Bangladesh J. Bot. **48**(3): 719-725, 2019 (September) Special

CLONING AND EXPRESSION OF CINNAMATE 4-HYDROXYLASE GENE FROM *DRYOPTERIS FRAGRANS* (L.) SCHOTT

LILI SUN, LIGUO ZHANG, XIAOFEI ZHANG[1], LINGLING CHEN[2], ZHIGANG BU[3], YING CHANG[2]*, FUDE WANG[4], YUNHUI MU, LIXIN CHEN AND LI GUO

Heilongjiang Academy of Agricultural Sciences, Harbin, Heilongjiang, 150069, China

Keywords: Gene clone, *Dryopteris fragrans*, Transformation, Cinnamate 4-hydroxylase, Overexpression

Abstract

The enzyme cinnamate 4-hydroxylase (C4H, EC 1.14.13.11), which belongs to the cytochrome P450 family, is involved in the phenylpropanoid pathway. Using RACE, full-length cDNA consisting of 2063 nucleotides encoding C4H (designated *DfC4H*) was isolated from *Dryopteris fragrans* (L.) Schott. A number of 1527 of nucleotides composed the complete open reading frame (ORF) of the cDNA (GenBank Accession Number KF830705.2). A protein that consisted of 508 amino acids with a calculated molecular mass of 58.1908 kDa was predicted to be encoded by this ORF. The overexpression of *DfC4H* was observed to increase significantly *p*-coumaric acid and anthocyanin levels in transgenic *Arabidopsis thaliana.*

Introduction

Dryopteris fragrans (L.) Schott, a perennial deciduous herb belonging to the family Dryopteridaceae is known as Xiang-Lin-Mao-Jue in China. This plant species is widespread in North America, Europe and Asia (Ao and Li 1990). Earlier evidence revealed that chemical compounds derived from the herb were of high medicinal importance (Wang *et al*. 2008, Chang 2009) due to their antioxidative and antibacterial effects. For these reasons, the properties of this *D. fragrans* have attracted considerable foreign and national research attention. The chemical composition of this medicinal plant is rather complex, including variety chemical substances, such as flavonoids, isophthalics, and terpenes (Widén *et al.* 2001). Using molecular biological techniques, fragments of both *DfC4H* and *DfPAL* gene family (Li *et al*. 2015), as well as the ORF of *DfCHS* (Sun *et al.* 2014) and a novel MADS-box gene, *DfMADS1* were cloned (Huang *et al*. 2014).

The enzyme cinnamic acid 4-hydroxylase, which is a representative of the family of cytochrome P450 (CYP73A) enzymes, is involved in the phenylpropanoid pathway (Teutsch *et al.* 1993, Achnine *et al.* 2004). C4H exerts multiple important functions; for example, its N-terminal hydrophobic region is involved in anchoring the complex of enzymes of general phenylpropanoid enzymes (C4H, 4-coumarate : coenzyme A ligase, and PAL) to the membrane of the endoplasmic reticulum (Winkel-Shirley 1999). Although the *C4H* gene ortholog was earlier cloned in *D. fragrans* (Li *et al.* 2015), the full-length sequence of the cDNA of the *C4H* gene has not yet been cloned. Therefore, to confirm the function of *C4H* in *D. fragrans*, in the present investigation, the full-length cDNA sequence of the *C4H* gene isolated from *D. fragrans* were cloned and characterized for the first time. In addition, bioinformatic analysis and transformation of this gene into *Arabidopsis thaliana* were also conducted.

*Author for correspondence: <suntianyi2006@126.com>. [1]College of Mathematics and Information Sciences of Guangxi University, Nanning, Guangxi, 530004, China. [2]Laboratory of Plant Research College of Life Sciences, Northeast Agricultural University, Harbin, Heilongjiang, 150030, China. [3]College of Biology, Hunan University, Changsha, Hunan, 410082, China. [4]Laboratory of Plant Research College of Life Sciences, Northeast Agricultural University, Harbin, Heilongjiang, 150030, China.

Materials and Methods

Description of the culture conditions and material used *Dryopteris fragrans* plants and spores were obtained after one year of growth in Heilongjiang, China (126°07′07″N, 48°42′38″E). Permissions were not explicitly required because the sampling was carried out in an experimental research area. Moreover, in China, *D. fragrans* is not protected or considered endangered.

Spores of *D. fragrans* were cultivated superficially on 1/2 MS culture medium under sterile conditions at a photoperiod of 12 hrs light/12 hrs dark and a temperature of 25 ± 1°C. The spores were allowed to grow into gametophytes. Upon fertilization and the subsequent development of the gametophytes into sporophytes, the latter were grown as secondary sporophyte cultures on the same medium.

Extraction of total RNA from *D. fragrans* sporophytes (100 mg) was performed by the RNA plant Plus Reagent (TIANGEN, Beijing, China) following the instructions of the manufacturer. Then, the total RNA concentration in each of the samples were measured spectrophotometrically. Next, the 5′- and 3′-ends of the corresponding cDNAs were obtained through a SMART^{TM}-RACE cDNA Amplification Kit (Clontech, Mountain View, CA, USA). The cDNA samples were then stored at –20°C for subsequent analysis.

RACE primers were developed based on the partial sequence of the *DfC4H* (KF830705) gene in *D. fragrans* (Li *et al*. 2015). Gene-specific primers (*DfC4H*-5′:5′-GCGTGTGTAGC ACCTCCTTGGCCAGCTC-3′ and *DfC4H*-3′: 5′-GGGAGAAGCCCCATGAATTTCG-3′) were designed according to the SMART^{TM} protocol (Clontech, Mountain View, USA). Universal primers were supplied in the kit.

Upon the completion of electrophoresis, the products of the PCR reactions from the agarose gels were recovered by a DNA gel extraction kit (HaiGene, Harbin, China). Next, the purified RACE products were cloned into the pEASY-T3 Cloning Vector (TransGen Biotech, Beijing, China) and positive clones from the two amplification reactions were selected randomly. These were then subjected to sequencing by Sangon Biotech Co., Ltd. (Shanghai, China). Analyses of the nucleotide and the deduced amino-acid sequences were performed by DNAMAN software.

Searches on BLAST using the website of NCBI and the NCBI program for identification of ORF were performed. Besides additional tools, including Motif Search and InterProscan were also employed to analyze the protein motif signature and the domain/functional sites, correspondingly. Finally, alignment of the protein sequences was implemented on BoxShade.

For construction of vectors, transformation, and transgenic plant growth, **s**tandardized techniques for DNA recombination using the strain DH5α of *Escherichia coli* were carried out using the method described by Sambrook *et al*. (1989). The fragment *Xba*I/*Bam*HI with the sequence *DfC4H* of cDNA were cloned from pEASY-T3-*DfC4H* in the place of the GUS-encoding gene of the binary vector plasmid pBI121 using the 35S promoter of the cauliflower mosaic virus (CaMV) and the terminator of nopaline synthase (NOS).

The aforementioned binary constructs and empty vector control plasmid were used from pBI121 to perform transformation of *Agrobacterium tumefaciens* strain LBA4404 by a freeze-thaw method (Chen *et al*. 1994). *Arabidopsis thaliana* was leaf-disc-transformed (Van *et al*. 1998) by kanamycin selection. Then, the rooted plantlets were cultivated in 1/2 MS culture medium under germ-free conditions and subsequently transferred to and grown in the soil of a greenhouse. Transgenic and control plants for the separate experiments grew together under identical environmental conditions and were finally harvested together at the same stages.

C4H catalyzes the regiospecific *para*-hydroxylation of *trans*-cinnamic acid resulting in the formation of *p*-coumaric acid. The content of *p*-coumaric acid was measured by HPLC (Liu *et al*. 2006). All samples were analyzed in triplicate.

Analysis for determination of the anthocyanin content was conducted following a previously reported method (Meng *et al.* 2004). The content of anthocyanin was established by calculation of A_{530} - $0.25A_{657}$. The results were presented as $A \cdot g^{-1} \cdot FW$, where FW is the fresh weight. All samples were analyzed in triplicate.

Results and Discussion

Using specific primers (ORF-F and ORF-R), full-length, 2063-bp-long cDNA of *DfC4H* were obtained. The cDNA clone contained a canonical polyadenylation signal sequence AATAAA, 437-bp 3′-terminal UTR, poly (A) tail, and 99-bp 5′-terminal UTR (Fig. 1). The length of the ORF portion of the *DfC4H* sequence was 1527 bp and encoded a 508-amino-acid-residue-long polypeptide with an isoelectric point (pI) value of 8.92 and a calculated molecular weight of 58.1908 kDa. The ORF was deposited in the GenBank database (Accession Number KF830705.2).

The results of the analysis by the InterProScan online tool revealed that the putative DfC4H protein might possess functional domains (IPR001128 and IPR002401) of cytochrome P450 and a conserved region (IPR017972) of cytochrome P450. In addition, the motif search analysis indicated that the DfC4H protein contains two motifs, a cytochrome P450 signature motif (PF00067) (42 - 494 bp), and the MCM N-terminal area (PF14551) (192 - 263 bp). The predicted amino-acid sequence of *D. fragrans* DfC4H contains functional and conserved domains of cytochrome P450 that are main characteristic features of the cytochrome P450 family and considered typical for the protein cytochrome P450 (Hasemann *et al.* 1995).

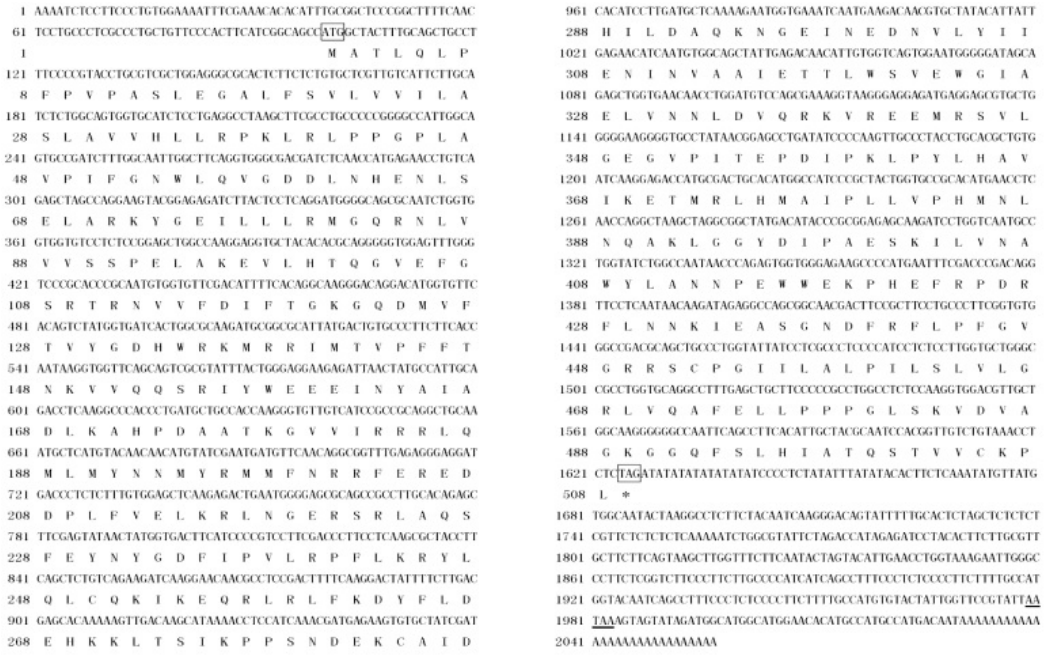

Fig. 1. Deduced amino-acid and nucleotide sequences of *DfC4H* cDNA from *D. fragrans.* Box: initiation codon ATG and termination codon TAG; Underlined: tail signal AATAA.

The results showed that the *DfC4H* gene was potentially capable for transcription into mRNA of *D. fragrans.* Until now, *C4H* genes have been found to exist as small gene families in a range of plants and have been successfully cloned. Four genes homologous of *C4H* were identified in

Populus kitakamiensis and *P. tremuloides* (Kawai *et al*. 1996, Lu *et al.* 2006). Nevertheless, just a single *C4H* gene was detected in the *Pisum*, *Arabidopsis*, and *Parthenocissus* genomes (Frank *et al.* 1996, Bell-Lelong *et al.* 1997, Liu *et al.* 2009). Therefore, various copy numbers of *C4H* exist in different plant species. The copy number of *C4H* in *D. fragrans* will be studied.

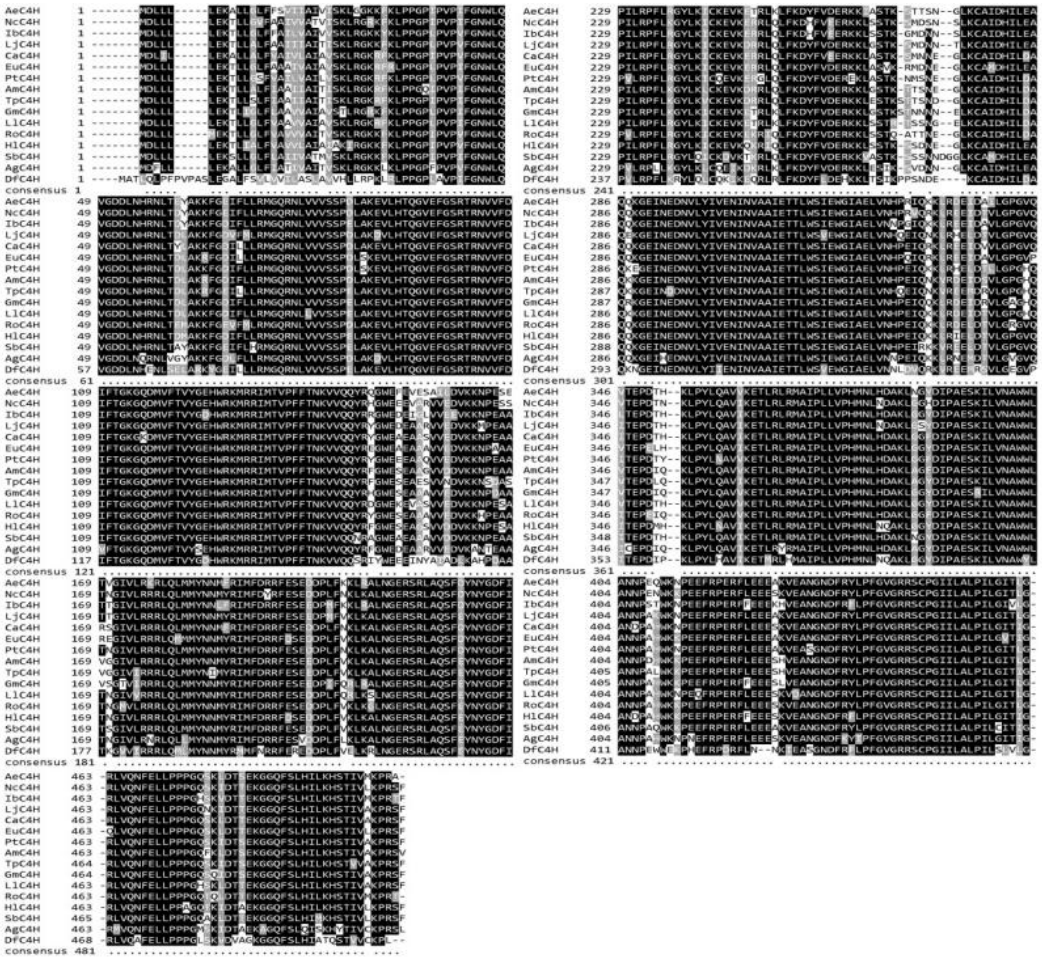

Fig. 2. Multiple amino-acid sequences aligned to DfC4H and C4Hs of other plants. GenBank accession numbers and sources of C4Hs are as follows (black shading indicates identical amino acids): AeC4H (*Arnebia euchroma*, ABD77493.2); NcC4H (*Neolamarckia cadamba*, AFN44010.1); IbC4H (*Ipomoea batatas*, ADB65927.1); LjC4H (*Lonicera japonica*, AGE10592.1); CaC4H (*Canarium album*, ACR10242.1); EuC4H (*Eucalyptus urophylla*, AGJ71350.1); PtC4H (*Populus tomentosa*, ACE95171.1); AmC4H (*Astragalus membranaceus*, AFD36436.1); TpC4H (*Trifolium pratense*, ACB78014.1); GmC4H (*Glycine max*, ACR44227.1); LlC4H (*Leucaena leucocephala*, AEM63593.1); RoC4H (*Rubus occidentalis*, ACM17896.1); HlC4H (*Humulus lupulus*, ACM69364.1); SbC4H (*Scutellaria baicalensis*, ADN32769.1); AgC4H (*Angelica gigas*, AEA72281.1); DfC4H (*Dryopteris fragrans*, AHI17493.2).

Online BLAST searches and analysis using the NCBI database revealed that the DfC4H protein is quite similar to the homologues of *Canarium album* (72%), *Rubus occidentalis* (73%), *Humulus lupulus* (73%), *Glycine max* (74%), *Scutellaria baicalensis* (75%), *Eucalyptus urophylla*

(76%), and *Neolamarckia cadamba* (78%) (Fig. 2). To further characterize the relationships among DfC4H of various species, alignment of the deduced amino-acid sequences was performed by Clustal W v. 1.83. Then, a phylogenetic tree was constructed (Fig. 3) using MEGA 5.0. Further the phylogenetic tree of DfC4H was divided into the following clusters: monilophytes, moss, Angiospermae, and Gymnospermae. The phylogenetic classification placed *D. fragrans* was between Gymnospermae and *Physcomitrella patens* (ADF28535.1). These results are consistent with previous findings obtained using the partial DfC4H protein sequence characterized previously (Li *et al*. 2015).

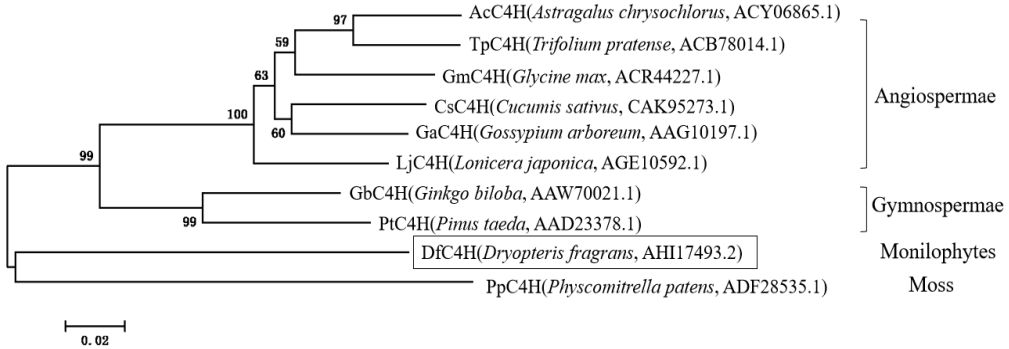

Fig. 3. Phylogenetic tree showing relationships between DfC4H and C4Hs of other plants.

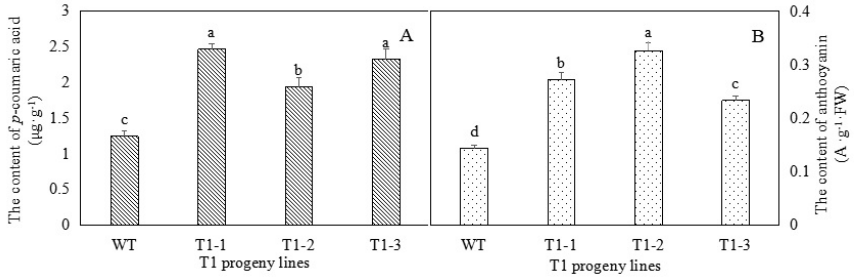

Fig. 4. Determinations of *p*-coumaric acid (A) and anthocyanin (B) in transgenic plants.

The transgenic plants infected with *Agrobacterium tumefaciens* harboring the *DfC4H* gene were subjected to an initial screening for resistance to 50 mg/l kanamycin, followed by PCR detection. The obtained by tissue culture three kanamycin-resistant plant lines (data not shown) were designated T_0, whereas the progeny of T_0 was denoted as T_1. Then, three lines, referred to as T_1-1, T_1-2, and T_1-3, were selected for content analysis of both the first metabolite of the phenylpropanoid pathway and anthocyanin content. PCR with specific primers (ORF-F and ORF-R), utilized for the amplification, was employed to confirm the T_1 plants resistant to kanamycin. The kanamycin-resistant plants yielded a band with a length of 1527 bp with a size corresponding to that of the *DfC4H* gene product was obtained from *D. fragrans*; in contrast, no band was obtained in the WT plants (data not shown).

The overexpression of biosynthetic genes, including *SbPALs*, *SbC4H*, *Sb4CL*, and *SbCHS*, may increase the secondary metabolite content of this plant (Xu *et al*. 2010). T_0 further assess the involvement of the *DfC4H* gene in the phenylpropanoid pathway, the content of the first metabolite (*p*-coumaric acid) in the transgenic and WT plants was determined. A significantly

higher quantity of *p*-coumaric acid (Fig. 4A) in the T_1-1, T_1-2, and T_1-3 plants than in the WT plants ($p < 0.05$) was observed. No considerable differences were detected between T_1-1 and T_1-3 plants. T_0 determine the effect of the *DfC4H* gene on the metabolic end products, the anthocyanin content in the WT and transgenic plants was measured. The content of anthocyanin in the T_1-1, T_1-2, and T_1-3 plants was significantly higher ($p < 0.05$) than that in the WT plants (Fig. 4B).

Earlier reports indicated that the promoted expression of *C4H* could lead to an increased formation of secondary plant metabolites. *C4H* was overexpressed in the roots of *Angelica gigas*, where pyranocoumarin was the most abundantly synthesized (Park *et al*. 2010). The overexpression of *C4H* promoted the acetosyringone production in elicited tobacco suspension cells (Blount *et al*. 2002). In the present study, it was discovered that the overexpression of *DfC4H* dramatically augmented the levels of *p*-coumaric acid and anthocyanin in transgenic *Arabidopsis thaliana* (Fig. 4).

In conclusion, here, molecular cloning and characterization of the *C4H* gene of *D. fragrans*, designated *DfC4H* was reported. Sequence comparison analysis via deduced amino acids of DfC4H was conducted, the results of which showed high homology to C4Hs of other plant species. It was also found that the overexpression of *DfC4H* increased the content of both the first metabolite (*p*-coumaric acid) and the metabolic end-product (anthocyanin) in transgenic *Arabidopsis thaliana.* Although the results of this investigation show that the expression of *DfC4H* is critically involved in the phenylpropanoid pathway, a more comprehensive understanding of its precise roles requires further studies.

Acknowledgements

The authors express their gratitude for the financial support provided by the National Natural Science Foundation of China (Project 31570189), Heilongjiang Postdoctoral Science Foundation (CN) (LBH-Z17203), Heilongjiang Province Science Foundation for Youths (CN) (QC2018027), and The science foundation of HAAS (2019JJPY003).

References

Achnine L, Blancaflor EB, Rasmussen S and Dixon RA 2004. Colocalization of L-phenylalanine ammonia-lyase and cinnamate 4-hydroxylase for metabolic channeling in phenylpropanoid biosynthesis. Plant Cell **16**: 3098-3109.

Ao Z and Li G 1990. Ferns of Heilongjiang Province. Northeast Forestry University Press, Harbin, p. 154.

Bell-Lelong DA, Cusamano JC, Meyer K and Chapple C 1997. Cinnamate-4-hydroxylase expression in *Arabidopsis.* Regulation in response to development and the environment. Plant Physiol. **113**: 729-738.

Blount JW, Masoud S, Sumner LW, Huhman D and Dixon RA 2002. Overexpression of cinnamate 4-hydroxylase leads to increased accumulation of acetosyringone in elicited tobacco cell-suspension cultures. Planta **214**: 902-910.

Chang Y 2009. Progress on research of *Dryopteris fragrans* (L.) Schott in domestic and aboard. Northern Horticulture **4**: 113-115.

Chen H, Nelson RS and Sherwood JL 1994. Enhanced recovery of transformants of *Agrobacterium tumefaciens* after freeze-thaw transformation and drug selection. Biotechniques **16**: 664-670.

Frank MR, Deyneka JM and Schuler MA 1996. Cloning of wound-induced cytochrome P450 monooxygenases expressed in pea. Plant Physiol. **110**: 1035-1046.

Hasemann CA, Kurumbail RG, Boddupalli SS, Peterson JA and Deisenhofer J 1995. Structure and function of cytochromes P450: a comparative analysis of three crystal structures. Structure **3**: 41-62.

Huang Q, Li W, Fan R and Chang Y 2014. New MADS-box gene in fern: cloning and expression analysis of *DfMADS1*from *Dryopteris fragrans*. PLoS One **9**: e86349.

Kawai S, Mori A, Shiokawa T, Kajita S, Katayama Y and Morohoshi N 1996. Isolation and analysis of cinnamic acid 4-hyroxylase homologous genes from a hybrid aspen, *Populus kitakamiensis.* Biosci. Biotechnol. Biochem. **60**: 1586-1597.

Li Y, Sun LL, Wang HM, Gao R, Zhang JZ, Hu BZ and Chang Y 2015. Cloning and expression analysis of phenylalanine ammonia-Lyase (PAL) gene family and cinnamate 4-hydroxylase (C4H) from *Dryopteris Fragrans*. Biologia **70** (5): 606-614.

Liu S, Hu Y, Wang X, Han L, Song S, Cheng H and Lin Z 2009. Isolation and characterization of a gene encoding cinnamate 4-hydroxylase from *Parthenocissus henryana*. Mol. Biol. Rep. **36**: 1605-1610.

Liu SS, Jin XQ and Gao D 2006. Study on extraction methods of *p*-coumaric acid in *Oldenlandia diffusa* (willd.) Roxb and determine the content by HPLC. Special Wild Economic Animal and Plant Research **1**: 21-24.

Lu Z and Ma K 2006. Spread of the exotic croftonweed (*Eupatorium adenophorum*) across southwest China along roads and streams. Weed Sci. **54**: 1068-1072.

Meng XC and Wang XJ 2004. Regulation of flower development and anthocyanin accumulation in *Gerbera hybrida.* J. Hort. Sci. Biotech. **79**: 131-137.

Park JH, Park NI, Xu H and Park SU 2010. Cloning and characterization of phenylalanine ammonia-lyase and cinnamate 4-hydroxylase and pyranocoumarin biosynthesis in *Angelica gigas.* J. Nat. Prod. **73**: 1394-1397.

Sambrook J, Fritsch EF and Maniatis T 1989. Molecular Cloning. A Laboratory Manual, Ed 2. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, p. 9.

Sun LL, Li Y, Li SS, Wu XJ, Hu BZ and Chang Y 2014. Identification and characterisation of DfCHS, a chalcone synthase gene regulated by temperature and ultraviolet in *Dryopteris fragrans*. Cell. Mol. Biol. **60** (6): 1-7.

Teutsch HG, Hasenfratz MP, Lesot A, Stoltz C, Garnier JM, Jeltsch JM, Durst F and Werck-Reichhart D 1993. Isolation and sequence of a cDNA encoding the Jerusalem artichoke cinnamate 4-hydroxylase, a major plant cytochrome P450 involved in the general phenylpropanoid pathway. Proc. Natl. Acad. Sci. USA **90**: 4102-4106.

Van der Graaff E and Hooykaas PJ 1998. Transformation of *Arabidopsis thaliana* C24 leaf discs by *Agrobacterium tumefaciens*. Methods in molecular biology (Clifton, N.J.) 245-258.

Wang HR, Wang R, Xu HL and Zhang YL 2008. Analgesic effect of *Dryoperis fragrans* (L.) Schott. Journal of Qiqihar Medical College **29**: 2443-2444.

Widén CJ, Fraser-Jenkins CR, Reichstein T and Sarvela J 2001. A survey of phenolic compounds in *Dryopteris* and related fern genera. Annales Botanici. Fennici. **38**: 99-138.

Winkel-Shirley B. 1999. Evidence for enzyme complexes in the phenylpropanoid and flavonoid pathways. Physiol. Plant **107**: 142-149.

Xu H, Park NI, Li XH, Kim YK, Lee SY and Park SU 2010. Molecular cloning and characterization of phenylalanine ammonia-lyase, cinnamate 4-hydroxylase and genes involved in flavone biosynthesis in *Scutellaria baicalensis*. Bioresource Technology **101**: 9715-9722.

(Manuscript received on 14 June, 2019; revised on 16 August, 2019)