ACADEMIA JOURNAL OF BIOLOGY 2023, 45(2): 19–26
DOI: 10.15625/2615-9023/18193

CHARACTERIZATION OF CHALCONE ISOMERASE 4A (*CHI4A*) GENE IN *Pueraria mirifica*

Cao Thi Thu Thuy[3], Nguyen Thi Bich Ngoc[1], Pham Thi Kieu Oanh[4], Cao Thi Thu Dung[5], Huynh Thi Thu Hue[1,2,*]

[1]Institute of Genome Research, VAST, Vietnam
[2]Graduate University of Science and Technology, VAST, Vietnam
[3] Vietnam National University of Agriculture,
Ministry of Agriculture and Rural Development, Vietnam
[4] Pollution Control Department, Ministry of Natural Resource and Environment, Vietnam
[5]Vinh University, Nghe An, Vietnam

Received 27 March 2023; accepted 29 May 2023

ABSTRACT

Pueraria mirifica is a popular medicinal plant and is rich in phytoestrogens, especially isoflavonoids. The chalcone isomerase (*CHI*) gene is well-known as a key factor in biosynthetic pathways such as flavonoid, isoflavonoid, and anthocyanin biosynthesis. Among four *CHI* subfamilies, group IV remained largely uncharacterized in *P. mirifica*. By isolation and gene cloning, the *CHI4A* gene was amplified and sequenced. A dataset that included *CHI4A* genes from isolation, transcriptome, and *Glycine max CHI4A* was constructed for analysis. A total of 31 single nucleotide polymorphism (SNPs) was detected among sequences and 14 SNPs were found in *P. mirifica* subgroups. Thirteen SNPs changed the amino acid sequences of *CHI4A* genes; however, they do not affect core active residues in the protein structure. Two active sites typical for group 4 *CHI* in the Fabaceae family: Ile50 and Tyr106; and one conserved site Asn113, were found in this study, which is similar to the result of the previous study on *CHI4*. A three-dimensional protein model of isolated *P. mirifica CHI4A* enzyme was constructed with a high confidence level. The result provides more information about the variation and protein profile of the *CHI4A* gene for further experiments.

Keywords: Chalcone isomerase, flavonoid, isolation, polymorphisms, *Pueraria mirifica*.

Citation: Cao Thi Thu Thuy, Nguyen Thi Bich Ngoc, Pham Thi Kieu Oanh, Cao Thi Thu Dung, Huynh Thi Thu Hue, 2023. Characterization of chalcone isomerase 4A (*CHI4A*) gene in *Pueraria mirifica*. *Academia Journal of Biology*, 45(2): 19–26. https://doi.org/10.15625/2615-9023/18193

[*]*Corresponding author email*: hthue@igr.ac.vn

INTRODUCTION

Pueraria candollei var. *mirifica* (shortened as *Pueraria mirifica*) belongs to the genus *Pueraria* of the Fabaceae family and has been widely known as a traditional medicinal herb (Malaivijitnond, 2012). Due to their high bioactive accumulation, *P. mirifica* is useful for multiple areas such as the food, cosmetic, and pharmaceutical industries (Wang et al., 2020). Research on plants' biological activities supports that numerous phytoestrogen constituents in *P. mirifica* are significantly beneficial to human health. Especially, flavonoids have antioxidant, anti-inflammatory, and pharmacological properties, which help in lowering blood lipid and sugar levels (Karabin et al., 2015). In plants, flavonoids are known as crucial secondary metabolites that improve tolerance to abiotic stress and participate in growth regulation, defense against pathogens, and pigment synthesis (Jiang et al., 2016).

The chalcone isomerase (*CHI*) gene encodes for the *CHI* enzyme, a key branch-point enzyme in the biosynthesis of flavonoids and isoflavonoids in the phenylpropanoid pathway. By evolution, *CHIs* are divided into four subfamilies depending on the substrate and structure of the protein. Each *CHI* group has a difference in biological function, in particular, the first two subfamilies (*CHI1*, *CHI2*) are able to convert chalcone into flavanone (Yin et al., 2019). In 2005, Ralston et al. identified two other *CHIs* (type III and type IV) in soybeans but their catalytic activity was not the same as *CHI1* and *CH2*. Ngaki et al. (2012) proved that *CHI* group 3 (*CHI3*) catalyzed fatty acid metabolism and then named them the fatty acid binding protein (FAP). In the evolutionary timeline, the secondary structure elements surrounding the fatty acid pocket underwent another shift, giving rise to CHI-like (*CHIL*) enzymes and they are named group IV *CHI* (Dastmalchi & Dhaubhadel, 2015).

Up to the present, there are two members of the *CHI4* subfamily in *Glycine max* were detected (Dastmalchi & Dhaubhadel, 2015) and analyzed in terms of sequence and phylogeny. Genomic studies reported a high genetic relationship similarity between *G. max* and *P. mirifica*. *G. max* genomic profile is a reliable reference in research of *Pueraria* plants transcriptome. This study reported that the *CHI4A* gene was isolated from *P. mirifica* to validate the *CHI4A* unigenes obtained from the *P. mirifica* transcriptomic database (data not shown). This work provides more information about the variation and protein profile of the *CHI4A* gene for further experiments.

MATERIALS AND METHODS

Sampling and RNA extraction

The tubes of approximately 2-year-old *P. mirifica* plants were collected from the experiment garden in Nghe An province. The sampled tubes were preserved in the RNAlater Stabilization Solution and stored at -80 ^{o}C until used. Total RNA was isolated using the TRIzol method. The total extracted RNA was diluted in DEPC water. The quality and quantity of RNA were analyzed at OD 260 nm.

Isolating and sequencing of *Pueraria mirifica CHI4A* genes

The Revert Aid Reverse Transcriptase Kit (ThermoFisher Scientific) was used to synthesize cDNA from purified RNA. A specific primer pair was designed based on the *GmCHI4A* sequence reported previously in Genbank (accession code NM001249853.2) and the sequence of *CHI4A* annotated unigenes (Table 1). *P. mirifica CHI4A* gene was isolated accordingly by RT-PCR.

The RT-PCR reaction mixture contained 12.5 µL Thermo Scientific 2X DreamTaq PCR Master mix, 1 µL of cDNA template, 1 µL of 10 µM each primer, and 9.5 µL ddH_2O in a total 25 µL reaction volume. The amplification condition was designed as follows: initial denaturation at 94 ^{o}C for 3 min; 30 cycles at 94 ^{o}C for 1 min; 58 ^{o}C for 30 min, 72 ^{o}C for 45 s; final extension at 72 ^{o}C for 5 min and hold at 4 ^{o}C. The product was checked by gel electrophoresis running in 0.8% agarose gel, 150 V, 300 A for 30 min. The nucleotide composition of the isolated

CHI4A gene was sequenced by the Sanger sequencing method, from a DNA recombinant vector using pJET1.2 forward and reverse primers (5'- CGACTCACTATAGGGAGAG CGGC-3' and 5'- AAGAACATCGATTTTC CATGGCAG -3', respectively).

Table 1. RT-PCR primers and conditions used for isolation of *CHI4A* gene from cDNA template

Primer name	Primer sequence	Tm (^{o}C)	Product length
CHI4A–F	5'-ACCCTCTTTGTCGAAACCTCT-3'	55.9	833 bp
CHI4A–R	5'-GTTCAAAAGCACAAACCCGC-3'	55.7	

Data analysis and protein modeling

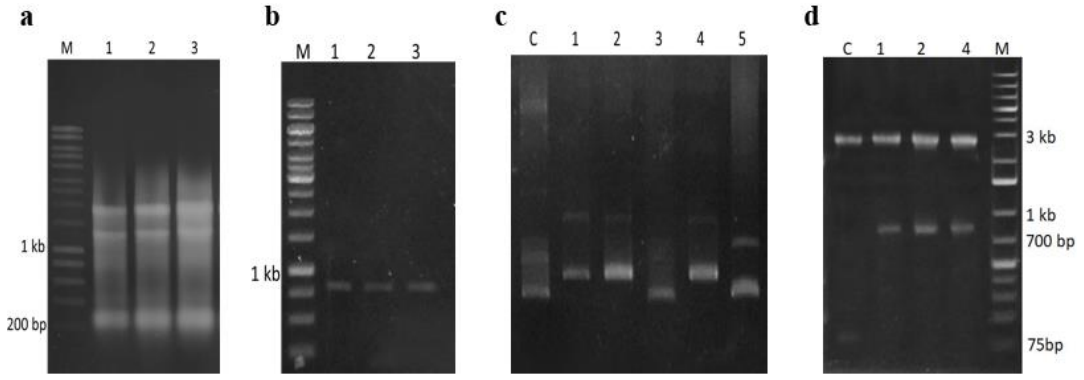
The obtained sequences were analyzed and aligned with Genbank's reference sequence (NM001249853.2) of *G. max* to identify the coding sequence. We constructed a *CHI4A* dataset consisting of the *CHI4A* sequences of *P. mirifica* by cloning, two unigenes annotated for *CHI4A* from the transcriptome database (TRINITY_DN39200, TRINITY_DN67223) (data not shown), and a sequence of *G. max*. The deduced amino acid sequence of *CHI4A* was translated using the ExPASy tool (https://web.expasy.org/translate/). MEGA version 11 software was used to process the Multiple Sequences Alignment (MSA) with input was this *CHI4A* dataset. The position and number of polymorphic sites were analyzed with DnaSP version 6.12.03.

Consequently, the conserved domains and motifs of *P. mirifica CHI4A* were predicted by using CDD software (https://www.ncbi.nl-m.nih.gov/cdd/) and motif finder software (https://www.genome.jp/). SWISS-MODEL software (https://swissmodel.expasy.org/) was used to construct a three-dimensional model for *P. mirifica* CHI4A protein based on its sequenced CDS.

RESULTS AND DISCUSSION

RNA extraction and isolation of *Pueraria mirifica CHI4A* gene

The total extracted RNA was sufficiently extracted, resulting in clear bands, and used as the template for cDNA synthesis (Fig. 1a). RT-PCR amplification product of the *CHI4A* coding sequence (CDS) region was a distinct band with approximately 700 bp in size (Fig. 1b), which is fitted with the expected length according to *CHI4A* CDS reference gene in *G. max* (NM001249853.2). The optimal RT-PCR annealing temperature was 58 ^{o}C for 30 cycles in 30 s. By contrast, there were no bands observed at other temperatures. The PCR product was used for the cloning experiment.



Figures 1. a) Total RNA extraction; b) RT-PCR amplification on *CHI4A* gene in *Pueraria mirifica*; c) Gel electrophoresis of recombinant plasmid extraction, c: control, 1, 2, 3, 4, 5: The randomly selected colonies for plasmid pJET1.2+CHI4A; d) Gel electrophoresis test for cutting plasmid pJET1.2+CHI4A result in a band approximately 3 kb for vector and band more than 800 bp for *CHI4A* gene (1,2,4), (c): vector pJET1.2 cut by B*gl*II restriction enzyme

CHI4A coding sequence was inserted into the pJET1.2 vector and then successfully transformed into competent cell DH5α. After screening in LB agar medium containing ampicillin, five colonies were selected for further studies, labeled from 1 to 5. To check the recombinant vectors, these plasmids were extracted. Figure 1c showed that colonies 1, 2, and 4 consisted of a recombinant plasmid, revealing a higher band than the control vector. The plasmids were cut by B*gl*II restriction enzyme and shown DNA bands with precise size (Fig. 1d). Therefore, plasmids 1, 2, and 4 were used for sequencing and analyzing.

Sequencing and polymorphism detection

Table 2. Variations of the *CHI4A* coding sequence of isolated *Pueraria mirifica*, two unigenes from *Pueraria mirifica* transcriptome, *Glycine max*

Variation description	Variation	Gene sequence				Amino Acid
		Pueraria mirifica CHI4A	TRINITY_ DN39200	TRINITY_ DN67223	*Glycine max CHI4A*	
Non-synonymous	c.59A>C	A	C	C	C	p.Asn20Thr
Synonymous	c.73C>T	C	C	C	T	Leu
Synonymous	c.75G>T	G	T	G	G	Leu
Synonymous	c.78C>T	T	T	T	C	Leu
Synonymous	c.81T>C	T	T	T	C	Gly
Non-synonymous	c.101A>T	A	T	A	A	p.Glu34Val
Non-synonymous	c.102G>T	G	T	G	G	p.Glu34Val
Synonymous	c.105C>T	C	C	C	T	Ile
Synonymous	c.111T>C	T	T	T	C	Phe
Synonymous	c.153T>A	T	T	T	A	Pro
Non-synonymous	c.172C>G	C	C	C	G	p.Gln58Asp
Non-synonymous	c.174G>C	G	G	G	C	p.Gln58Asp
Synonymous	c.180T>C	T	C	T	C	Phe
Non-synonymous	c.210G>T	G	G	G	T	p.Glu70Asp
Non-synonymous	c.216T>A	T	T	T	A	p.Asp72Glu
Synonymous	c.225T>C	T	T	T	C	Phe
Synonymous	c.243T>C	T	T	T	C	Ala
Synonymous	c.245G>T	G	T	G	G	Pro
Non-synonymous	c.319T>G	T	G	T	G	p.Ser107Ala
Non-synonymous	c.321T>G	T	T	T	G	p.Ser107Ala
Synonymous	c.393G>A	G	A	G	A	Glu
Synonymous	c.396C>T	C	T	T	T	Phe
Synonymous	c.405C>T	C	T	C	C	Ser
Non-synonymous	c.413C>T	C	T	T	T	p.Ser138Phe
Non-synonymous	c.422A>T	A	A	A	T	p.His141Leu
Synonymous	c.462A>T	A	A	A	T	Thr
Non-synonymous	c.511G>A	G	G	G	A	p.Val171Ile
Synonymous	c.570A>G	A	G	A	G	Ala
Synonymous	c.573G>T	G	G	G	T	Val
Non-synonymous	c.608C>T	C	T	T	T	p.Ser203Phe
Synonymous	c.612G>A	G	A	G	G	Ser

Notes: CHI4A *Pueraria mirifica* cloned gene, TRINITY DN67223 and TRINITY DN39200 two unigenes of *Pueraria mirifica* transcriptomic database, and *Glycine max* CHI4A gene (NM001249853.2).

A full-length *CHI4A* gene of 630 bp comprising an ORF was sequenced from the *CHI4A* plasmids, encoding 210 amino acids. The MSA result revealed *P. mirifica* cloned *CHI4A* nucleotide sequence has a higher similarity with two unigenes annotated from *P. mirifica* transcriptomic database than the *G. max CHI4A* gene. In particular, it was 99.4% and 97.78% genetic similarity with the unigenes DN67223 and DN39200 respectively while it was 96.03% compared with the *CHI4A* gene in *G. max*.

The sequence alignment showed 31 polymorphic sites, including 18 synonymous single nucleotide polymorphisms (SNPs), and 13 non-synonymous SNPs. Within the *P. mirifica CHI4A* group, there were 14 SNPs consisting of six non-synonymous SNPs (c.59A>C, c.101A>T, c.102G>T, c.319T>G, c.413C>T, c.608C>T), which means they cause the difference in amino acid chains. The consequences of these variations were confirmed by screening the conserved motifs in the *CHI4A* protein. It is notable that one of two substitutions between two unigene sequences is also detected in the isolated *P. mirifica* sequence.

Analysis of conserved motif and catalytic residues

The amino acid chain of *P. mirifica* isolated *CHI4A* sequence was analyzed using the Motif finder tool (GenomeNet), with both Pfam and NCBI-CDD as the database sources. Accordingly, a total of six non-significant predicted motifs was found in *P. mirifica CHI4A*: PLN02804, PF02431 (pfam02431), PLN02311, PLN02559, smart00189 (PF00715), PF16035 (pfam16035). Throughout evolution, the type IV *CHIs* subfamily was formed by the CHI-like enzymes (Ralston et al., 2005). After being given rise from type III *CHIs*, these proteins lack the chalcone-to-flavanone catalytic ability and active site residue conservation (Ngaki et al., 2012).

For the chalcone isomerase mechanism, the role of hydrogen bonds is a crucial factor in catalysis (Jez et al., 2002), several conserved regions were mentioned in previously published articles, such as the active sites in alfalfa including Arg-36, Gly-37, Leu-38; Phe-47, Thr-48, Ile-50; Tyr-106, Lys-109, Val-110, Asn-113; and Thr-190, Met-191 (Jez et al., 2000). We aligned the amino acid sequences of the *CHI4A* dataset with other members of the Fabaceae family and screened the conserved and active sites among sequences. Only Tyr106 - one of four core residues in the *CHI* family and Asn113 - conserved sites were detected in all four sequences. This is similar to the report from Dastmalchi & Dhaubhadel (2015) about the evolution of *CHI* subfamily 4 in *G. max*. Regarding the three active sites typical for group 4 *CHIs* of Fabaceae (Lin et al., 2021), Ile50 and Tyr106 were found in our sequences. As mentioned, the *CHI4A* in *P. mirifica* has a high similarity with the *CHI4* subfamily of *G. max*. When comparing with other reports about *CHI* subfamily 4, our MSA outcome proved that *P. mirifica CHI4A* has substitutions at several catalytic residues with the other three *CHI* subfamily. It is interesting to note while there were 31 SNPs in the dataset, none of them affect the Tyr106 core residues in *CHI4A*.

Construction of the three-dimensional *CHI* protein model

Figure 3 illustrated the deduced three-dimensional protein model for cloned *CHI4A* gene. The model was conducted based on *Arabidopsis thaliana* which is a crystal structure of chalcone isomerase from *A. thaliana* chalcone-isomerase-like protein At5g05270 (AtCHIL), which is a typical presenter for CHI4 subfamily (Lin et al., 2021) (Table 3). The monomer of 224 amino acid residues was solved by X-ray crystallography with a pairwise sequence identity of 63.01% and a high resolution of 1.7 Å. To this aim, deduced protein building relied on structural homology modeling and evaluated using Global Model Quality Estimate (GMQE) as a number between 0 and 1, and Quaternary Structure Quality Estimate (QSQE) score should be around zero (or > -4) (Table 3). In general, a high GMQE score of 0.88 is expressed indicates higher reliability

and accuracy of the resulting, with 0.98 the coverage of the target sequence. In terms of QMEAN, the value 0.81 is calculated indicating good consistency between the model structure and experimental structures of similar size.

There was a quite high number of SNPs in nucleotide sequence and amino acid sequences of isolated *CHI4A*, however, they do not affect the core active residue of the enzyme. Therefore, the predicted model quality estimation has significantly contributed to the understanding and narrowed the gap between known protein sequences and experimentally determined structures.

Our research proved that there was a significant number of variations in the *CHI4A* gene sequenced by the next generation sequencing method (transcriptomic) and Sanger sequencing. However, these variations do not affect the core active sites and conserved regions in the *CHI4A* amino acid sequence. Hence, with the deduced amino acid chain, its constructed three-dimensional protein model has a high confidence level. This work provided more knowledge for further studies in the function of *CHI4A* in *P. mirifica*.

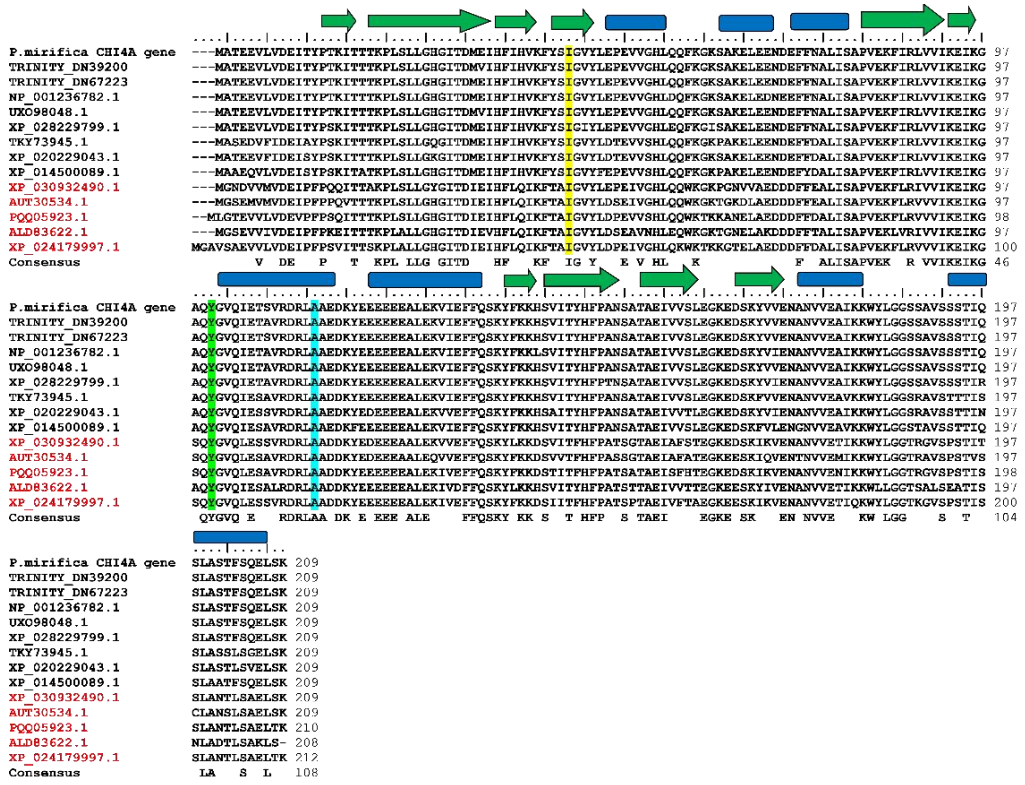


Figure 2. Sequence alignment of *CHI4A* amino acid sequence of isolated *Pueraria mirifica*, two unigenes from *Pueraria mirifica* transcriptome, *Glycine max* and other members in the Fabaceae family. The yellow, green, and blue shade indicates the conserved active site residues in *Pueraria mirifica CHI4A*. The blue rectangular and green arrow indicates for alpha-helix and beta-sheet, in that order. UXO98048.1: *Pueraria montana* var. *lobata*; NP_001236782.1: *Glycine max*; XP_028229799.1: *Glycine soja*; TKY73945.1: *Spatholobus suberectus*; XP_020229043.1: *Cajanus cajan*; XP_014500089.1: *Vigna radiata* var. *radiata*; RDX69997.1: *Mucuna pruriens*; XP_030932490.1: *Quercus lobata*; AUT30534.1: *Camellia fraternal*; PQQ05923.1: *Prunus yedoensis* var. *nudiflora*; ALD83622.1: *Morus alba*; XP_024179997.1: *Rosa chinensis*

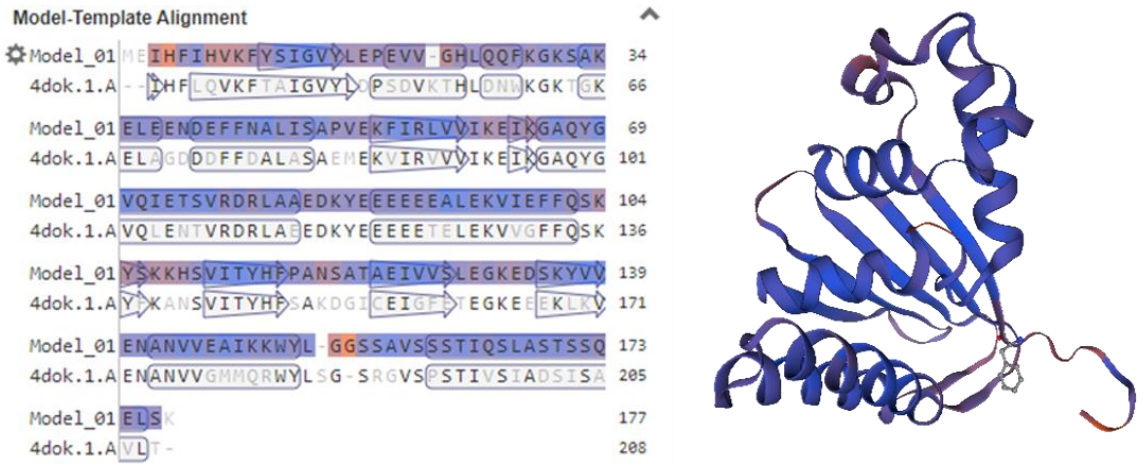


Figure 3. Three-dimensional structure of cloned *Pueraria mirifica CHI4A* gene. The rounded rectangles show the alpha-helix, while the right arrows represent the beta-pleated sheets. The dark blue on the sequence area indicates the highest reliability, which was followed by the lighter blue and the red, in that order

Table 3. The model quality estimation parameters

Template	Seq identity	Oligo-sate	Method	Coverage	GMQE	QSQE	QMEAN	Description
4dok.1.A	63.01%	monomer	X-ray, 1.7Å	0.98	0.88	0.00	0.81	Similarity to chalcone-flavone isomerase Crystal structure of *Arabidopsis thaliana* chalcone- isomerase like protein At5g05270 (AtCHIL)

CONCLUSION

In summary, the *CHI4A* gene from *P. mirifica* has been identified and analyzed in the length of 630 bp encoding 210 amino acids. The gene has 31 polymorphic sites, including 18 synonymous single nucleotide polymorphisms (SNPs), and 13 non-synonymous SNPs. Two active sites are typical for group 4 *CHIs* of Fabaceae - Ile50 and Tyr106 and a conserved site - Asn113, were found in the sequence. Among 31 SNPs in the dataset, none of them affects the Tyr106 core residues in *CHI4A*. The deduced *CHI4A* protein model was constructed showing many typical three-dimensional structures same as the model *A. thaliana CHI4A* protein.

Acknowledgements: This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 106.02-2019.13.

REFERENCES

Dastmalchi M. & Dhaubhadel S., 2015. Soybean chalcone isomerase: evolution of the fold, and the differential expression and localization of the gene family. *Planta*, 241(2): 507–523. https://doi.org/10.1007/s00425-014-2200-5

Jez J. M., Bowman M. E., Dixon R. A. & Noel J. P., 2000. Structure and mechanism of the evolutionarily unique plant enzyme chalcone isomerase. *Nature Structural Biology*, 7(9): 786–791. https://doi.org/10.1038/79025

Jez J. M., Bowman M. E. & Noel J. P., 2002. Role of Hydrogen Bonds in the Reaction Mechanism of Chalcone Isomerase.

Biochemistry, 41(16): 5168–5176. https://doi.org/10.1021/bi0255266

Jiang N., Doseff A. & Grotewold E., 2016. Flavones: From Biosynthesis to Health Benefits. *Plants*, 5(2): 27. https://doi.org/10.3390/plants5020027

Karabin M., Hudcova T., Jelinek L., & Dostalek P., 2015. Biotransformations and biological activities of hop flavonoids. *Biotechnology Advances*, 33(6): 1063–1090. doi; 10.1016/j.biotechadv.2015.02.009

Lin L. M., Guo H. Y., Song X., Zhang D. D., Long Y. H. & Xing Z. B., 2021. Adaptive Evolution of Chalcone Isomerase Superfamily in Fagaceae. *Biochemical Genetics*, 59(2): 491–505. https://doi.org/10.1007/s10528-020-10012-z

Malaivijitnond S., 2012. Medical applications of phytoestrogens from the Thai herb Pueraria mirifica. *Frontiers of Medicine*, 6(1): 8–21. https://doi.org/10.1007/s11684-012-0184-8

Ngaki M. N., Louie G. V., Philippe R. N., Manning G., Pojer F., Bowman M. E., Li L., Larsen E., Wurtele E. S. & Noel J. P., 2012. Evolution of the chalcone-isomerase fold from fatty-acid binding to stereospecific catalysis. *Nature*, 485(7399): 530–533. https://doi.org/10.1038/nature11009

Ralston L., Subramanian S., Matsuno M. & Yu O., 2005. Partial Reconstruction of Flavonoid and Isoflavonoid Biosynthesis in Yeast Using Soybean Type I and Type II Chalcone Isomerases. *Plant Physiology*, 137(4): 1375–1388. https://doi.org/10.1104/pp.104.054502

Wang S., Zhang S., Wang S., Gao P. & Dai L., 2020. A comprehensive review on Pueraria: Insights on its chemistry and medicinal value. *Biomedicine & Pharmacotherapy*, 131: 110734. doi: 10.1016/j.biopha.2020.110734

Yin Y., Zhang X., Gao Z., Hu T. & Liu Y., 2019. The Research Progress of Chalcone Isomerase (*CHI*) in Plants. *Molecular Biotechnology*, 61(1): 32–52. https://doi.org/10.1007/s12033-018-0130-3