## STUDY ON CHARACTERIZATION OF CHALCONE SYNTHASE GENE FROM Pueraria lobata AND Pueraria mirifica IN VIETNAM

### Huynh Thi Thu Hue<sup>1,2,\*</sup>, Nguyen Minh Phuong<sup>3</sup>, Nguyen Xuan Canh<sup>3</sup>

<sup>1</sup>Institute of Genome Research, VAST, Vietnam <sup>2</sup>Graduate University of Science and Technology, VAST, Vietnam <sup>3</sup>Vietnam National University of Agriculture, Ha Noi, Vietnam

Received 17 December 2020, accepted 18 July 2021

### ABSTRACT

Two species of genus Pueraria ((Pueraria lobata (synonym: Pueraria montana var. lobata) and Pueraria mirifica (synonym: Pueraria candollei var. mirifica)) are traditional plants used in medicine since ancient times. These plants have been used and became commercially crucial indigenous medicinal plants. Currently, both roots and flowers of P. mirifica are used as a dietary supplement and functional food for women because of their rich source of phytoestrogen and nutrition. However, little information of genes on both species of Pueraria genus (P. lobata and P. mirifica) are known in Vietnam. The purpose of this research is to support more understanding about Chalcone synthase (CHS) genes by determining and sequence analyzing an encoding region of CHS genes that were isolated from P. lobata and P. mirifica. The full-length open reading frame (ORF) sequence CHS was identified with 1170 bp which encodes 389 amino acids by Sanger sequencing. The isolated CHS gene of P. lobata has no difference in sequence with CHS reported on GenBank (D10223.1), whereas a difference of 26 nucleotide positions in CHS sequence of P. mirifica compared with the published gene sequence (JQ409456.1) as consequential having 97.78% genetic similarity. The CHS genes sequence of P. lobata and P. mirifica are homologous with 98.4% because of having 19 nucleotide differences. Chalconestilbene synthase N-C terminal, PLN03173, CHS-like, BH0617, fabH are some important domains predicting the CHS genes. Especially, the family signature 'GVLFGFGPGLTI' motif of CHS gene as a part of the active-site scaffold contributes to decide the product of cyclization reactions performing the stereochemistry of cyclization which was also observed in P. lobata and P. mirifica, but it was not included for all members in Fabaceae family. With in sillico analysis, the P. lobata and P. mirifica CHS sequences have highly conserved regions to maintain their structure and function, so that it needs further studies to clarify these points.

Keywords: Chalcone synthase, CHS, gene analysis, P. lobata, P. mirifica.

*Citation*: Huynh Thi Thu Hue, Nguyen Minh Phuong, Nguyen Xuan Canh, 2021. Study on characterization of chalcone synthase gene from *Pueraria lobata* and *Pueraria mirifica* in Vietnam. *Academia Journal of Biology*, 43(3): 47–58. https://doi.org/10.15625/2615-9023/15763

<sup>\*</sup>Corresponding author email: hthue@igr.ac.vn

<sup>©2021</sup> Vietnam Academy of Science and Technology (VAST)

## INTRODUCTION

of Two species genus Pueraria ((Pueraria lobata (synonym: Pueraria montana var. lobata) and Pueraria mirifica (synonym: Pueraria candollei var. mirifica)) are widely used as a traditional medicinal herb. It is very useful for multi-purpose and functions because of its high bioactivities accumulation. P. lobata is seemed to be one of the earliest plants listed in traditional Chinese medicine (Keung & Vallee, 1998). The vast majority of Asian countries applying dried *P. lobata* roots as an important ingredient of medicinal kudzu root tea to treat influenza fever, dysentery, etc. Moreover, P. lobata root powder and extract as the single ingredient or in combination with other herbs such as Ginkgo biloba, miltiorrhiza, Salvia Bacopa monnieri, Silvbum marianum, and Salix alba in some countries (the US, the UK, and Australia) are developed for relieving hangover symptoms, fever and flu, improving liver functions and aiding weight loss purposes since these products are fewer effect sides (Healthstore, 2011; TGA, 2011; Cheung et al., 2012). P. mirifica is an indigenous herb, mostly found and applied in Thailand because of its rejuvenating properties or antioxidant activities for beauty benefits in prepostmenopause. menopause and Furthermore, the purposes are similar to modern hormone replacement therapy, good effects on reproductive organs, enhanced breast size enhancement, support for the bone, hair, and fingernails and many other special functions (Wanadorn, 1933; Manonai et al., 2000, Cherdshewasart et al., 2006). In Vietnam, P. lobata is almost used as a summer beverage, nutritional food as well as herbal remedy to treat diarrhoea, muscle stiffness, thirst, and diabetes. To date, both roots and flowers of P. mirifica are synthesized and used as the main ingredients in a dietary supplement and functional food products for women. Nevertheless, these products are considered to be less popular and advanced in the Vietnamese market than in other countries. P. lobata and P.mirifica

are known for the (iso)flavonoid accumulation at a high level. Biological activities, the therapeutic potential of (iso)flavonoid and their health benefits have been determined and confirmed by many previous studies.

(Iso)flavonoids play a crucial role function in plants and also in human health (Birt et al., 2001). Isoflavones such as puerarin or daidzein have the potential to replace estrogenic drugs with steroid frames such as HRT with fewer side effects (Dewick, 2009). Chalcone synthases are pivotal enzyme in the (iso)flavonoid biosynthesis process in plants, which is part of secondary metabolite biosynthesis being studied for its medicinal properties. The chalcone synthase superfamily of type III polyketide synthases (Austin et al., 2003) are primary enzymes that also participate in many important biosynthesis pathways, especially (iso)flavonoids for some prominent roles: antioxidant, antimutagenic, antiproliferative activities. estrogen-like activity, etc (Birt et al., 2001; Miadokova et al., 2002; Ryan-Borchers et al., 2006; Iwasaki et al., 2008; Scarpato et al., 2008). The chalcone synthase are noted to be the most important factor, particularly in flavonoid and isoflavonoid, catalyzing the first committed step in the process. Flavonoids are performed via the phenylpropanoid and polyketide pathway releasing a naringenin chalcone by enzyme chalcone synthase (CHS)(Dao et al., 2011).

In the world, there have been many studies on the Chalcone synthase genes. A series of studies on the characterization of the CHS gene has been published on buckwheat (*Fagopyrum esculentum* M), *Phaseolus vulgaris*, and *Hordeum vulgare* (Hrazdina et al., 1986; Ryder et al., 1987; Rohde et al., 1991). Because of its characteristics, chalcone synthase cDNAs in *Petunia hybrid*, soybean, *Pisum sativum* are carried out for molecular cloning with different research purposes, and fields (Shao et al., 1995, M. & C. Sengupta-Gopalan 1991, Ichinose et al., 1992). However, the characteristics of CHS genes in *Pueraria* plants of Vietnam have rarely been investigated so far. In this research, to characterize of CHS gene of the two species, we isolated and sequenced these genes from tuber's mRNA, its main domains were predicted and compared with references by using bioinformatics tools.

#### MATERIALS AND METHODS

### Materials

One-year tube of the *P. lobata* was collected from Phu Binh, Thai Nguyen, Vietnam. And two-year tube of the *P. mirifica* was collected from Do Luong, Nghe An, Vietnam. Each specimen of the tubes were preserved in the RNA later Reagent. Several reference sequences using in the comparative analysis are listed (Table 1).

No.	Species	AC number (NCBI)
1	Glycine max	NP_001304585.2
2	Cicer arietinum	CAA10131.1
3	Vigna radiata	AJZ72657.1
4	Trifolium pretense	PNY03318.1
5	Lupinus luteus	ABF59866.1
6	Glycyrrhiza uralensis	ABM66532.1
7	Trifolium subterraneum	AAA67701.1
8	Onobrychis viciifolia	AAB81987.1
9	Arachis hypogaea	AAU43217.1
10	Pisum sativum	BAA01512.1
11	Vigna unguiculata	CAA52819.1
12	Mucuna pruriens	RDY14063.1
13	Glycyrrhiza inflate	ACH67480.1
14	Senna tora	ACB78187.1
15	Arachis duranensis	XP_015971138.1
16	Medicago truncatula	KEH27377.1
17	Onobrychis viciifolia	AEF14414.1
18	Acacia confuse	AFA55180.1
19	Vigna unguiculata	QCE15713.1
20	Cajanus cajan	XP_020230031.1
21	Spatholobus suberectus	TKY62964.1
22	Prosopis alba	XP_028780252.1
23	Glycine soja	XP_028189397.1
24	Caragana korshinskii	AYE88587.1

<i>Table 1</i> . Reference sequences in	Table	1. Re	ference	seq	uences	list
---	-------	-------	---------	-----	--------	------

## Methods

#### Total RNA extraction

Total RNA was extracted using TRIzol reagent which has been innovated appropriately with *P. lobata* and *P. mirifica* tubes and the concentration of RNAs was

later determined by using NanoDrop<sup>TM</sup> One (US).

# Amplification of CHS genes by RT-PCR method

Thermo Fisher KIT (Revert Aid Reverse Transcriptase) was used to synthesize the first-strand cDNA from the extracted RNA and amplify the CHS 5' and 3' transcript. The forward primer 5' TTGAGTTCGATCAAAT CGCAG 3', reverse primer 5' TAGGCA TCTCAGATGGCC 3' for the amplification of CHS gene were designed based on the reference sequence in GeneBank (D10223.1), then synthesized and provided by the PHUSA Biochem company. The condition of amplification was optimized for 25 µl of PCR including 1 µl of cDNA, 2.5 µM of each primer, 0.5 unit of Dream Taq polymerase (Thermo Scientific), 1 mM of each dNTP and 2.5 µl of Taq PCR buffer. RT-PCR reaction conducted with the thermal cycling: initial denaturation at 94 °C for 3 min; 30 cycles at 94 °C for 1 min; 59 °C for 30 min, 72 °C for 1 min 20 s; final extension 72 °C for 8 min and hold at 4 °C in a total 25 µL reaction volume. The products of PCR were run on 0.8% agarose electrophoresis gel at 150 V, 300 A for 30 min.

### Sanger sequencing

The amplified products were purified by OMEGA biotek KIT before sequencing with the Big Dye Terminator kit (ABI Foster City, USA) on an Applied Biosystems<sup>TM</sup> 3500 system, and the primers for Sanger sequencing which listed in RT-PCR method were used. After that, the obtained sequences were analyzed and assembled by the BioEdit program. BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) were performed to search the sequence homology. Consequently, the conserved domain and motifs of each amino acid sequence were predicted by using software of NCBI (https://www.ncbi.nlm.nih.gov/Structure/cdd/ wrpsb.cgi) and motif software (https://www.genome.jp/tools/motif/).

### **RESULTS AND DISCUSSION**

# Amplification of CHS genes by using RT-PCR

The total RNA samples extracted from *P.lobata* and *P.mirifica* tubes and used as templates for cDNA synthesis (Fig. 1a). RT-PCR amplification was performed by following the description of Sambrook (Sambrook et al., 2001), with the specific primer and conditions as described above at three different temperatures in order to determine the optimal DNA condition. Finally, the PCR products were purified and checked by 0.8% agarose gel (Fig. 1d).

As the results (Figs. 1b, 1c), DNA distinct bands appeared very clearly, with approximately 1.24 kb at the best annealing temperature 59 °C of 30 cycles for 30 s, but no bands were observed at 61 °C and 63 °C annealing temperature. In addition, the PCR products were purified by using Thermofisher Scientific DNA purification kit, which have more accurate readable DNA sequences. Thus, CHS genes of P. lobata and P. mirifica were successfully amplified with the expected size, which is a prerequisite for sequencing.



*Figure 1*. Agarose Gel Electrophoresis of CHS gene product. a.Result of total RNA extraction;
b. Result of RT-PCR reaction CHS gene of *P. lobata* at different temperatures (59 °C, 61 °C, and 63 °C);
c. Result of RT-PCR reaction CHS gene of *P. mirifica* at different temperatures (59 °C, 61 °C, and 63 °C);
d. Result of purified PCR product. M. Marker 1kb (iNtRON)

# Sequence analysis of CHS gene from *P.lobata* and *P.mirifica*

After sequencing, the isolated CHS genes from *P. lobata* and *P. mirifica* were determined with 1170 bp in length, encoded for 389 amino acids containing a start codon ATG, stop codon TGA and then utilized to compare the two published genes on GenBank (D10223.1-*P. lobata*, JQ409456.1-*P. mirifica*).

The results of Blast analysis showed that the *P. lobata* CHS gene and D10223.1 (*P. lobata*) were obtained with 100% of genetic similarity, and no differences have been found between them. Besides, a pairwise sequence alignment between the *P. mirifica* CHS gene and JQ409456.1 (*P. mirifica*) showed that 97.78% of genetic similarity was identified, with the list of 26 different nucleotide positions. In particular, there were 16 amino acid variants which were appeared more frequently at the 5'end of the coding sequence (Table 2).

According to Dixon et al. (1999) and Pandith et al. (2019), the majority of CHS gene in legume species belonged to multigene families such as 6–8 members in green bean (P. vulgaris), seven in pea, six or seven in P. lobata, at least nine in soybean (G. max), and other species. Especially, Wani concluded that the two isoforms GaCHS1 (an ORF of 1176 bp) and GaCHS2 (an ORF of 1170 bp) (G. asiatica L) have appeared the differential expression pattern at different stages (Wani et al., 2017). Besides that, the *P. mirifica* CHS gene, with 1170 bp in length encoding 389 amino acid residues is also demonstrated its characteristics belonging to a multigene family (Wiriyaampaiwong et al., 2012). In a recent study, Suntichaikamolkul and his colleagues (2019) based on transcriptome and phylogenetic analysis of *P. mirifica* show that two CHS genes (CHS1, CHS2 encoding chalcone synthase) were involved in isoflavone biosynthesis in a total of 14 putative genes (Suntichaikamolkul et al., 2019). As mentioned before, several CHS genes in other plants have been found in different isoforms with many different nucleotides and amino acid positions.

The red circles represent the different nucleotides making different amino acids, while the yellow circles show similar amino acids.

	published on GenBank (JQ409456.1)																										
	Position	47	51	55	57	60	63	69	77	78	81	84	144	281	348	451	452	472	479	507	515	520	523	533	559	627	818
P.mirifica	Nucleotide	С	А	А	С	С	С	А	А	С	А	А	С	G	G	С	G	А	А	G	G	G	А	С	С	Т	G
	Amino acid	P	А	]	[	L	А	G	]	N	Р	Р	Н	R	E	]	R	М	Y	Т	R	A	K	A	V	L	G
10400456 1	Nucleotide	Т	С	G	Т	А	Т	С	С	Т	Т	С	Т	А	С	G	А	С	С	С	С	С	С	Т	Т	А	Α
JQ409456.1	Amino acid	L	А	I	/	L	А	G		Т	Р	Р	Н	K	D	]	D	L	S	Т	P	P	Q	V	V	L	E
V. (., Different emine exidence it and encoded in black hold letters																											

 Table 2. Number of nucleotide differences P. mirifica CHS gene and CHS gene

 published on GenBank (JQ409456.1)

Note: Different amino acid positions are marked in black bold letters.

Table 3. Number	of	nuc	cleo	otid	e s	eq	uenc	ce	differ	ence	es P.	lob	ata	and	<i>P. r</i>	nir	rifica	CHS	gen	es

	Position	29	30	31	33	36	39	43	60	559	621	657	684	786	804	909	1150	1154	1164	1167
Dlohata	Nucleotide	С	А	С	А	G	А	G	Т	Т	Т	G	А	С	С	С	С	G	С	С
<i>F.100aia</i>	Amino acid	A	l	(	)	R	А	G	L	V	L	А	Р	G	L	Н	V	R	А	Ι
P.mirifica	Nucleotide	G	G	А	С	Т	Т	А	С	С	С	А	Т	G	Т	Т	Т	А	Т	А
	Amino acid	(	J	N	N	S	А	S	L	V	L	А	Р	G	L	Н	V	H	А	Ι

The pairwise sequence alignments using BLAST (Fig. 2) illustrated CHS genes from *P. lobata* and *P. mirifica* were shared a genetic similarity of 98.4% and discovered only 7 variants in total 19 different nucleotide positions (Table 3) leading to different amino acids. Furthermore, these differences were observed much higher at the 5' and 3' end of the encoding sequence and may able to create their own characteristics

of *P. mirifica* compared to *P. lobata* species and others. However, different amino acid positions affect CHS activities and modify functions need to be further researched.

			10		20		30		40		50		60
ana	D labata		.	.	.		- MA-			.	.	.	· · ·
спъ	P.IODala	M V S	V	A E	I	R Q		R A	E	G P	A I		L
снѕ	P.mirifica		• • • • •			· · · · · ·	.GGA.	г т <mark>т</mark>	• • • <b>z</b>				<mark>c</mark>
		M V S	V.	A E	I	RQ	_¶,/,₹	SA	E	S P	A I	Ξ	чV
			70		80		90	1	.00		110		120
			.		.					.	.	.	•••
CHS	P.lobata	GCCATTGG	AACTG	CAAAC			TGTGT	<b>FGATCAG</b>	GAGC	CCTAT		TTAC	TAC
снѕ	P.mirifica		· · · · ·	•••••					· · · · ·			· • • • •	· · ·
		A I G	т.	A N	P	P N	C V	DQ	S	T Y	P E	Y (	Y
			130		140		150	1	60		170		180
												.	
снѕ	P.lobata	TTCAGAAT	CACCA	ACAGI	GAGC	ACATG	ACCGAC	GCTCAA/	GAG	AATTC	CAGCG	CATG	TGT
снз	P.mirifica	FRI	т.	N S	E	н м	TE	ь к	E	K F	QF	M	C
		FRI	т	n s	Е	H M	тЕ	LK	Е	K F	QF	м	С
			1 0 0				010	_					0.4.0
					200		210		220		230		240
снѕ	P.lobata	GACAAGTC	TATGA	TCAAG	AAGA	GATAC	ATGTA	TTAACO	GAAG	AGATC	TTGAA	AGAG	AAT
CHE	D minifico	DKS	M	I K	ĸ	R Y	MY	L T	E	ΕI	LK	E	N
спъ	P.miriiica	D K S	м	I K	к	R Y	MY	L T	Е	E I	LK	Е	N
		1	250		260		270	2	280		290		300
снѕ	P.lobata	CCAAACAT	GTGTG	CTTAC		CACCI	TCTTTC	GATGC1			ATGGI	GGTG	GTG
		P N M	C.	A Y	м	A P	S L	D A	R	Q D	MV	v	V
снѕ	P.mirifica	P N M	· · · · ·	 a v	 M	<b>D D</b>	S T.		 P	· · · · ·	M V		· · · ·
		1 14 13		I	1.1		0 1	DI	10	<u> </u>	1.1 0	v	Č.
			310		320		330	3	340		350		360
снз	P.lobata	GAGGTACC				AGGCI	GCAAC						
_		EVP	к	L G	к	ΕA	АТ	K A	I	K E	W G	Q	P
снѕ	P.mirifica		· · · · ·	••••			· · · · ·	· · · · · · · · · · · · · · · · · · ·	•••••••••••••••••••••••••••••••••••••••	·····		•••••	•••
		E V E	IX .		IX.	L A	A 1	K A	-	IX IS	vv	Q	L
			370		380		390	4	100		410		420
CHS	P lobata												
		KSK	I	т н	L	I F	СТ	T S	G	V D	M F	G	A
снз	P.mirifica		•••••	•••••	• • • •	· · · · · ·	••••	••••••	• • • • •	•••••	••••	••••	•••
		K S K	. 1	т н	Ц	T F	C T	TS	G	V D	M	G	A
			430		440		450	4	160		470		480
CHG	P lobata			·   · · ·							.   TACAT		
CIID	I.IOData	D Y Q	L	тĸ	Q	L G	L R	ΡY	v	K R	Y M	I M	Y
снз	P.mirifica		••••••	<u>.</u>	• • • •	· · · · ·	• • • • • •	•••••••••••••••••••••••••••••••••••••••	· · · · ·				• . •
		DĭQ	<u>і</u> Г	TR	Q	L G	LR	Pĭ	~	K K	Y IV	1 141	ĭ
			490		500		510	5	520		530		540
CHS	P lobata		.   	·   · · ·									
CIID	I.IOData	QQG	C	F A	G	G T	V L	R L	A	K D	L A	E	N
CHS	P.mirifica	• • • • • • • •	· · · · ·	<u>.</u>		· · · <u>·</u> ·	· · · · ·	••••••			· · · · ·	· · <u>·</u> ·	· · ·
		Q Q G	C	F, Y	G	G T	V L	RL	A	кD	LA	E	N
			550		560		570	5	580		590		600
CHE				$\cdot \mid \cdot \cdot \cdot$				$\cdot \cdot \mid \cdot \cdot \cdot \cdot$			.	.	
	r.robata	N K G	A	R V	L	V V	C S	E I	T	A V	T F	R	G
снѕ	P.mirifica				с								· · ·
1		N K G	A	r v	VL	v v	C S	ΕI	т	A V	T F	R	G
1			610		620		630	e	540		650		660
				.	· 10 ·							· ! · ^	
CHS	P.Lobata			ACCTI H T.	GATA	S T.	V C		TTG1	F G	GATGG	AGC C	GCT
снз	P.mirifica				<mark>c</mark> .		•••••	· · · · · · · ·		•••••		<b>z</b>	
1		P S D	т	H L		S L	V G	Q A	L	F G	DG	; A	A

Figure 2. Pairwise sequence alignment of CHS genes between P. lobata and P. mirifica

		670	600	C00	700	710	700
		670	680	690	700	/10	120
CHC	B lobata						
спъ	F.IODala				F V D T	V F I	37
CHS	P mirifica	AVIVG		L F Q V			v
CIID	1.11111104	AVTVG	S D P	T P O V	E K P I	. у в т.	v
		AVIVG		I I Q V			v
		730	740	750	760	770	780
		1 1	, 10	, 30	, 00	1 1	1
CHS	P.lobata	TGGACTGCACAAAC	ATTGCTCCAG	ACAGTGAAGGG	GCTATTGATGO	ACACCTTCGT	GAA
		ω τ Α Ο τ	ТАР	DSEG	ATDG	H T, R	E
CHS	P.mirifica						
_		WTAOT	IAP	DSEG	AIDO	HLR	Е
		~					
		790	800	810	820	830	840
		<u>^</u>		.			
CHS	P.lobata	GTTGGCCTCACATT	CATCTCCICA	AGGATGTTCCT	GGGATTGTCTC	AAAGAACATT	GAT
		VGLTF	HLI	K D V P	GIVS	KNI	D
CHS	P.mirifica	<mark>G</mark>					
		VGLTF	H L L	K D V P	GIVS	KNI	D
		V	V				
		850	860	870	880	890	900
				.			•••
CHS	P.lobata	AAGGCACTTTTTGA	GCATTCAACC	CACTGAACATC	TCTGATTACAA	CTCCATCTTT	TGG
		K A L F E	AFN	PLNI	S D Y N	SIF	W
CHS	P.mirifica	•••••••••••	•••••	• • • • • • • • • • •	•••••	•••••	• • •
		KALFE	AFN	PLNI	S D Y N	SIF	W
		910	920	930	940	950	960
		· · · ·   · · · <mark>/</mark>   · · · ·		.			•••
CHS	P.lobata	ATTGCACA (CCTGG	GGGCCTGCAA	ITTTGGACCAA	GTTGAGCAGAA	GTTGGGTCTC	AAA
		IAHPG	G P A	ILDQ	VEQK	L G L	K
CHS	P.mirifica	· · · · · · · · · · · · · · · · · · ·	•••••			•••••	• • •
		I A H P G	GPA	LLDQ	VEQK	. L G L	K
		070	000	0.0.0	1000	1010	1000
		970	980	990	1000	1010	1020
CHC	D lobata						
спъ	F.IODala						
CHG	D mirifica		AIK	О V Ц Б	DIGN	MSS	A
CIID	r.miririca						7
						11 0 0	11
		1030	1040	1050	1060	1070	1080
							1000
CHS	P.lobata	TGTGTTCTTTCAT	TTGGATGAGA	GAGGAGGAAA	TCAGCTGAAAA	CGGACTTAAA	ACC
		CVLFI	LDEI	M R R K	SAEN	GLK	Т
CHS	P.mirifica						
		CVLFI	LDEI	MRRK	SAEN	GLK	т
		1090	1100	1110	1120	1130	1140
				.			
CHS	P.lobata	ACAGGTGAAGGACT	GAATGGGGTG	TGTTGTTCGGT	TTTGGACCTGG	ACTTACTATT	GAG
		TGEGL	EWG	V L F G	F G P G	LTI	Е
CHS	P.mirifica	••••••	•••••		•••••		• • •
		TGEGL	EWG	V L F G	F G P G	LTI	Е
1							
1		1150	1160	1170			
1		•••• <b>/</b> ••• <b>/</b>	· · · ·   · · · <mark>A</mark>	· <u>^</u> · ·			
снѕ	P.lobata	ACTGTTGTT	AGTGTGGC <mark>C</mark> A	I <mark>C</mark> TGA			
1		T V V L R	SV A	1 *			
CHS	P.mirifica	AI.	<mark>.</mark> .	. <mark>.</mark>			
1		TVVLH	s v a	₩ *			
1		vv	• • • • •	•			
1							

Figure 2. Pairwise sequence alignment of CHS genes between P. lobata and P. Mirifica (next)

# Putative amino acid sequence analysis of CHS gene (*P. lobata* and *P. mirifica*)

The two CHS sequences and five sequences of relatives' species of *Fabaceae* family which have the full-length coding sequences (CDS) were aligned using BLAST and BioEdit (Figure 3). The result showed 42 different amino acids, which may have directed specifically characteristics and their functions, as well as increasing species diversity.

In fact, the highest genetic similarity was found between the CHS gene of P. lobata and CHS8 Glycine max (NP 001304585.2) with 96.24% which was followed by that of CHS7 Glycine max (NP\_001340309.1), Glycine soja (ACT32034.1), Phaseolus vulgaris (XP\_007158815.1), and Vigna radiata (AJZ72655.1) with 96.15%, 95.98%, 91.81% and 89.49%, respectively. Likewise, the most significant match was detected between the CHS gene of P. mirifica and CHS8 G. max (NP 001304585.2), with a similarity of 95%, followed by G. soja (ACT32034.1) with 94,7%, the next V. radiate (AJZ72655.1) being 89.2%, CHS7 G. max of 74.4%, and P. vulgaris, with 73% the lowest one among five selected species. In conclusion, the two CHS gene sequences share similarities to the references, ranging from 73% to 96.24%.

Amino acid sequences of the two genes are analyzed using Conserved Domains software (NCBI) to show the location of several key conserved domains such as Chalcone and stilbene synthase N-terminal, Cterminal, PLN03173, CHS-like, BH0617, and fabH and predict these functions (Fig. 3). PLN03173 (from 1-388aa) seemed to be a provisional chalcone synthase, while the BH0617 (from 17-388aa) region is expected for naringenin-chalcone synthase. CHS\_like (from 16-384aa) are described as Chalcone stilbene synthases; plant-specific and polyketide synthases (PKS) and related enzymes. FabH (from 17-385aa), which is well-known for 3-oxoacyl-(acyl-carrierprotein) synthase III, and is a determining branched-chain factor in fatty acid

biosynthesis (https://www.ncbi.nlm.nih.gov/ Structure/cdd/wrpsb.cgi).

In addition, some motifs supported more understanding of isolated CHS genes in this study. For instance, the conserved CHS active site residues ( $C^{164}$ ,  $F^{215}$  H<sup>303</sup>, and N<sup>336</sup>) were identified functions as the active site of the enzyme, which is essential for the catalytic activity of both enzymes (Chalcone and stilbene) and probably represents the binding site for the 4-coumaryl-CoA group (Ferrer et al., 1999). Especially, C<sup>164</sup> acts as the active site nucleophile and the attachment site for polyketide intermediate in polyketide formation, whereas H<sup>303</sup> and N<sup>336</sup> are known as the decarboxylation of malonyl CoA (Dao et al., 2011 and Ferrer et al., 1999). F<sup>215</sup> are emphasized with the aim of the binding substrate through elongation of the polyketide intermediate (Jez et al., 2000) and the other residues K55, R58 and K62 are also introduced as the CoA binding active sites on the figure (Dao et al., 2011; Ferrer et al., 1999).

The leucine zipper motif, L<sup>310</sup>, L<sup>317</sup>, L<sup>319</sup> and L<sup>331</sup> (Claudot et al., 1999) play an important role as the functional active site and was also found in the PcCHS gene product (Wiriyaampaiwong et al., 2012). The results are shown in Figure 3 that all residues remained unaltered in the two sequences of P. lobata and P. mirifica. And another, it is interesting to note that the family signature "GVLFGFGPGLTI" loop (Suh et al., 2000; Dao et al., 2011) of CHS genes are determined including the unique amino acid Pro-375 (P<sup>375</sup>) in the G<sup>372</sup>FGPG loop appears as a strictly conserved region in all member of CHS superfamily, without other condensing enzymes (Suh et al., 2000). As shown in Figure 3, this motif has also been observed in the two isolated P. lobata and P. mirifica CHS genes. As indicated above, "GVLFGFGPGLTI" was the family signature of the CHS gene and the region belongs to Chalcone - stilbene synthase C-terminal of P. lobata, P. mirifica and 24 species in the Fabaceae family. Therefore, the signature was used to check the conserved of these sequences (Fig. 4).



Figure 3. Amino acid sequence alignment of P. lobata, P. mirifica, and CHS protein of the five most significant similarity species in Fabaceae family with full-length cds. (NP\_001304585.2 (G. max), NP\_001340309.1 (G. max), ACT32034.1 (G. soja), XP\_007158815.1 (P. vulgaris), AJZ72655.1 (V. radiate) are downloaded from NCBI web. Chalcone and stilbene synthase N-terminal, C-terminal are shaded with right and left brackets. The main conserved CHS active site residues are marked in yellow while the putative family signature is indicated by the blue box. The black boxes are shown the CoA binding active sites, whilst the leucine zipper motifs are indicated in the red boxes. A filled black rectangle represents a unique residue. The big black dot is shown the different amino acid positions)

	360	370	380	
		· · · · · · · · · · · · · · · · · · ·	<del>  </del>	
CHS P.lobata	RKSAENGLKTTGEGLE	WGVLFGFGP	GLTIETVVL-	-RSVAI
CHS P.mirifica	RKSAENGLKTTGEGLE	WGVLFGFGP	GLTIZTVVL-	HSVAI
D10223.1 P. lobata	RKSAENGLKTTGEGLE	WGVLFGFGP	GLTIETVVL-	-RSVAI
JQ409456.1 P. mirifica	RKSAENGLKTTGEGLE	WGVLFGFGP	GLTIZTVVL-	HSVAI
NP_001304585.2	RKSAENGLKTTGEGLE	WGVLFGFGP	GLTIETVVL-	-RSVAI
CAA10131.1	KKSAKDGLKTTGEGLE	WGVLFGFGP	GLTIETVVL-	HSVAI
AJZ72657.1	RKSAENRLKTTGEGLE	WGVLFGFGP	GLTIETVVL-	HSVTI
PNY03318.1	RKSKEDGLATTGEGLE	WGVLFGFGP	GLTVETVVL-	HSVAT
ABF59866.1	RKSKEDGLKTTGEGLE	WGVLFGFGP	GLTIETVVL-	HSVAI
ABM66532.1	KKSAQDGLKTTGEGLE	WGVLFGFGP	GLTIETVVL-	HSVAI
AAA67701.1	KKSAQNGLKTTGEGLE	WGVLFGFGP	GLTIETVVL-	HSVAI
AAB81987.1	-KSAQNGLKTTGEGLE	WGVLFGFGP	GLTI TVVLI	LRSVAI
AAU43217.1	NKSLQQGLQTTGEGLE	WGVLFGFGP	GLTIETVVL-	-HSVAI
BAA01512.1	KKSTQDGLNTTGEGLE	WGVLFGFGP	GLTIETVVL-	HSVAI
CAA52819.1	EKSVENGLKTTGKDLE	WGVLFGFGP	GLSLETVVL-	-HSVAI
RDY14063.1	<b>RKSAQNGQKTTGEGLE</b>	WGVLFGFGP	GLTIETVVL-	-RSVAI
ACH67480.1	KKSAQNGLKTTGEGLE	WGVLFGFGP	GLTIETVVL-	-HSVAI
ACB78187.1	-KSIEDGLKTTGEGLD	WGVLFGFGP	GLTVETVVL-	-RSVGVN
XP_015971138.1	EEGKATTGEGFD	WGVLFGFGP	GLT <mark>VE</mark> TVVL-	-HSLPLENRS
KEH27377.1	KRSAQEGLETTGEGLK	WGVLFGFGP	GLTIETVVL-	-HSMVI
AEF14414.1	RKSKENGLATTGEGLE	WGVLFGFGP	GLI <mark>VE</mark> TVVL-	-RSVAA
AFA55180.1	<b>RRSVKEGKATTGEGLE</b>	WGVLFGFGP	GLTVETVVL-	-HSVPV
QCE15713.1	EEGKSSTGEGLK	WGVLYGFGP	GLTMETIVL-	-HSATIDTNN
XP_020230031.1	RKSAKDGLKTTGEGLE	WGVLFGFGP	GLTIETVVL-	HSVAI
TKY62964.1	SKEEGKGTTGEGLE	WGVLFGFGP	GLTVETVVL-	-HSVPLEG
XP_028780252.1	KRSMNGGKATTGEGLE	WGVLFGFGP	GLT <mark>VE</mark> TVVL-	HSVPV
XP_028189397.1	RKSAENGLKTTGEGLE	WGVLFGFGP	GLTIETVVL-	-RSVAI
AYE88587.1	EEGKSTTGEGLN	WGVLFGFGP	GLTMETIAL-	-HSANIDTGY

*Figure 4*. Multiple alignments of an amino acid-specific region of *P. lobata*, *P. mirifica* and other species in the Fabaceae family

There are 16 out of 26 species of the Pueraria genus (including P. lobata and P. mirifica) contained "GVLFGFGPGLTI motif in the Chalcone - stilbene synthase C terminal. However, the last amino acid position of putative CHS family signature was changed to Valine (V) instead of Isoleucine (I) in PNY03318.1 T. pretense, ACB78187.1 (S. tora); XP\_015971138.1 (A. duranensis), AEF14414.1 (O. viciifolia), AFA55180.1 (A. TKY62964.1 (*S*. confuse), suberectus), XP\_028780252.1 (*P. alba*), while only AYE88587.1 (C. korshinskii), QCE15713.1 (V. unguiculata) modified to Methionine (M). Moreover, CAA52819.1 (V. unguiculata) has two amino acids Threonine- Ileucine (TI) at the last position of the 'GVLFGFGPGLTI' motif but has none of Serine - Leucine (SL). Consequently, the 'GVLFGFGPGLTI' amino acid residue region is a family signature of the CHS gene, which was also observed in P. lobata and P. mirifica, but it was not included for all members in the Fabaceae family. With this in sillico analysis, the P. lobata and P.

*mirifica* CHS sequences have highly conserved regions to maintain their structure and function, since, it needs further studies to clarify these points.

#### CONCLUSION

CHS genes from *P. lobata* and *P. mirifica* have been identified and analyzed in the length of 1170 bp encoding for 389 amino acids. These genes have some differences in nucleotide making changes in putative amino acid with references. Additionally, both predicted amino acid sequences of the CHS gene from *P.lobata* and *P.mirifica* were significantly higher similar comparing with published CHS sequences on GenBank (100% and 97.78%, respectively) and also have highly conserved regions to maintain their structures and functions.

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

## REFERENCES

- Austin M. B., Noel J. P., 2003. The chalcone synthase superfamily of type III polyketide synthases. *Nat. Pro. Rep.*, 20(1): 79–110.
- Birt D. F., Hendrich S., Wang W., 2001. Dietary agents in cancer prevention: flavonoids and isoflavonoids. *Pharmaco Ther*, 90(2-3): 157–177.
- Cherdshewasart W., Subtang S., Dahlan W., 2007. Major isoflavonoid contents of the phytoestrogen rich-herb Pueraria mirifica in comparison with *Pueraria lobata*. J *Pharm .Biomed. Anal .*, 43(2): 428–434.
- Cheung D. W. S., Koon C. M., Ng C.F., Leung P. C., Fung K. P., Poon S. K. S., Lau C. B. S., 2012. The roots of *Salvia miltiorrhiza* (Danshen) and *Pueraria lobata* (Gegen) inhibit atherogenic events: a study of the combination effects of the 2-herb formula. J. Ethnopharmacol., 143(3): 859–866.
- Claudot A. C., Ernst D., Sandermann H., Drouet A., 1999. Cloning and characterization of two members of the chalcone synthase gene family from walnut. *Plant. Physiol. Biochem.*, 37(10): 721–730.
- Dao T.T. H., Linthorst H. J. M., Verpoorte R., 2011. Chalcone synthase and its functions in plant resistance. *Phytochem. Rev.*, 10(3): 397–412.
- Dewick, P. M., 2009. The mevalonate and methylerythritol phosphate pathways: terpenoids and steroids. Medicinal natural products: a biosynthetic approach. UK, Wiley: 187-306.
- Dixon R.A., Steele C. L., 1999. Flavonoids and isoflavonoids-a gold mine for metabolic engineering. *Trends. Plant. Sci.*, 4(10): 394–400.
- Estabrook E. M., Gopalan C. S., 1991. Differential Expression of Phenylalanine Ammonia-Lyase and Chalcone Synthase during Soybean Nodule Development. *Plant. Cell.*, 3(3): 299–308.

- Ferrer J. L., Jez J. M., Bowman M. E., Dixon R. A., Noel J. P., 1999. Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis. *Nat. Struct. Biol.*, 6(8): 775–783.
- Healthstore. FSC Kudzu Capsule, 2020. Available from: http://www.healthstore.uk.com/c43206/ku dzu.html. Accessed on 28 November 2020.
- Hrazdina G., Lifson E., Weeden N. F., 1986. Isolation and characterization of buckwheat (*Fagopyrum esculentum* M.) chalcone synthase and its polyclonal antibodies. *Arch. Biochem. Biophys.*, 247(2): 414–419.
- Ichinose Y., Kawamata S., Yamada T., An C. C., Kajiwara T., Shiraishi T., Oku H., 1992. Molecular cloning of chalcone synthase cDNAs from *Pisum sativum*. *Plant. Mol. Biol.*, 18(5): 1009–1012.
- Iwasaki M., Inoue M., Otani T., Sasazuki S., Kurashi N., Miura T., Yamoto S., Tsugane S., 2008. Plasma isoflavone level and subsequent risk of breast cancer among Japanese women: a nested case-control study from Japan Public Health Centerbase prospective study group. J. Clin. Oncol., 26(10): 1677–1683.
- Keung W. M., Vallee B. L., 1998. Kudzu root: an ancient Chinese source of modern antidipsotropic agents. *Phytochemistry* 47(4): 499-506.
- Manonai J., Chittacharoen A., Theppisai U., Theppisai H., 2007. Effect of *Pueraria mirifica* on vaginal health. *Menopause* 14(5): 919–924.
- Miadokova E., Masterova I., Vlckova V., Duhova V., Toth J., 2002. Antimutagenic potential of homoisofl avonoids from Muscari racemosum. J. Ethnopharmacol., 81(3): 381–386.
- Nakajima O., Akiyama T., Hakamatsuka T., Shibuya M., Noguchi H., Ebizuka Y., Sankawa U., 1991. Isolation, sequence and bacterial expression of a cDNA for chalcone synthase from the cultured cells

of Pueraria lobata. Chem. Pharm. Bull ., 39(7): 1911–1913.

- Nakajima O., Shibuya M., Hakamatsuka T., Noguchi H., Ebizuka Y., Sankawa U., 1996. cDNA and Genomic DNA Clonings of Chalcone Synthase from *Pueraria lobata. Biol. Pharm. Bull.*, 19(1): 71–76.
- Pandith S. A., Ramazan S., Khan M. I., Reshi Z. A., Shah M. A., 2019. Chalcone synthases (CHSs): the symbolic type III polyketide synthases. *Planta*, 251(1):15
- Rohde W., Dörr S., Salamini F., Becker D., 1991. Structure of a chalcone synthase gene from *Hordeum vulgare*. *Plant. Mol. Biol.*, 16(6): 1103–1106.
- Ryan- Borchers T. A., Park J. S., Chew B. P., McGuire M. K., Fournier L. R., Beerman K. A., 2006. Soy isoflavones modulate immune function in healthy postmenopausal women. *Am. J . Clin* .*Nutr.*, 83(5): 1118–1125.
- Ryder T. B., Hedrick S. A., Bell J. N., Liang X. W., Clouse S. D., Lamb. J., 1987.
  Organization and differential activation of a gene family encoding the plant defense enzyme chalcone synthase in *Phaseolus vulgaris*. *Mol. Gen. Genet.*, 210(2): 219–233.
- Sambrook J., Russell D. W., 2001. Molecular Cloning. A Laboratory Manual, 3rd ed. Cold.
- Scarpato R., Paganucci L., Bertoli A., Fiore L., Pistelli L., Federico G., 2008. Licoflavone C attenuates the genotoxicity of cancer drugs in human peripheral

lymphocytes. *Phytother. Res.*, 22(12): 1650–1654.

- Shao L., Li Y., Pan A., Cheng Z., Chen M., 1995. Molecular cloning, sequencing, and expression in *Escherichia coli* of the chalcone synthase gene. *Chin. J. Biotechnol.*, 11(2): 131–135.
- Suh D. Y., Fukuma K., Kagami J., Yamazaki Y., Shibuya M., Ebizuka Y., Sankawa U., 2000. Identification of amino acid residues important in the cyclization reactions of chalcone and stilbene synthases. *Biochem. J.*, 350(Pt 1): 229–235.
- TGA. Australian Register of Therapeutic Goods Medicine. 2020; Available from: https://www.ebs.tga.gov.au/ Accessed on 28 November 2020.
- Wanadorn P. W., 1933. A reputed rejuvenator. J. Siam. Soc. Nat. Hist., 9(1): 145–147.
- Wani T. A., Pandith S. A., Gupta A. P., Chandra S., Sharma N., Lattoo S. K., 2017. Molecular and functional characterization of two isoforms of chalcone synthase and their expression analysis in relation to flavonoid constituents in Grewia asiatica L. PloS One, 12(6): e0179155. https://doi.org/ 10.1371/journal.pone.0179155
- Wiriyaampaiwong P., Thanonkeo P., Thanonkeo S., 2012. Cloning and Characterization of Chalcone Synthase Gene from *Pueraria candollei* var. *mirifica. J. Med. Plans. Res.*, 6(42): 5469–5479.