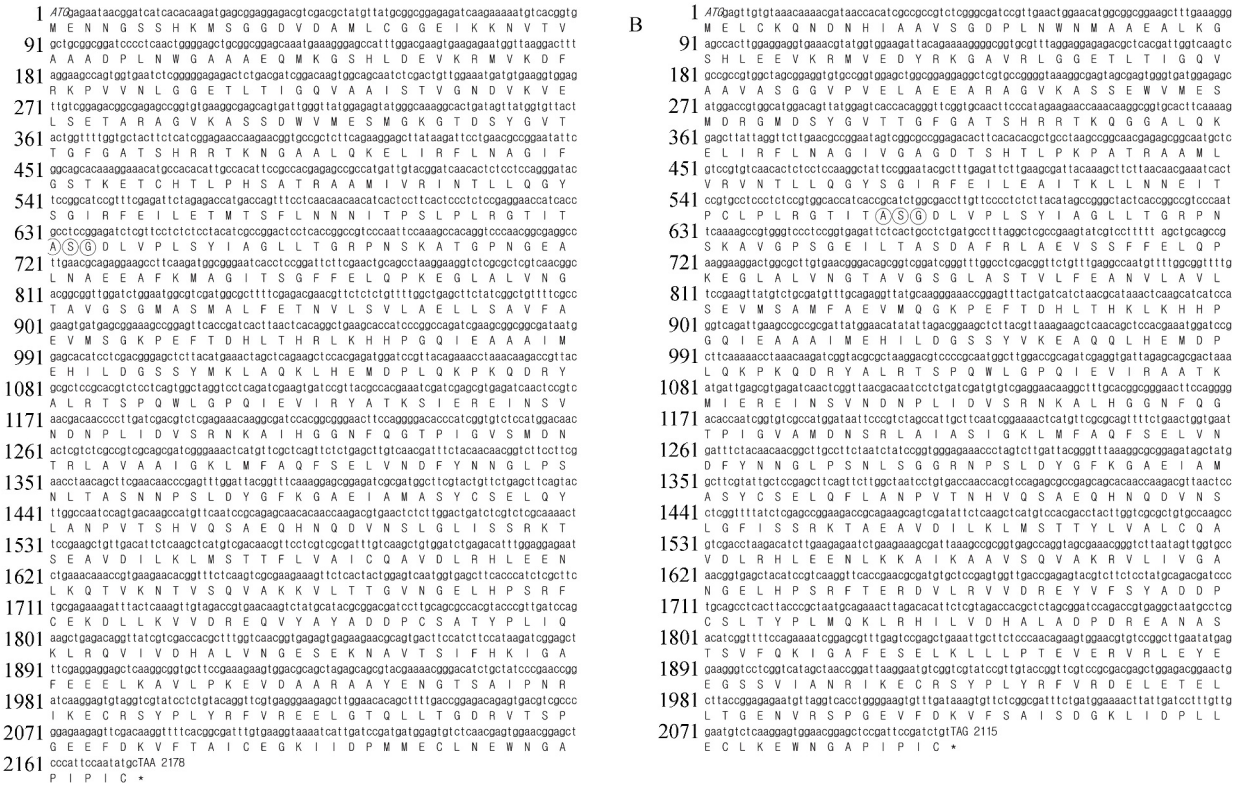

Fig. 2 Nucleotide sequences and deduced amino acid sequences of *IiPALs*. (A) *IiPAL1*, (B) *IiPAL2*. The start codon (ATG) is in italics and the stop codon (TGA or TAA) is indicated by an asterisk. The conserved active sites motifs are circled

The secondary structures are shown in Fig. 4A and the differences between *IiPAL1 and IiPAL2* are listed in Table 1. The three-dimensional structures of *IiPALs* were predicted using Discovery Studio 2.5 with *P. Crispum PAL* as a template. All of the *IiPALs* proteins contained four domains (Fig. 4B), including the N-terminus residues 1−24, numbered according to *IiPAL1*, the MIO domain (residues 25−262), the core domain (residues 263−529 and 652−716), and the inserted shielding domain (residues 530−651).

Phylogenetic analysis of the IiPALs

To investigate the evolutionary relationships among *IiPALs* and *PALs* from other plant species, the phylogenetic tree was constructed using the neighbor-joining method. The gymnosperm *PAL* protein from *Pinus taedawas* selected as outgroup, and the phylogenetic tree was grouped into two main branches, dicotyledons and monocotyledons. The *IiPALs* in the present study were most closely related to the dicot *PALs*, and they were classified as members of this group, with

a closest relationship with *AtPAL1* and *AtPAL2,* forming a Cruciferae cluster (Fig. 4C). A total of 20 protein sequences used for analysis were from following plant species: *IiPAL1* (DQ468345), *AtPAL1* (L33677), *AtPAL2* (L33678), *NtPALB* (AB008200), *PtPAL* (P52777), *GmPAL1* (X52953), *CsPAL* (D26596), *MsPAL* (X58180), *ClPAL6* (U43338), *DoPAL1* (AY450643), *PsPAL* (D10003), *RiPAL2* (AF237955), *ShPAL*(L36822), *MePAL1* (AY036011), *BnPAL2* (AY795080), *TaPAL* (X99705), *PaPAL1* (AF036948), *OsPAL* (XM473196), *ZeaPAL* (L77912), and *TpPAL* (AB236800).

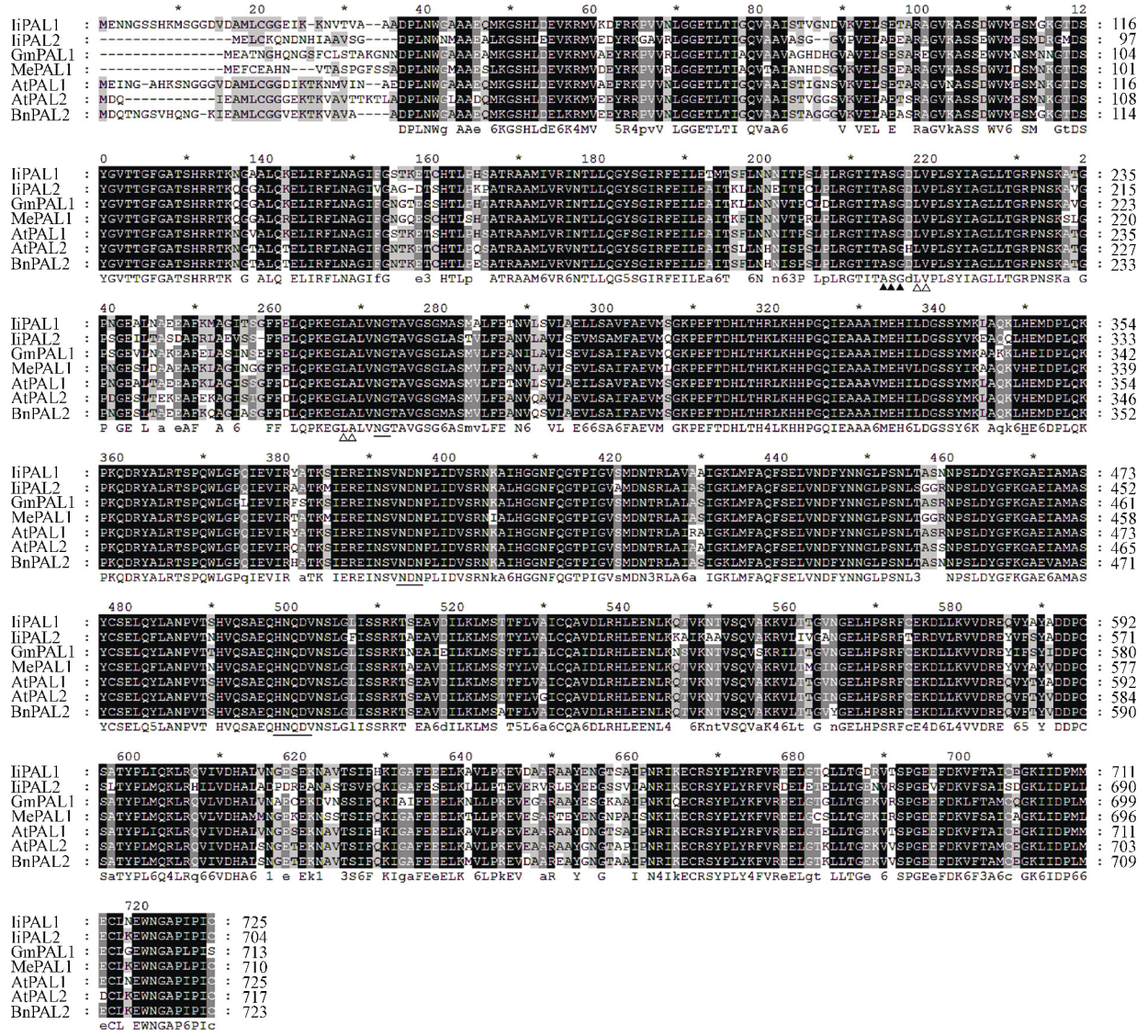

Fig. 3 Multi-sequence alignment of *IiPALs* with other plant PAL proteins. The sequences shown here are from *AtPAL1* (*A. thaliana*, L33677), *AtPAL4* (*A. thaliana*, NP187645), *VvPAL* (*Vitis vinifera*, XP 002268732), *PtPA*L (*Populus trichocarpa*, ACC63889.1), and *BnPAL* (*Brassica napus*, AY795080)

Tissue-specific and induced expression profile of IiPALs

The expression profile of *IiPAL1* and *IiPAL2* in different tissues of *I. indigotica* showed that *IiPAL1* and *IiPAL2* expression could be detected in all tissues with different expression levels. The transcript level of *IiPAL1* was the highest in flowers, whereas that of *IiPAL2* was highest in roots. The expressions of *IiPAL1* and *IiPAL2* in leaves and flowers were almost the same (Fig. 5A).

To understand the role of *IiPALs* in responses to plant defense, the plants were treated with MeJA and UV-B respectively. The results revealed that the transcription levels of *IiPALs* were responsive to different treatments to various degrees. As shown in Fig. 5B, the *IiPAL1* expression under the MeJA induction was rapidly and strongly induced, peaked at 4 h (5-fold of original value), and gradually decreased at 6−8 h. Interestingly, the expression level of *IiPAL1* increased again at 12 h, and then decreased a little after 24 h. For *IiPAL2*, the transcript abundance in the hairy root was also increased, peaked at 8 h, and then declined. After UV-B treatment (Fig. 5C), *IiPAL1* and *IiPAL2* displayed the highest expression levels at 10 min with about 4- and 9-fold increases, respectively, and then decreased after 30 min. After UV-B

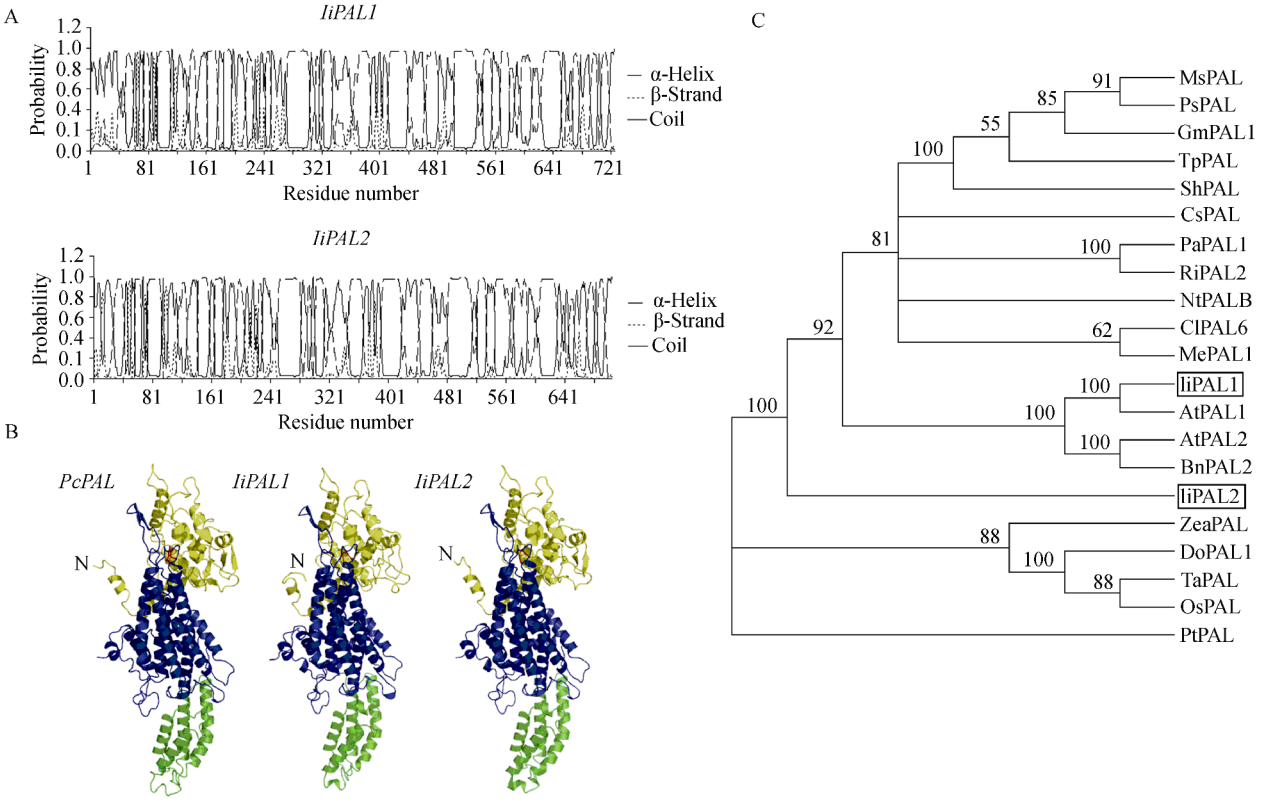

Fig. 4 Bioinformatics analysis of *IiPALs*. (A) predicts secondary structures of *IiPAL1* and *IiPAL2*. The short-term, green, dotted line and random coil represent alpha helix, beta turn, random coil, respectively; (B) Three dimensional protein model of *IiPALs* protein. The chains corresponding to the MIO domain (golden), core domain (blue), inserted shielding domain (green) and the MIO group (red) are highlighted; (C) Phylogenetic tree of *IiPALs* and other plant *PAL* genes using the Clustal X software and MEGA 5.0 software based on the Neighbour-joining method. *IiPAL1* and *IiPAL2* are marked by black box

Table 1 The differences in secondary structures between *IiPAL1* and *IiPAL2*

Gene	*α*-helices	*β*-turns	Extended strands	random coils
IiPAL1	366 (50.48%)	60 (8.28%)	81 (11.17%)	218 (30.07%)
IiPAL2	388 (55.11%)	36 (5.11%)	51 (7.24%)	299 (32.53%)

exposure was terminated, the transcription levels remained increased to certain degrees.

Subcellular localization of IiPALs

As shown in Figs. 5D−5G, the rice protoplast expressing *IiPAL1*-GFP showed green fluorescent signals, which was localized in the cytoplasm (Fig. 5D). The fluorescence of the *IiPAL1*-GFP fusion was exclusively distributed in cytoplasm when merged with signals shown in panels D and E (Fig. 5F). The results of *Ii*PAL2-GFP were the same as that of *Ii*PAL1-GFP.

Functional analysis of IiPAL1 and IiPAL2

The *in vitro* functional activities of *IiPAL1 and IiPAL2* were investigated by expressing the genes in *E. Coli* BL21 (DE3). The fractionation analysis using SDS-PAGE showed molecular masses of about 97 kDa and 95 kDa (including the tags), respectively (Fig. 6). It was in good agreement with that predicted by the bioinformatics method. Enzyme assays were carried out under standard conditions, using purified protein and L-Phe as the substrate. The control reaction with L-Phe using the same buffer without the PAL enzyme gave a peak of L-Phe (Fig. 7A). After L-Phe catalyzed by recombinant proteins ht-*IiPAL1* and ht-*IiPAL2*, respectively, only one peak at about 23 min could be seen, which should be ascribed to the L-Phe conversion to *trans*-cinnamic acid (Figs. 7C−7D).

Moreover, the time course for expression of the target proteins was also examined. From Fig. 8A, it was evident that, in a short time (6 min), the amount of the protein production increased over time. And the promoting reaction rate of *IiPAL2* rose quickly at 30 s, from which *IiPAL2* had a stronger catalytic ability than *IiPAL1*.

Meanwhile, the protein concentration was proportional to enzyme promoting reaction rate (Fig. 8B). The initial activity of *IiPAL1* was higher than that of *IiPAL2,* and the effects of different concentrations of *IiPAL2* on the enzymatic reaction were higher than that of *IiPAL1*. When the protein concentration of *IiPAL1* and *IiPAL2* reached 100 mmol·L^{-1}, the promoting reaction rate had a rapid increase and reached the peak at 200 mmol·L^{-1}, followed by a gradual decrease.

Integration of transcript and metabolite accumulation analyses

A canonical correlation analysis was performed to ex-

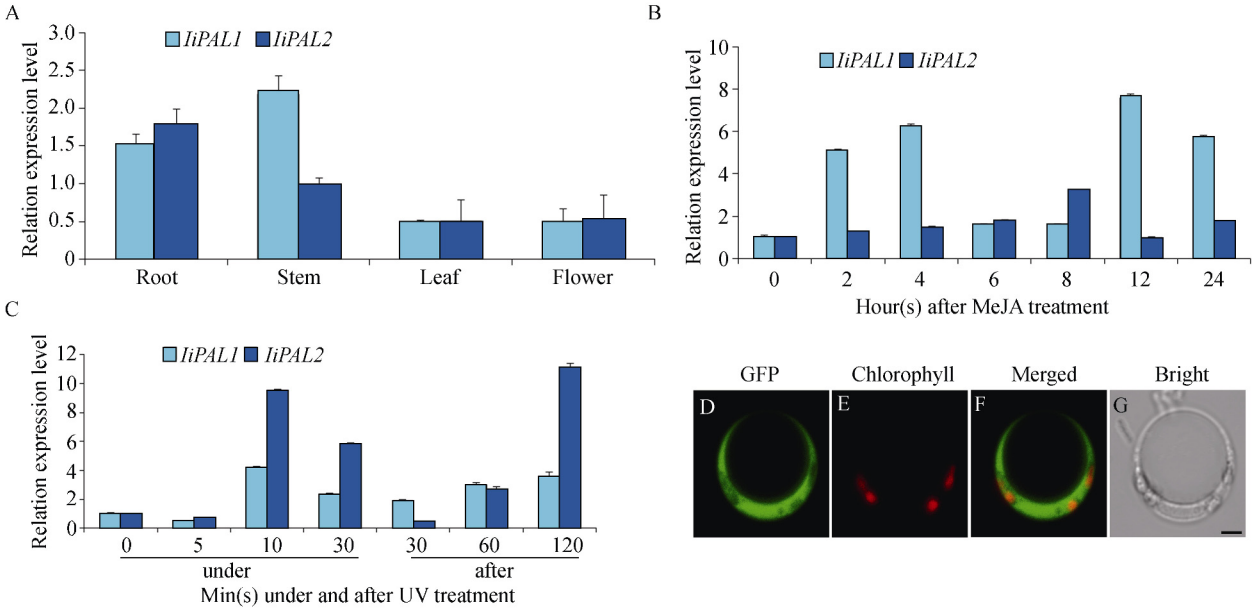

Fig. 5 Characterization of *IiPAL*s. (A) Transcript abundance of *IiPAL*s in different tissues of *I. Indigotica*; (B) *IiPAL*s expression under the induction of MeJA; (C) Fold changes of *IiPAL*s in hairy roots of *I. indigotica* treated with UV-B; (D-G) Analysis of *IiPAL1* subcellular localization (Bars = 3.0 μm)

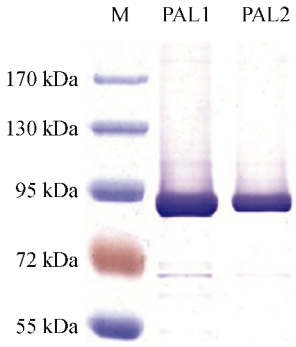

Fig. 6 SDS-PAGE analysis of the purified proteins. (M, protein marker; Lane 1, purified ht-*IiPAL1*; Lane 2, purified ht-*IiPAL2*)

plore possibly correlations between the accumulations of four lignans and the expression profiles of *IiPAL1* and *IiPAL2* (Fig. 9) with the variable correlation coefficient cut-off values being set at 0.6. The variable correlation coefficients between *IiPAL1* transcript and four metabolites accumulation (coniferin, lariciresinol, secoisolariciresinol, and pinoresinol) were −0.184, 0.954, 0.573, and 0.873, respectively, whereas, that of *IiPAL2* were −0.136, 0.200, 0.503 and 0.072, respectively. These results suggested that *IiPAL1* was most likely to be involved in the biosynthesis of lariciresinol and its precursor (pinoresinol). However, *IiPAL2* seemed to be unrelated to any of the measured lignans.

Discussion

PAL is one of the branch point enzymes that link primary metabolism to secondary metabolism [35]. It catalyzes the first step in the formation of cinnamic acid, a precursor, to a variety of phenylpropanoid derivatives. In the present study, we isolated a novel *I. indigotica PAL* gene, *IiPAL2*. The existence of a small multigene family in *I. indigotic* was consistent with the results from other plant species such as *S. Viminalis, S. Miltiorrhiza and C. canephora* [15, 19-20].

IiPAL1 and *IiPAL2* encoded 725 and 711 amino-acid proteins, respectively, whose lengths were similar to that of other reported *PAL*s [36]. Sequence analysis and homology modeling revealed that *IiPAL*s shared identical characteristics with many other *PAL*s. *IiPAL*s also possessed a conserved Ala- Ser-Gly (216−218) catalytic triad (Figs. 2 and 3). As reported in other *PAL*s, the Ala-Ser-Gly triad can be converted into the MIO prosthetic group by cyclization and the elimination of water [37]. Meanwhile, the conserved deamination and catalytic sites may participate in substrate selectivity and binding, catalysis and/or the formation of the MIO prosthetic group. This analysis suggested that *IiPAL1* and *IiPAL2* are members of *PAL* family and the proteins may have the same catalytic function as other *PAL* proteins.

The members of the *PAL* gene family in a plant are usually expressed differently in tissues and appear to be functionally distinct. The *RiPAL1 in Rubus idaeus* is associated with early fruit ripening events, whereas expression of *RiPAL2* is correlated more with later stages of flower and fruit development [38]. In *C. lanatus*, only six of the 12 *ClPAL*s may have the potential roles of developing fruit color and flavor [23]. Consistently, differential expression patterns were observed from the two *IiPAL*s, although they were expressed in all of the tissues analyzed. *IiPAL1* was found to be highly expressed in the stems, while *IiPAL2*

predominately showed expressions in the roots. The root of *I. indigotica* is the well-known Chinese medicine “Banlangen” (Radix *Isatidis*), and the distribution of *IiPAL*s in the tissues suggests that *IiPAL2* may play possible roles in roots through lignin formation and participate in the developmental processes of *I. indigotica*. However, the role of *IiPAL1* cannot be denied, because some genes may mediate active pharmaceutical compounds biosynthesis in aerial organs [39].

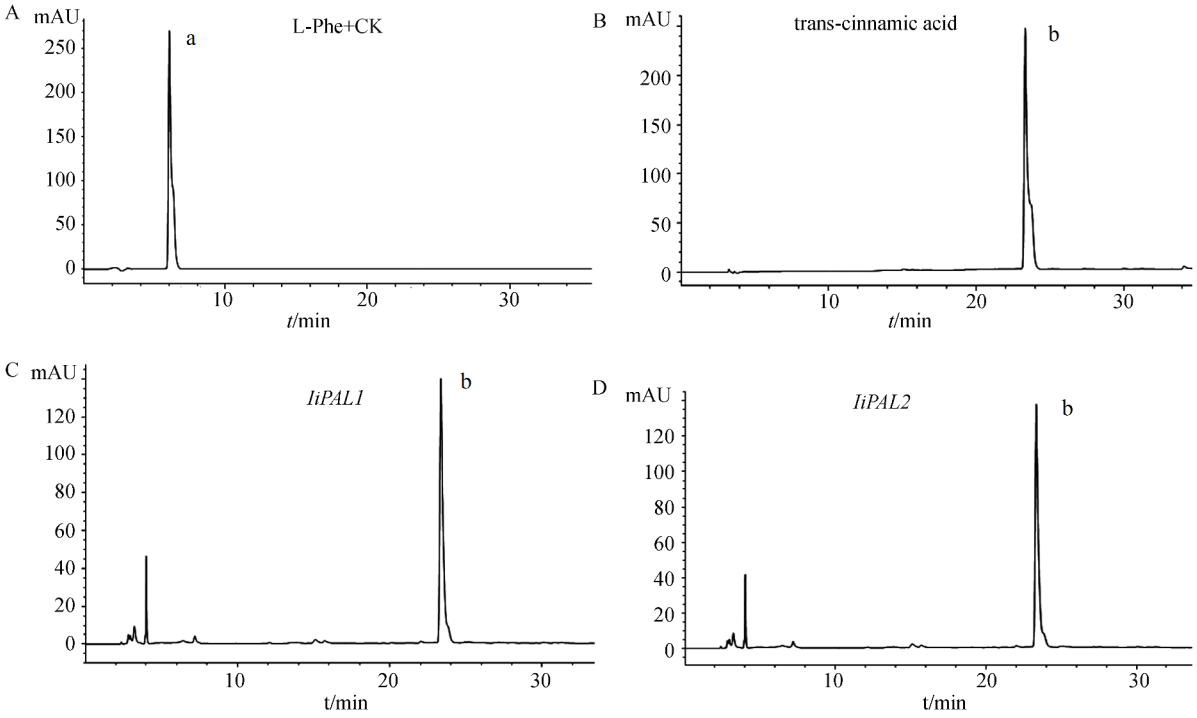

Fig. 7 Representative HPLC profiles of incubation of L-Phe without (A) or with (C, D) purified *ht-IiPAL*s enzyme. (B) trans-cinnamic acid standard (Peak a: L-Phe; peak b: trans-cinnamic acid)

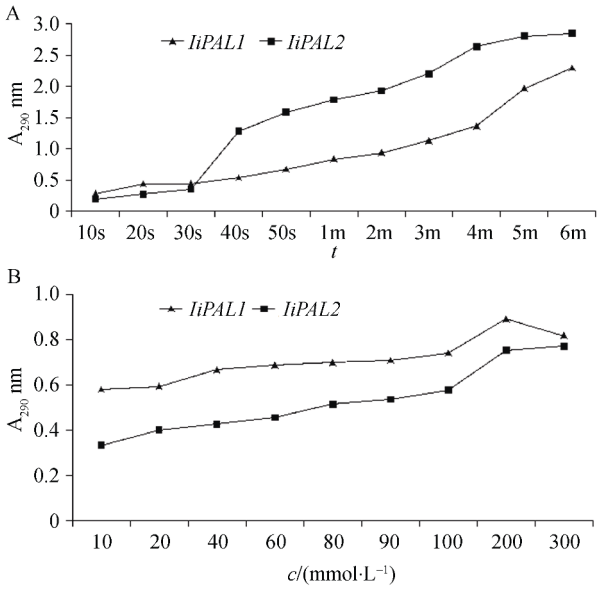

Fig. 8 The effects of the reaction time (A) and protein concentration (B) on the enzyme activity. Trans-cinnamon acid has the maximum absorption at 290 nm

PAL is one of the most extensively studied enzymes with respect to plant responses to biotic and abiotic stresses. Expression analyses demonstrated that *PAL* could be widely stimulated by environmental elicitors such as low temperature, dehydration, and UV irradiation [40]. Plant signaling molecules, including ABA, salicylic acid (SA), and JA [41], have also been demonstrated to elicit *PAL* activity. In general, all the elicitors chosen in the present study could escalate the *IiPAL*s expression, but the up-regulations may be caused by different mechanisms.

Consistently, *IiPAL*s transcripts in hairy roots were enhanced after MeJA and UV-B treatments (Fig. 5). However, differences were observed for two *IiPAL*s in the time and degree of responses after treated with MeJA. Similar results were previously found in three *PAL*s from *Salvia Miltiorrhiza*. These differences in response to MeJA might be explained by the complexity and diversity of MeJA. On the other hand, the expressions of *IiPAL1* and *IiPAL2* were similar after irradiation with UV-B. They showed the highest induction at the same time under UV-B induction and the increase didn’t disappear after the removal of UV-B, as seen with *P. Kurrooa*. Many elicitors could modulate the production of many important secondary metabolites of the phenylpropanoid pathway, such as bisbibenzyl in *Plagiochasma appendiculatum* [17], lignins and anthocyanins in *A. thaliana* [42], and phenolics and flavonoids in *Hypericum perforatum* [43]. So it may be used as a new strategy to improve the content of phenylpropanoid compounds.

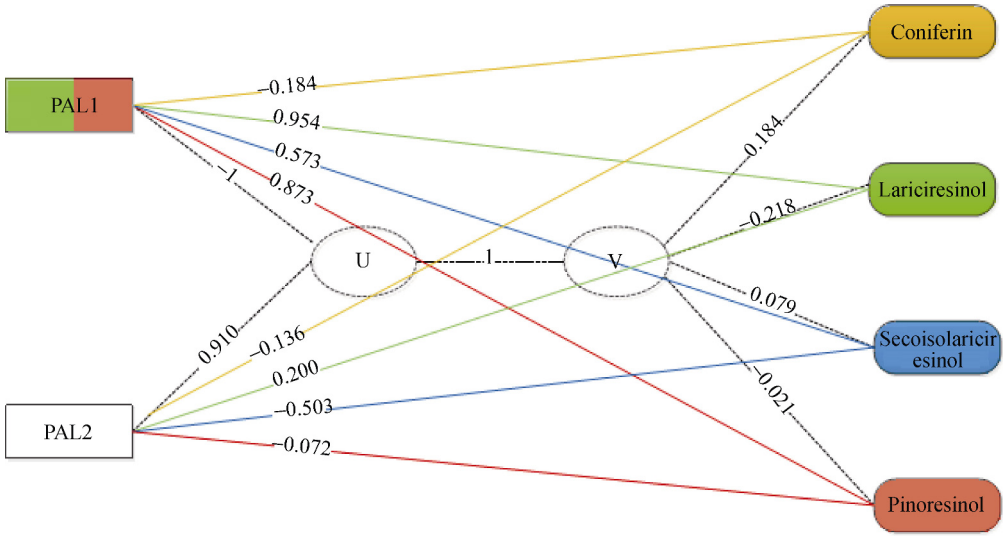

Fig. 9 Correlation network for the metabolites (right) and IiPALs (left) with a cut-off |r| > 0.6. The canonical correlation coefficient between two canonical correlation variables (U and V) was 1.00. The gene color in the gene-squares represents the level of gene-to-metabolites correlation: green area represents correlation to lariciresinol, red area represents correlation to pinoresinol, and the white color represents no correlation to four lignans

Subcellular localization analyses indicated that *IiPALs* were localized in cytoplasm, which was consistent with the results obtained with *S. viminalis* [15]. At subcellular level, *PAL* was mainly located in cytoplasm and chloroplast, mitochondria, glyoxysome, peroxisoame, and other membrane organelles [44]. Jin Nakashima has proven that *PAL* activity in the cytosolic is the best [45]. The results of subcellular localization suggested that *IiPALs* could be localized in cytoplasm, indicating that *IiPALs* may have a high PAL activity.

To date, many plant PAL genes have been cloned and successfully expressed *in vitro*, such as *PALs* in *Artemisia annua* [46]. To further confirm its function, *IiPALs* were expressed in *E. coli* BL21 (DE3) in the present study. The recombinant protein could catalyze the elimination of ammonia from L-phenylalanine to form cinnamic acid, revealing that *IiPALs* encoded a functional protein with a high PAL activity. As the gateway enzyme, PAL plays a key role in mediating carbon flux from primary metabolism in the phenylpropanoid pathway. It may be a new target for enhancing the content of pharmaceutical compounds in *I. Indigotica*, just as Junli Chang reported with *Arabidopsis thaliana* [47].

In previous studies, the integration of transcriptomics and metabolomics data is used as an important mean to dig crucial genes involved in the synthesis of target compounds [44]. In the present study the correlation analysis indicated that *IiPAL1* transcript was coincident with the accumulation of lariciresinol, suggesting that *IiPAL1* was relevant with the biosynthesis of active pharmaceutical compounds in *I. Indigotica*. However this result needs to be validated by metabolic engineering in the future.

Taken together, our results showed the existence of another *IiPAL* gene in *I. Indigotica*. The cloning and characterization of two distinct *I. Indigotica* PAL genes provided a basis for further undertaking a detailed molecular and genetic analysis of the regulation of this key gene family in *I. Indigotica*. The different expression profiles showed that roles of each *IiPAL* may be different in the biosynthesis of secondary metabolites. As a key step, further studies concerning the roles of *IiPALs* using *Agrobacterium* tumefaciens-mediated genetic transformation would help better understand the metabolic network of secondary metabolites in *I. Indigotica*, which may also provide a new strategy for enhancing the content of pharmaceutical compounds.

References

[1] Al-Shehbaz IA, Beilstein MA, Kellogg EA. Systematics and phylogeny of the *Brassicaceae* (Cruciferae): an overview [J]. *Plant Syst Evol*, 2006, **259**: 89-120.

[2] *Pharmacopoeia of the Peoples Republic of China* (2010). Vol 1 [S]. 2010: 161

[3] Deng TT. Discussion on treatment of SARS by TCM [J]. *Tianjin J Traditional*, 2003, **3**: 4-8.

[4] Wang YT, Yang ZF, Zhan HS, *et al*. Screening of anti-H1N1 active constituents from Radix Isatidis [J]. *Int J Mol Sci*, 2011, **28**: 419-422.

[5] Li B. *Active ingredients and quality evaluation of Isatis Indigotica* [M]. Shanghai: Second Military Medical University, 2003.

[6] Ahmad I, Waheed A, Tahir NB, *et al*. Anti-inflammatory constituents from *Perovskia atriplicifolia* [J]. *Pharm Biol*, 2015, **53**(11): 1628-1631.

[7] Hwang B, Cho J, Hwang IS, *et al*. Antifungal activity of lariciresinol derived from *Sambucus williamsii* and their membrane-active mechanisms in Candida albicans [J]. *Biochem Biophys Res Commun*, 2011, **410**(3): 489-493.

[8] Adlercreutz H, Mousavi Y, Clark J, et al. Dietary phytoestrogens and cancer: *in vitro* and *in vivo* studies [J]. *J Steroid Bio chem*

Mol Biol, 1992, **41**(3-8): 331-337.

[9] Verpoorte R, Alfermann AW, Johnson TS. *Applications of Plant Metabolic Engineering* [M]. Springer, 2007: 14-17.

[10] Vogt T. Phenylpropanoid biosynthesis [J]. *Mol Plant*, 2010, **3**(1): 2-20.

[11] Jones D. Phenylalanine ammonia-lyase: regulation of its induction, and its role in plant development [J]. *Phytochemisty*, 1984, **23**(7): 1349-1359.

[12] Koukol J, Konn EE. The metabolism of aromatic compounds in higher plants. Ⅳ. Purification and properties of the-phenylalanine deaminase of *Herdeum vulagare* [J]. *J Biol Chem*, 1961, **236**: 2692-2698.

[13] Ejtahed RS, Radjabian T, Hoseini TS. Expression analysis of phenylalanine ammonia lyase gene and rosmarinic acid production in *Salvia officinalis* and *Salvia virgata* shoots under salicylic acid elicitation [J]. *Appl Biochem Biotechnol*, 2015, **176**(7): 1846-1858.

[14] Jin Q, Yao Y, Cai Y, *et al*. Molecular cloning and sequence analysis of a phenylalanine ammonia-lyase gene from *Dendrobium* [J]. *PLos One*, 2013, **8**(4): e62352.

[15] de Jong F, Hanley SJ, Beale MH, *et al*. Characterisation of the willow phenylalanine ammonia-lyase (PAL) gene family reveals expression differences compared with poplar [J]. *Phytochemistry*, 2015, **117**: 90-97.

[16] Bhat WW, Razdan S, Rana S, *et al*. A phenylalanine ammonia-lyase ortholog *(PkPAL1*) from *Picrorhiza kurrooa* Royle ex. Benth: Molecular cloning, promoter analysis and response to biotic and abiotic elicitors [J]. *Gene*, 2014, **547**(2): 245-256.

[17] Yu H, Liu X, Gao S, *et al*. Molecular cloning and functional characterization of a phenylalanine ammonia-lyase from liverwort *Plagiochasma appendiculatum* [J]. *PCTOC J Plant Biotechnol*, 2014, **117**(2): 265-277.

[18] Yun YH, Koo JS, Kim SH, *et al*. Cloning and expression analysis of phenylalanine ammonia-lyase gene in the mycelium and fruit body of the edible Mushroom *Flammulina velutipes* [J]. *Mycobiology*, 2015, **43**(3): 327-332.

[19] Hou X, Shao F, Ma Y, *et al*. The phenylalanine ammonia-lyase gene family in *Salvia miltiorrhiza*: genome-wide characterization, molecular cloning and expression analysis [J]. *Mol Biol Rep*, 2013, **40**(7): 4301-4310.

[20] Lepelley M, Mahesh V, McCarthy J, *et al*. Characterization, high-resolution mapping and differential expression of three homologous PAL genes in *Coffea canephora* Pierre (Rubiaceae) [J]. *Planta*, 2012, **236**(1): 313-326.

[21] Shi R, Shuford CM, Wang JP, *et al*. Regulation of phenylalanine ammonia-lyase (PAL) gene family in wood forming tissue of *Populus trichocarpa* [J]. *Planta*, 2013, **238**(3): 487-497.

[22] Tonnessen BW, Manosalva P, Lang JM, *et al*. Rice phenylalanine ammonia-lyase gene OsPAL4 is associated with broad spectrum disease resistance [J]. *Planta Mol Biol*, 2015, **87**(3): 273-286.

[23] Dong C, Shang Q. Genome-wide characterization of phenylalanine ammonia-lyase gene family in watermelon (*Citrullus lanatus*) [J]. *Planta*, 2013, **238**(1): 35-49.

[24] LU B. Cloning and characterization of a differentially expressed phenylalanine ammonialyase gene (*IiPAL*) after genome duplication from tetraploid *Isatis indigotica* Fort. [J]. *J Integr Plant Biol*, 2006, **48**(12): 1439-1449.

[25] Chen J, Dong X, Li Q, *et al*. Biosynthesis of the active compounds of *Isatis indigotica* based on transcriptome sequencing and metabolites profiling [J]. *BMC Genomics*, 2013, **14**: 857.

[26] Porebski S, Bailey L, Baum B. Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components [J]. *Plant Mol Biol Rep*, 1997, **15**: 8-115.

[27] Finn RD, Bateman A, Clements J, *et al*. Pfam: the protein families database [J]. *Nucleic Acids Res*, 2014, **42**(Database issue): D222-D230.

[28] Schultz J, Copley RR, Doerks T, *et al*. SMART: a web-based tool for the study of genetically mobile domains [J]. *Nucleic Acids Res*, 2000, **28**(1): 231-234.

[29] Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees [J]. *Mol Biol Evol*, 1987, **4**(4): 406-425.

[30] Hussain MR, Shaik NA, Al-Aama JY, *et al*. In silico analysis of single nucleotide polymorphisms (SNPs) in human BRAF gene [J]. *Gene*, 2012, **508**(2): 188-196.

[31] Lois R, Dietrich A, Hahlbrock K, *et al*. A phenylalanine ammonia-lyase gene from parsley: structure, regulation and identification of elicitor and light responsive cis-acting elements [J]. *EMBO J*, 1989, **8**(6): 1641-1648.

[32] Tan H, Xiao L, Gao S, *et al*. Trichome and artemisinin regulator 1 is required for trichome development and artemisinin biosynthesis in artemisia annua [J]. *Mol Plant*, 2015, **8**(9): 1396-1411.

[33] Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding [J]. *Anal Biochem*, 1976, **72**: 248-254.

[34] Yan Q, Shi M, Ng J, *et al*. Elicitor-induced rosmarinic acid accumulation and secondary metabolism enzyme activities in *Salvia miltiorrhiza* hairy roots [J]. *Plant Sci*, 2006, **170**: 853-858.

[35] Butland SL, Chow ML, Ellis BE. A diverse family of phenylalanine ammonia-lyase genes expressed in pine trees and cell cultures [J]. *Plant Mol Biol*, 1998, **37**(1): 15-24.

[36] Cramer CL, Edwards K, Dron M, *et al*. Phenylalanine ammonia-lyase gene organization and structure [J]. *Plant Mol Biol*, 1989, **12**(4): 367-383.

[37] MacDonald MJ, D'Cunha GB. A modern view of phenylalanine ammonia lyase [J]. *Biochem Cell Biol*, 2007, **85**(3): 273-282.

[38] Kumar A, Ellis BE. The phenylalanine ammonia-lyase gene family in raspberry, structure, expression, and evolution [J]. *Plant Physiol*, 2001, **127**(1): 230-239.

[39] Cui G, Duan L, Jin B, *et al*. Functional divergence of diterpene syntheses in the medicinal plant *Salvia miltiorrhiza* [J]. *Plant Physiol*, 2015, **169**(3): 1607-1618.

[40] Zhu Q, Xie X, Lin H, *et al*. Isolation and functional characterization of a phenylalanine ammonia-lyase gene (SsPAL1) from Coleus (*Solenostemon scutellarioides* (L.) Codd) [J]. *Molecules*, 2015, **20**(9): 16833-16851.

[41] Munoz-Espinoza VA, Lopez-Climent MF, Casaretto JA, *et al*. Water stress responses of tomato mutants impaired in hormone biosynthesis reveal, abscisic acid, jasmonic acid and salicylic acid interactions [J]. *Front Plant Sci*, 2015, **6**: 997.

[42] Pietrowska-Borek M, Nuc K, Guranowski A. Exogenous adenosine 5'-phosphoramidate behaves as a signal molecule in plants; it augments metabolism of phenylpropanoids and salicylic acid in *Arabidopsis thaliana* seedlings [J]. *Plant Physiol Biochem*, 2015, **94**: 144-152.

[43] Gadzovska SS, Tusevski O, Maury S, *et al*. Effects of polysaccharide elicitors on secondary metabolite production and antioxidant response in *Hypericum perforatum* L. shoot cultures [J]. *Sci World J*, 2014, **2014**: 609649.

[44] Hanson K, Havir E. *Phenylalanine ammonia-lyase* [M]. New York: Acadamic Press, 1981: 577-625.

[45] Jin N, Tatsuya A, Keiji T. Immunocytochemical localization of phenylalanine ammonia-lyase and cinnamyl alcohol dehydrogenase in differentiating tracheary elements derived from *Zinnia* mesophyll cells [J]. *Plant Cell Physiolo*, 1997, **38**(2): 113-123.

[46] Zhang Y, Fu X, Hao X, *et al*. Molecular cloning and promoter analysis of the specific salicylic acid biosynthetic pathway gene phenylalanine ammonia-lyase (AaPAL1) from *Artemisia annua* [J]. *Biotechnol Appl Biochem*, 2016, **63**(4): 514-524.

[47] Chang J, Luo J, He G. Regulation of polyphenols accumulation by combined overexpression/silencing key enzymes of phenylpropanoid pathway [J]. *Acta Biochim Biophys Sin* (*Shanghai*), 2009, **41**(2): 123-130.

Cite this article as: MA Rui-Fang, LIU Qian-Zi, XIAO Ying, ZHANG Lei, LI Qing, YIN Jun, CHEN Wan-Sheng. The phenylalanine ammonia-lyase gene family in *Isatis indigotica* Fort.: molecular cloning, characterization and expression analysis [J]. *Chin J Nat Med*, 2016, **14**(11): 801-812.