Journal of Animal & Plant Sciences, 34(5): 2024, Page: 1303-1316
ISSN (print): 1018-7081; ISSN (online): 2309-8694
https://doi.org/10.36899/JAPS.2024.5.0812

COMPARISON IN DNA BARCODING, GENE EXPRESSION AND CHEMICAL OF *PUERARIA MIRIFICA* PLANTS

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

[1]Institute of Genome Research, Vietnam Academy of Science and Technology, 100000, Hanoi, Vietnam
[2]Department of Biotechnology, Graduate University of Science and Technology, Vietnam Academy of Science and Technology, 100000, Hanoi, Vietnam
[3]Faculty of Biotechnology, Vietnam National University of Agriculture, 100000, Hanoi, Vietnam
[4] School of Agriculture and Natural Resources, Vinh University, 460000, Nghe An, Vietnam
[5]Pollution Control Department, Ministry of Natural Resource and Environment, 100000, Hanoi, Vietnam
*Corresponding author's: Email: hthue@igr.ac.vn; hueigr76@gmail.com

ABSTRACT

The *Pueraria mirifica* plant is used as a precious material in healthcare and herbal supplement production as it is rich in phytoestrogen content. The wide cultivation gradually directs to the differences among cultivars based on their local original planting region. Besides variation in morphology, *P. mirifica* plants are different in phytoestrogen content. With barcoding markers located on nuclear and chloroplast, our phylogenetic analysis proved that using the group of *mat*K, *rpo*C, *rps*16, and *trn*H-*psb*A concatenated sequences provides better discriminating power than using a single marker of common *P. mirifica* cultivars. Only the maximum likelihood tree based on marker combination separate *P. mirifica* plants from other members of the Pueraria family and separated tribes in the Fabaceae family. In addition, daidzein, genistin, and puerarin contents were recorded from the tuber of all cultivars. These results showed that some are potential cultivars for extracting phytoestrogens from the tuber. Among the five varieties studied, NA has the highest content of Daidzein and Puerarin, and TLBYT has the most difference in the sequences of *rps*16 and *trn*H-*psb*A and the lowest content of Puerarin, Genistin, and Daidzein. Together, qRT-PCR was performed to evaluate eight phytoestrogen biosynthesis gene expressions in leaf and tuber, in which four novel genes *CHS11*, *CHS13*, *CHI2A*, and *UGT2* were examined for the first time in *P. mirifica*. These gene expressions varied depending on organs and cultivars. The differences in both the content of substances and in DNA barcodes as well as gene expression showed a clear polymorphism between varieties. The information is valuable for the selection of material sources for production.

Keywords: *DNA barcoding, P. mirifica, CHS, CHI, UGT.*

Published first online September 23, 2024 Published final October 22, 2024

INTRODUCTION

White Kwao Krua, with the scientific name *Pueraria candollei* var. *mirifica* or *Pueraria mirifica*, plays a critical role in the healthcare and economic sectors. This plant serves as a traditional medicine to treat symptoms of estrogen deficiency and as a rejuvenating therapy, due to its high content of vigor phytoestrogens as isoflavone and chromenes (Yazaki *et al*., 2008; Intharuksa *et al*., 2020; Suntichaikamolkul *et al*., 2023). Notably, miroestrol and its precursor deoxymiroestrol are the two chromenes with the highest estrogenic capacity that were proven to accumulate in *P. mirifica* (Yusakul *et al*., 2011). Currently, there are several *P. mirifica* cultivars that are cultivated and used as materials in medicine and herbal supplement production. Inconsistent cultivation and nurturing of *P. mirifica* may result in changes in its morphology, even the organic content. Correctly identifying the high-quality *P. mirifica* is necessary for safety monitoring input of applications as well as protecting the customer's health. Besides, some species belonging to the Pueria genus have Puerarin in their roots, while other species do not, but in terms of morphology they are quite similar. For example, research on three species *P. montana*, *P. lobata*, and *P. phaseoloides* showed that *P. lobata* contains Puerarin while *P. phaseoloides* and *P. montana* do not have it. In another study, two markers ITS2 and *mat*K were analyzed, which showed that ITS2 is better for distinguishing these species using DNA barcodes (Adolfo *et al*., 2022). Hence, the chemical composition combined with DNA barcode analysis of herbal dietary supplements helped to accurately distinguish different samples (Pawar *et al*., 2017). This raises the need to classify the origin

and identify phytoestrogen concentration among cultivars.

In the past, scientists discovered and classified Pueraria based on morphological methods via characteristics such as stipule, calyx, pod, and vexillum callosities together with flowers per node (Lackey, 1977 a, b; G Van Der Maesen, 1985). Since 2001, molecular studies have provided rigorous evidence in the classification of these plants, and more than morphological tools are needed (Ohashi, 2005; Ohashi and Iokawa, 2006; Pan *et al*., 2015). Other molecular and phylogenetic works were conducted to investigate the evolutionary history and clear taxonomy of *P. mirifica* (Egan *et al*., 2016). The term "DNA barcode" was first coined in 2003, standing for a short 400 - 800 bp DNA sequence that can be isolated and characterized for all plant species (Hebert *et al*., 2003). When combined with sequencing technology and bioinformatics, DNA barcodes reveal accurate and reliable evidence to discriminate between species/varieties, especially in herbal plants (Kress and Erickson, 2007; Coutinho *et al*., 2015). A huge number of DNA barcoding markers were screened among medicinal plants, resulting in different efficiencies of each marker at each level (Hollingsworth *et al*., 2009). Previous studies elucidated the difference in taxa discrimination between chloroplast and nuclear genes (Soltis and Kuzoff, 1995; Zhang *et al*., 2020). DNA barcoding profiles of ITS, *mat*K, *rbc*L, and *trnH-psbA* are used to support HPLC profiling to identify *M. pruriens* (L.) DC plants. It is an efficient method for the identification of the herbal for safe commercial products (Intharuksa *et al*., 2023).

Internal transcribed spacer (ITS) and the *trn*H-*psb*A spacer are known for their barcodes, which are useful for identifying widely angiosperm species (Chen *et al*., 2010). Additionally, the chloroplast marker *mat*K gene has been used as an effective molecular marker in medicinal plants, with wide application across angiosperms and legumes (Chen *et al*., 2010; Egan *et al*., 2016). Both *mat*K and ITS markers were employed and showed the efficiency in discriminatory and phylogenetic studies of Pueraria and other Fabaceae plants. Besides, chloroplast *rps*16 region is an effective tool in phylogenetic studies due to its slower evolution than the nuclear gene (Baker *et al*., 2000) and its ability to separate in sub-tribe Glycininae level (Lee and Hymowitz, 2001);however, working ability in under level is not clear. The rpoC1 marker is known to provide high-quality sequences and have a lower evolution rate in barcoding studies (Hollingsworth *et al*., 2011; El-Sherif and Ibrahim, 2020). To date, DNA barcoding using *rbc*L, ITS, ITS2, *mat*K, *rpo*B, *rpo*C, and *trn*H-*psb*A provides effective quality control for successful species-level identification of the herbal plants (Mahima *et al*., 2022; Chen *et al*., 2023). Altogether, every single marker or coupled marker is effective at a certain level and for certain species, but the potential when combining them is still a conundrum. The DNA barcoding analysis was widely conducted between *Pueraria* species but not much in-tra species.

In this work, the objectives are to clarify the divergence in DNA barcodes and the phylogenetic relationships between some Pueraria cultivars using nuclear and cpDNA sequences and compare the outcome of these means. These parameters will serve as testimonials contributing to further sub-classification, harvesting of raw materials, and ecology-plant relationship research.

MATERIALS AND METHODS

Plant materials: Five *P. mirifica* samples consisted of three Vietnamese-originated cultivars (Nghe An – NA, Son La – SL, Thanh Hoa – TH) and two Thailand-originated cultivars (TLBYT, TLCNX). These four-year-old plants were grown under the same conditions in the botanical garden belonging to Vinh University, Vietnam. For DNA barcode sequencing, three different samples were collected in every cultivar. These samples were labeled from 1 to 3 with their cultivar names (NA1 – NA3, SL1 – SL3, TH1 – TH3, TLBYT1 – TLBYT3, TLCNX1 – TLCNX3). The pieces of tuber and leaf tissues were collected for the RNA extraction procedure and stored in the RNA-later stabilization reagent (Qiagen, Germany) under -80°C. The botanical characteristics of the whole plants were recorded carefully for further classification and study correlation with phytoestrogen concentration.

Molecular markers: A nuclear marker (ITS) and four chloroplast markers (*mat*K, *rpo*C1, *trn*H-*psb*A, *rps*16) were used to assess their suitability as DNA barcodes across all 15 samples. Based on the conserved region in two ends of each marker and the reference sequence of *P. mirifica*, primer pairs were specific designed to amplify in selected plants (Table 1).

DNA extraction, PCR amplification, and DNA sequencing: Total genomic DNA was isolated using the protocol by Doyle and Doyle (1990) with a minor modification. Grind 0.1g of fresh leaf tissue in 2 ml CTAB buffer at 60º C (preheated CTAB buffer: 2% (w/v) CTAB, 1.4 M NaCl, 0.2% (v/v) 2-mercaptoethanol, 20 mM EDTA, 100 mM Tris-HCl, (pH 8.0). Incubate the sample at 60º C for 30 min. Extract twice with chloroform-isoamyl alcohol (24:1). Recover the nucleic acid with 100% (v/v) ethanol (with 1/10V 5 mM sodium acetate). The concentration of total DNA was determined using a Nanodrop Spectrophotometer 2000 (Thermo Fisher Scientific, USA). Each 25 µL PCR reaction contained 12.5 µL 2X Thermo Scientific DreamTaq PCR Master Mix, 2.5 µL template DNA (25-50 ng/ µL), 2.5 µL of primer (10 pmol each primer), and 7.5 µL ddH_2O. The

amplification program was set as 3 minutes at 95°C followed by 35 cycles of 30 s at 95°C, 1 minute at the primer annealing temperature, 5 minutes at 72°C extension; and a final hold at 4°C. The amplification products were checked by electrophoresis running on 1% agarose gel and then purified by GeneJET PCR Purification Kit (Thermo Fisher Scientific, USA). Purified PCR products were sequenced by the Sanger sequencing method, with 1st BASE service. The sequences were verified again by the BLAST tool (NCBI).

Table 1. The designed primers for amplification of the DNA barcoding markers

Region	Primer name	Sequence (5'→3')	Approximate size (bp)	Tm (°C)
ITS	Pu ITS-F	AAGTCGTAACAAGGTTTC	730	54°C
	Pu ITS-R	TCCTCCGCTTATTGATATGC		
*mat*K	Pu matK-F	CGTACAGTACTTTTGTGTTTACGAG	850	61°C
	Pu matK-R	AGTCCATCTGGAAATCTTGGTTC		
*rpo*C1	Pu rpoC1-F	GGCAAAGAGGGAAGATTTCG	505	56°C
	Pu rpoC1-R	CCATAAGCATATCTTGAGTTGG		
*rps*16	Pu rps16-F	AGACATGATTAGATTAGCTGTGG	860	58°C
	Pu rps16-R	GATTCGATAAACGGCTCATTGG		
*trn*H-*psb*A	Pu trnH-psbA-F	GTTATGCATGAACGTAATGCTC	320	56°C
	Pu trnH-psbA-R	ATGGTGGATTCACAATCC		

Sequences analysis: Amplified DNA regions were aligned with the reference sequences using the MUSCLE tool on MEGAv11 software. Variations among sequences and other sequence indicators were then detected via DNasp software. The genetic distance both within and between species was evaluated by the Kimura 2-Parameter (K2P) distance matrix running in MEGAv11 software. The nucleotide sequence of each barcode region of each cultivar for three sequenced times was nearly identical so the represented sequences of each marker were submitted into GENBANK with the accession numbers OR186628 - OR186631, OR203621 - OR203628, OR223818, OR223819.

Phylogenetic construction: Nucleotide sequences of five markers of other Pueraria members as well as the Fabaceae family were retrieved from Genbank for phylogenetic analysis. A sequence matrix for four chloroplast markers was conducted for each species individually. The concatenated sequence datasets were edited and aligned using the MUSCLE algorithm in MEGAv11 software. ITS and concatenated datasets were used to conduct two separate phylogenetic trees under Maximum Likelihood (ML) approaches. The evolution model was determined, and the Taruma 3-Parameter model with a discrete Gamma distribution was selected as the best-fit one.

Determine the phytoestrogen content: Weighing 100 g of each cultivar tuber cut at the largest diameter position, put on separated centrifugation tubes with 500 mL 100% methanol, shake by vortex for a minute. The samples were then ultrasonicated at 40 °C for 30 min. After centrifugation at 6000 rpm for 5 min, the solution was separated into a clean flask. The supernatant was extracted again with 250 mL methanol. The isolated solution was grouped into the previous flask with 100% methanol. The solution was filtered through a filter membrane and then analyzed by the HPLC Alliance system and C18 reverse phase column (250 x 4.6 mm^2; 5 µm) with a GES C18 guard column (Vertical, USA). The mobile phase followed the gradient program with a 1% acid phosphoric mixture in water and methanol. The solvent flow rate was 1 mL/ min, and the injection volume was 20 µL. The photodiode array detector (PDA) wavelength was 260 nm.

Total RNA extraction, cDNA synthesis, and qRT-PCR amplification: To compare the expression pattern of phytoestrogen synthesis genes among cultivars, total RNA from the leaf was isolated by the TRizol reagent (Invitrogen) then the quantity and the quantity and integrity (High sensitivity RNA ScreenTape, Agilent Technology). To extract total RNA from the tube, continuously grind 0.1 g of tuber tissue in liquid nitrogen with PVP 2% , mix in 2ml EB buffer (preheated EB buffer: 1% (w/v) SDS, 25 mM EDTA, 100 mM Tris-HCl, (pH 8.0), 1% β-mercaptol ethanol). Using 3ml acid phenol: chloroform and vortex the mixture, and keep at room temperature for 10 minutes. Extract twice with acid phenol: chloroform to complete the dissociation of nucleoprotein complexes. Recover the nucleic acid with 0.1 volume of 3M sodium acetate and an equal volume of isopropanol (Deepa *et al*., 2014).

Total RNA was used to synthesize cDNA with the Reverse Aid Transcriptase Kit (Thermo Fisher Scientific). Eight primer pairs were designed to amplify eight key genes in the phytoestrogen synthesis route. The

qRT-PCR reaction mixture contained 10 µL of Luna Universal qPCR Master Mix, 20 ng of cDNA template, and 0.5 µL of 10 µM for each primer in a total of 20 µL. The amplification condition was set in the LightCycler 96 System (Roche, Switzerland) as initial denaturation at 95°C for 3 minutes; 45 cycles at 95 °C for 20 minutes, and 60 °C for 30 minutes. Melting curves were obtained by the default setup of the Light Cycler 96 system. The assay was conducted with three techniques with one gene and biological replicates with one sample (tuber/leaf cultivar). We used the *EF1* alpha gene as an endogenous control. The mean expression level of the genes was compared to *EF1* control using Student' s t-test with * for $P < 0.05$, and ** for $P < 0.01$. Statistical analysis was conducted following the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001). All real-time primer information and amplification conditions are indicated in Table 2.

Table 2. qRT-PCR primers to validate the phytoestrogen gene expression

Primer name	Sequence (5′-3′)	Size (bp)	Tm (°C)
CHI2A F	ACAATGATCTTGCCCCTGAC	154	55.0 °C
CHI2A R	CTGGTGGGAATGTTTCGTCT		
CHS11 F	TCGTGTGCTTGTCGTTTGTT	150	55.5 °C
CHS11 R	GGGTTTTCAACTGGCAGAGG		
CHS13 F	GAAGCGATACATGCACCTGA	180	54.6 °C
CHS13 R	TGCGTGATCTTTGACTTTGG		
UGT2 F	TCCAAATGAAGCCAAAGACC	210	53.8 °C
UGT2 R	CCTTTGTCATCTTCCCCGTA		
IFR6 F	CGATGCAATTGAGCCAGTTA	202	53.3 °C
IFR6 R	CGCTCCTTTCACATTTCCAT		
HID F	CACCCTTGGCTGCTCTAAGT	162	59.5 °C
HID R	GGGTTGAAGAGCTGGAAAGC		
IFS F	CCAAAGCCTCGTCTTCCCTT	232	60.0 °C
IFS R	AAGTGAGGCGCCTTATGGCT		
CYP81E63 F	TTTGGGATGGGAAGAAGGGC	180	55.6 °C
CYP81E63 R	CTTGCACATGGCCTCCAATG		
EF1α F	GCTATGCACCTGTTCTTGAT	117	58.0 °C
EF1α R	GCTCCTTCTCAAGTTCCTTAC		

RESULTS

Botanical morphology: All five cultivars were vein plants with deltoid leaves and trifoliate leaf structures. Trichome distributed in both plant stem and leaves, especially much more in NA's leaves. The NA stem color was also distinct from the samples; it was violet-olive in NA plants and sage-gray in others. On average, the leaf number ranged from 10-30 in the three Vietnamese cultivars while it was 60-80 in the TLCNX and raised to 80-120 in the TLBYT cultivar. Regarding the tuber, there was also a slide difference in yield and size among cultivars. Vietnamese cultivars yield 10-17 tubers per plant with 7-30 cm in length and 4.5-20 cm in diameter. On the other hand, Thailand cultivars gave a yield of 24-30 tubers per plant with smaller size than Vietnamese ones: 15-30 cm in length and 4.5-10 cm in diameter. The TLBYT tubers were ovoid while the other cultivar tubers were globoid with indent at two ends of the tuber. Therefore, NA and TLBYT plants had distinct botanical characteristics among the varieties (Fig. 1).

The efficiency of marker amplification: It was found that the primers of *rpo*C1, *trn*H-*psb*A, and *rps*16 provided a quality amplicon whilst those of ITS and *mat*K were optimized in *P. mirifica*. As shown in Table 3, after trimming and alignment, the whole sequenced amplicon of each marker was 565 bp for ITS, 791 bp for *mat*K, 492 bp for *rpo*C1, and 759 bp for *rps*16. Therefore, the alignment length for the combined chloroplast dataset was 2310 bp. Our ITS amplified region had the highest GC% content with 57%, followed by *rpo*C1 with 41.4%. This parameter was 28% in *mat*K, *trn*H-*psb*A, and *rps*16, which were 29.3%, and 30.6%, in that order.

DNA barcoding analysis: Several nucleotide differences in each alignment set for *P. mirifica* cultivars are indicated in Table 3. Overall, *trn*H-*psb*A has the shortest amplicon length while it provides the highest variable sites and parsimony-informative (PI) sites (nine sites). *Rps*16 sequences range about 800bp and result in high variable points (seven positions). The three TLBYT plants contain two stretches of five nucleotides and a stretch of 10 nucleotides in the *trn*H-*psb*A region that is divergent from the rest 12 plants (Table 4). A similar outcome was observed for *rps*16 when only TLBYT cultivars consisted of a five-nucleotide stretch and a four-

nucleotide stretch (Table 4). The parameters for the rest of the single markers were significantly smaller. The mean pairwise distance according is the highest number for *trn*H-*psb*A and *rps*16 (0.0105 and 0.0032, respectively). The concatenated cpDNA hence had the highest number of variation, PI, and gap sites and the middle mean pairwise distance (0.0040).

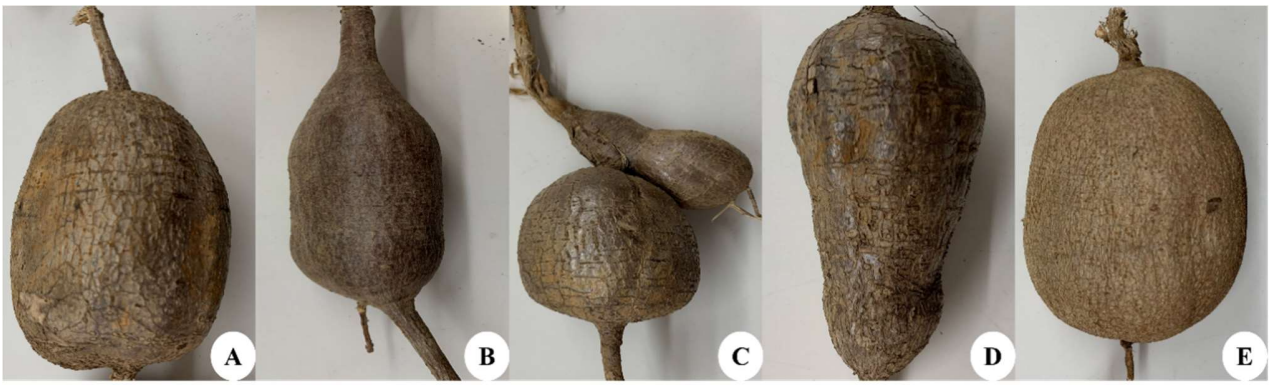

Fig. 1. The represented tuber root specimens of the *Pueraria mirifica* cultivars (A) NA cultivar, (B) SL cultivar, (C) TH cultivar, (D) TLBYT cultivar, (E) TLCNX cultivar

Table 3. DNA marker dataset statistics

	ITS	**matK**	**rpoC1**	**trnH-psbA**	**rps16**	**Chloroplast concatenated sequence**
Length range (bp)	730	850	505	320	860	-
Aligned/analysis length (bp)	565	791	492	316	759	2358
Average %GC content	57	28	41.4	29.3	30.6	32.2
Conserve site (%)	99.65	99.49	99.19	92.09	97.89	98.87
Variable site	3	4	4	9	7	26
Parsimony informative (PI) site	2	4	4	9	7	21
Gap/ missing site (%)	0	1	0	15	9	15
Mean pairwise distance	0.0017	0.0020	0.0030	**0.0105**	0.0032	0.0040
Intraspecific distance (min-max/ mean)	0.0006-0.2242 (0.0565)	0.0000-0.0030 (0.0009)	0.0000-0.0051 (0.0022)	0.0000-0.0175 (0.0075)	0.0088-0.0856 (0.0294)	**0.0011-0.0037 (0.0024)**
Interspecific distance (min-max/ mean)	0.0016-0.2146 (0.1069)	0.0000-0.0434 (0.0125)	0.0000-0.0105 (0.0064)	0.0000-0.0497 (0.0270)	0.0111-0.1320 (0.0589)	**0.0347-0.1538 (0.1124)**

In the present study, intraspecific and interspecific distance values overlapped in the discrimination power of five separated barcodes. This overlap might be due to the diversity in the *Pueraria* genus and the similarity in phenotype characteristic of the Fabaceae members. However, the mean interspecific distance values were higher than intraspecific values in all markers. For the concatenated barcodes, the lowest value of interspecific was higher than the intraspecific one, which proved this dataset's reliable discrimination power. This dataset was used for the phylogenetic analysis of five *P. mirifica* cultivars.

There were numerous variations among plants in the same cultivar and variations among different cultivars (Table 4). Three nucleotide sequences for each region of each cultivar had a highly significant identity. In more detail, three SNPs in the ITS region resulted in four haplotypes among *P. mirifica* samples, four SNPs in *mat*K, and two SNPs in *rpo*C1 both gave only three haplotypes. Meanwhile, nine and seven SNPs were observed respectively in *trn*H-*psb*A and *rps*16 regions, and two haplotypes were recorded for each marker. Notably, almost variations were detected in the Thai cultivars (TLBYT and TLCNX), concentrating more on the three TLBYT samples. The sequence characteristic initially showed the divergence between Vietnamese and Thai *P. mirifica* plants. There are nucleotide differences in DNA barcodes between varieties especially in the *trn*H-*psb*A and *rps*16.

Phylogenetic analysis: To elucidate the aboriginal cultivars, we found the best model and parameters before constructing the phylogenetic tree. There were differences in branching between trees based on the ITS region and trees based on concatenated cpDNA markers. In the ITS tree, five selected cultivars were grouped with the *P. candollei, P. candollei* var. *mirifica* sequences, and this group separated from clusters for *P. montana*, *P. montana* var. *lobata*, and *P. montana* var. *thomsonii* (Figure 2). A similar result was described in the tree by other single cpDNA markers. By contrast, in the combined cpDNA tree, *P. mirifica* sequences concentrated into a distinct cluster, which was separated from the *P. lobata* cluster. Remarkably, the three samples of TLBYT cultivars were also distributed into their branch, laying between *P. mirifica*'s and *P. lobata*'s clusters. As mentioned above, TLBYT plants carry differences in both botanical characteristics and barcoding sequences from other cultivars. In a wider taxonomy range, the concatenated dataset tree discriminated clusters in the tribe rank while there was chaos in the sub-tribe order in the ITS tree. That might be due to putting all the PI from cpDNA markers together providing a better potential to separate members at the cultivar level. Moreover, it can be seen in Fig. 3 that the concatenated barcode set could be discriminated at both inter- and intraspecific levels regarding all other members in the fabaceae family.

Table 4. Variations in the DNA barcoding sequences among the *Pueraria mirifica* cultivars

	Site	NA1	NA2	NA3	SL1	SL2	SL3	TH1	TH2	TH3	TLBYT1	TLBYT2	TLBYT3	TL CNX1	TL CNX2	TL CNX3
ITS	138	**A**	**.**	.	.	.	.	**G**	.	.	G	G	G	.	.	**.**
	227	**T**	**C**	C	C	C	C	**C**	C	C	C	C	C	.	.	**.**
	541	**C**	**.**	.	.	.	.	**.**	.	.	.	.	.	.	.	**G**
matK	416	**G**	.	.	.	.	.	.	.	.	**A**	A	A	**.**	.	.
	453	**T**	.	.	.	.	.	.	.	.	**G**	G	G	**G**	G	G
	526	**C**	.	.	.	.	.	.	.	.	**G**	G	G	**.**	.	.
	655	**C**	.	.	.	.	.	.	.	.	**.**	.	.	**T**	T	T
rpoC1	257	**C**	.	.	.	.	.	.	.	.	**A**	A	A	**A**	A	.
	292	**C**	.	.	.	.	.	.	.	.	**A**	A	A	**.**	.	.
trnH-psbA	86..90	CTTTT	.	.	.	.	.	.	.	.	AAAAC	AAAAC	AAAAC	.	.	.
	108	A	.	.	.	.	.	.	.	.	C	C	C	.	.	.
	111	-	.	.	.	.	.	.	.	.	TTATT	TTATT	TTATT	.	.	.
	128	T	.	.	.	.	.	.	.	.	G	G	G	.	.	.
	142	-	.	.	.	.	.	.	.	.	TCGTTTTTGATCGTTTTTGATCGTTTTTGA			.	.	.
	173	A	.	.	.	.	.	.	.	.	C	C	C	.	.	.
	206	T	.	.	.	.	.	.	.	.	A	A	A	.	.	.
	225	G	.	.	.	.	.	.	.	.	A	A	A	.	.	.
rps16	48	T	.	.	.	.	.	.	.	.	C	C	C	.	.	.
	196	-	.	.	.	.	.	.	.	.	TTGAT	TTGAT	TTGAT	.	.	.
	210	A	.	.	.	.	.	.	.	.	G	G	G	.	.	.
	329	A	.	.	.	.	.	.	.	.	T	T	T	.	.	.
	428	A	.	.	.	.	.	.	.	.	T	T	T	.	.	.
	449	C	.	.	.	.	.	.	.	.	T	T	T	.	.	.
	484	A	.	.	.	.	.	.	.	.	C	C	C	.	.	.
	654	G	.	.	.	.	.	.	.	.	T	T	T	.	.	.
	730	-	.	.	.	.	.	.	.	.	TTTG	TTTG	TTTG	.	.	.

The dot "." indicates the same nucleotide as the first sequence (NA1), the dash "-" indicates the aligned gap

Phytoestrogen content: HPLC analysis showed the concentration of phytoestrogens varied among cultivars (Table 5). The three compounds Daidzein, Genistin, and Puerarin were determined to be over the Limit of Quantification (LOQ, LOQ = 10 µg/g) in just the SL plant. By contrast, none of them recorded to be over LOQ in TLBYT plants and some of the compounds were under LOQ in the remaining samples. NA plant had the highest content of Daidzein and Puerarin (44.2 µg/g and 28.8 µg/g, respectively) whilst its genistin was lower than LOQ. We detected only Puerarin at TH (22.2 µg/g) and only Genistin at TLCNX (11.6 µg/g) among the three selected compounds.

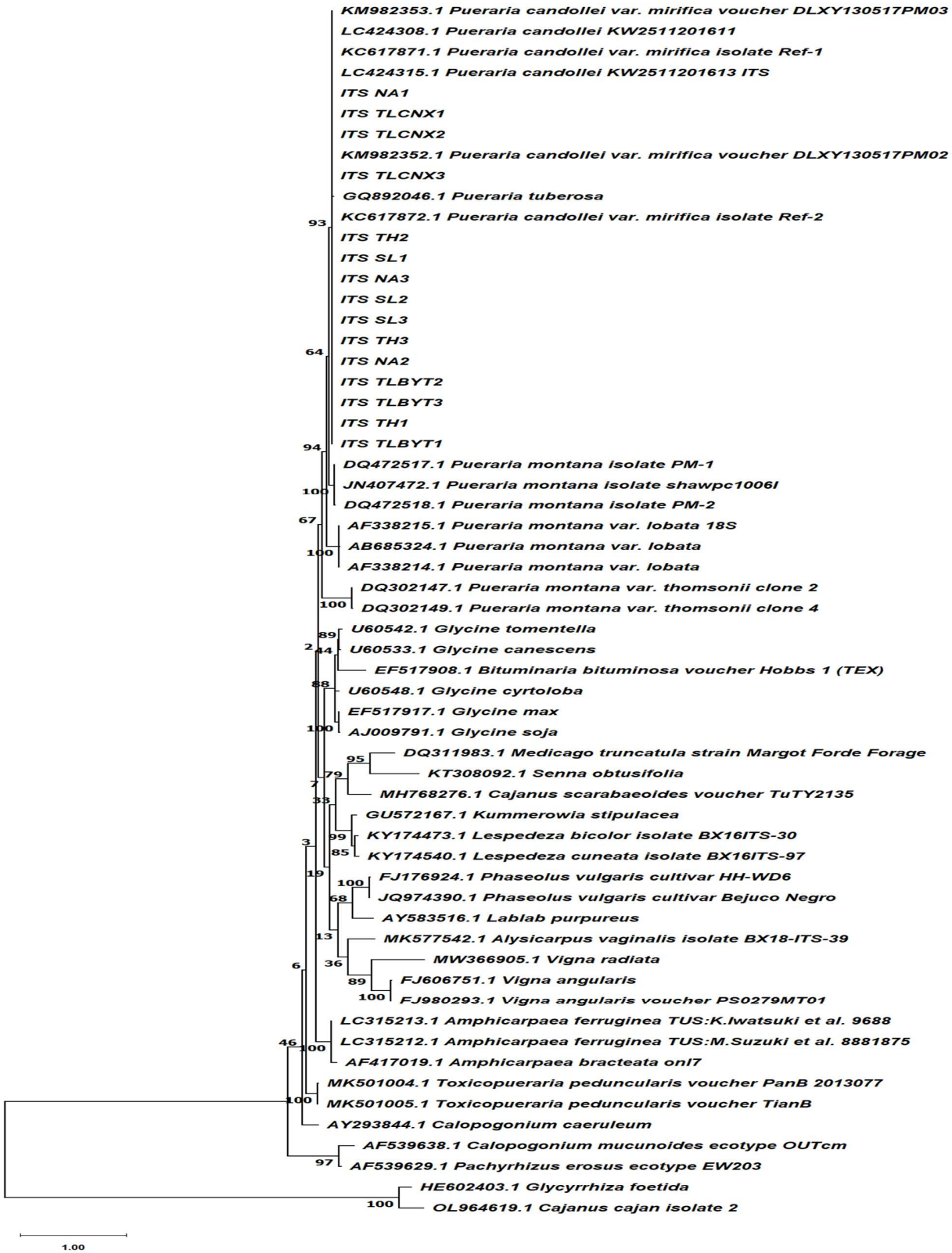

Fig. 2. The maximum likelihood tree resulting from Taruma 3-Parameter + Gamma model analysis of the ITS region, 1000 bootstrap replicate.

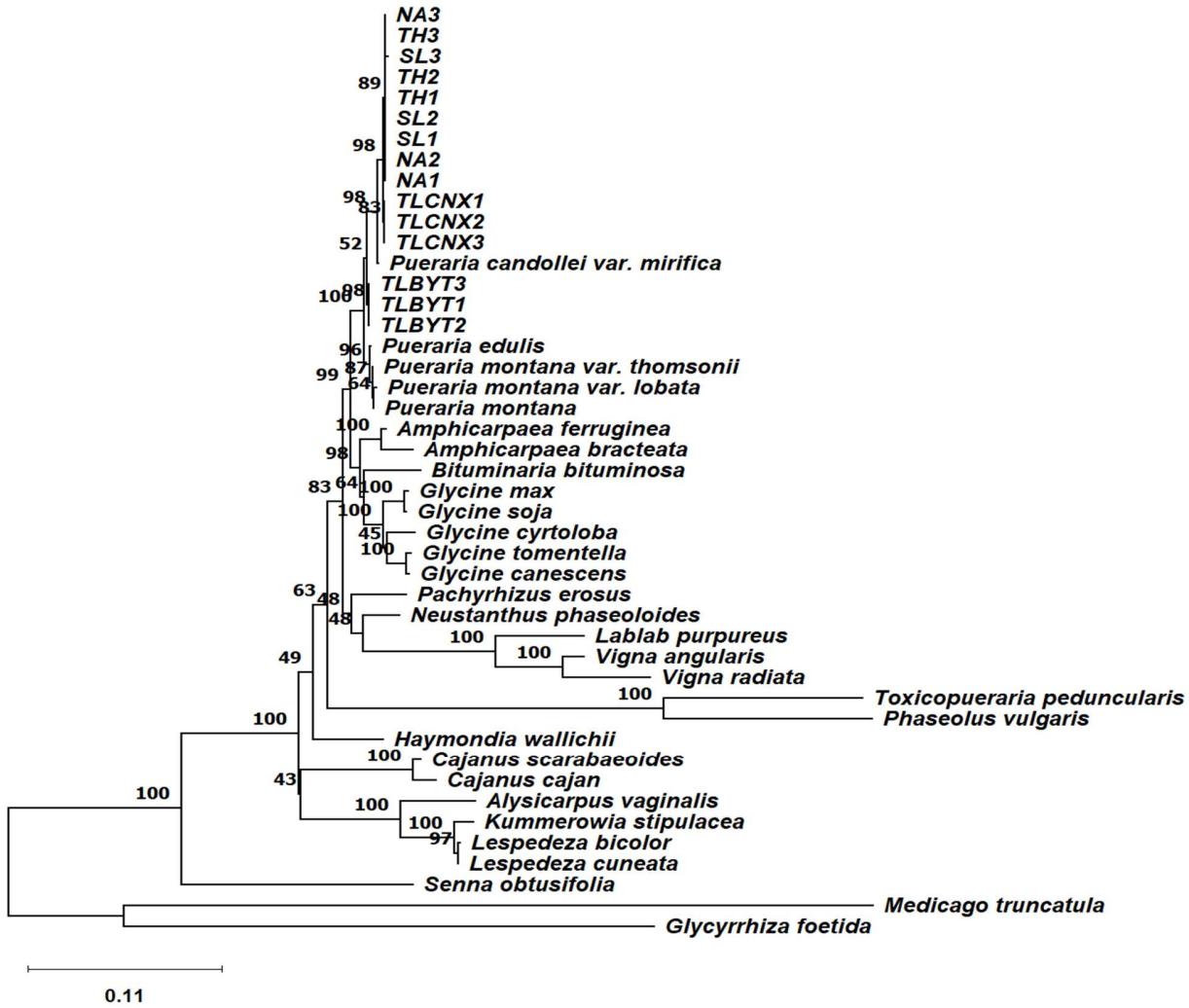

Fig. 3. The maximum likelihood tree resulting from Taruma 3-Parameter + Gamma model analysis of the concatenated matK, rpoC1, trnH-psbA, rps16 sequences, 1000 bootstrap replicate

Table 5. The concentration of daidzein, genistin, and puerarin in tubers of the *Pueraria mirifica* cultivars.

Cultivars	Concentration (µg/g)		
	Daidzein	Genistin	Puerarin
NA	44.2	-	28.8
SL	34.1	16.9	25.4
TH	-	-	22.2
TLBYT	-	-	-
TLCNX	-	11.6	-

"-" denotes under the Limit of Quantification (LOQ = 10 µg/g).

Gene expression pattern among samples: To validate the gene expression involved in the phytoestrogen production in *P. mirifica* cultivars, qRT-PCR was conducted and eight genes in both leaf and tuber specimens were successfully amplified. These genes belong to seven gene families, which were indicated to function in converting intermediates in the (iso)flavone and miroestrol biosynthesis pathway. We especially examined the expressions of *CHS11*, *CHS13*, *CHI2A*, and *UGT2*, which are novel genes that have not been studied in the *P. mirifica* phytoestrogen synthesis route before (Fig. 4).

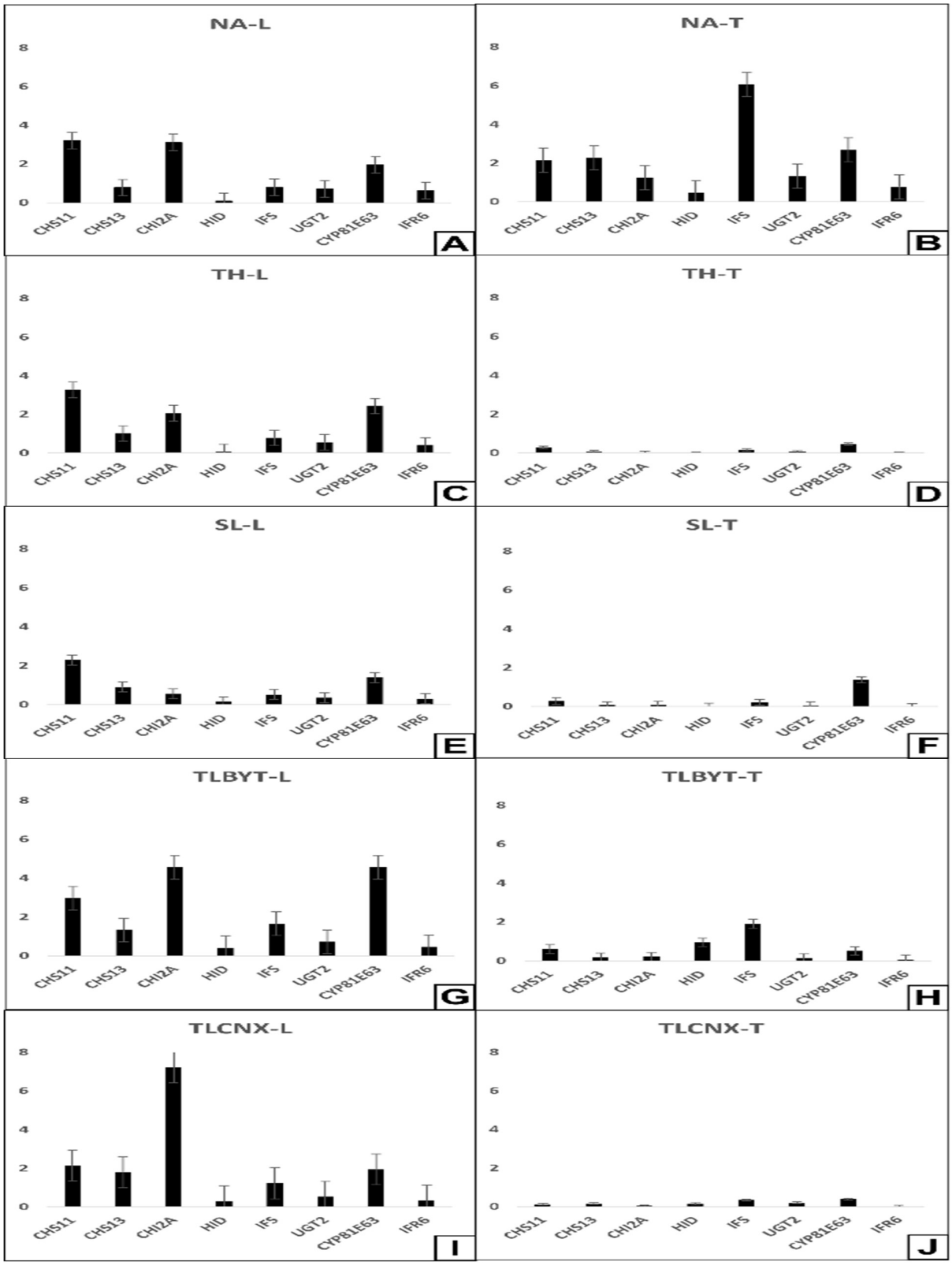

Fig. 4. Expression of the genes function in phytoestrogen biosynthesis pathway in the leaf and tuber root of five Pueraria mirifica cultivars. "L" denotes the leaf sample, and "T" denotes the tuber root sample.

The *CHSs* and *CHIs* are the initial genes that function in the (iso)flavone biosynthesis pathways, which catalyze the conversion of (iso)liquiritigenin. The genes *CHS11*, *CHS13*, and *CHI2A* tend to be much more abundant in the leaves than in the tuber roots of all cultivars. More specifically, these genes reduce expression strongly at the roots of TLBYT and almost do not activate at the roots of TH, SL, and TLCNX. Likewise, *HID* and *IFS* express significantly in the tuber roots of NA and TLBYT, and a reverse trend was observed in the TH, SL, and TLCNX samples. This outcome is similar to previous results from the studies on these gene expressions in *P. mirifica* and *P. lobata.* Remarkably, *IFS* expression was extremely elevated in the NA tuber root in comparison with other cultivars. These two genes are active upstream of daidzein biosynthesis, whose concentration was detected in the upper LOQ merely in the NA and SL roots. These outcomes imply the more complicated biosynthesis of Daidzein concerning the pathway and the gene families involved. Considering the NA plants, the expression of genes *UGT2*, *CYP81E63*, and *IFR6* were much higher in the roots while a reverse trend was seen in the remaining four cultivars. In addition, the overall expression trend of these genes is dominant in the NA cultivar.

DISCUSSION

In the research, five barcoding markers were amplified in a total of 15 four-year-old *P. mirifica* plants, giving an outcome of 28 polymorphism sites distributed across samples. The variations in barcoding sequences among different cultivars and in plants of the same cultivars were mostly distributed in the Thai cultivars, certainly the TLBYT plants. *Trn*H-*psb*A, *rps*16, and *mat*K accounted for the vast majority of substitution numbers. Given that *mat*K and *rps*16 are coding regions, both synonymous SNP and non-synonymous SNP potentially affect the molecular pathways, which are capable of change in plant phenotype. Although synonymous SNP does not change in amino acid sequence, they still can interfere with messenger RNA splicing and the stability and the structure of the protein, and the protein folding process by preventing RNA-binding protein from recognizing transcript (Hunt *et al*., 2014). However, the higher numbers of variable sites and pairwise distances do not correlate with the high classification ability of the marker (Manzanilla *et al*., 2018). These plants were collected and cultivated separately under similar conditions so that the slight difference in the TLBYT morphology and barcoding may be due to the previous planting and environmental conditions.

Using only a single marker is insufficient to make DNA barcodes a core tool in botanical systematic (Kress, 2017). Previous barcoding studies in *Pueraria spp*. combined ITS2 with *mat*K marker (Adolfo *et al*., 2022) and *mat*K with *trn*D-*trn*T (Egan *et al*., 2016) provided a clear classification system in this wide genus; however, *P. mirifica* was not differentiated. Although a single concatenated tree that includes four chloroplast genes could provide a better-resolving strength to show the relatedness of Vietnamese and Thai *P. mirifica* cultivars, there was a lack of marker sequences of other Pueraria plants in Genbank. Our research provides more for this data bank.

The phylogenetic analysis made clear the aboriginal relation of contemporary *P. mirifica* plants; however, the diversity in barcodes and complicated breeding selection prevented us from classifying conspecific plants (Hollingsworth *et al*., 2011). In the taxonomy of the *Pueraria* genus, employing single markers and a combination of two markers was efficient in differentiating at interspecies levels (Munger, 2002; Chen *et al*., 2010; and Adolfo *et al*., 2022), but the intraspecific rank classification needs stronger tools. Previous studies carried out on barcoding sequences, transcriptomic data, HPLC analysis, and nuclear markers provided more detail about the evolution and classification of medicinal plants (Adolfo *et al*., 2022; Intharuksa *et al*., 2023). Based on the genetic distance analysis, statistical data about intraspecific and interspecific distance demonstrated that a combination of cpDNA barcodes has the potential to classify at cultivar or intraspecies level. In this work, TLBYT members were successfully separated in the phylogenetic tree by concatenated cpDNA barcoding markers; meanwhile, the single marker constructed the tree with equivocal resolution. Considering the morphological characteristics, the TLBYT produces a root system with long thin tuber-like roots, slightly similar to the root system of kudzu (Munger, 2002; Lindgren *et al*., 2013). As mentioned above, TLBYT plants carry differences in both botanical characteristics and barcoding sequences from other cultivars. TLBYT cultivars might be crossbreeding with other Pueraria species based on their divergence in botanical phenotype and phylogenetic characteristics.

Chemical identification using the HPLC method was taken to record the content of three important phytoestrogens in the five *P. mirifica* cultivars. The results show that the NA and SL cultivars are potential breeds for medicinal and pharmaceutical purposes related to extracting phytoestrogens from the tuber. Previous studies demonstrated that morphological features, certainly root diameter, correlate with secondary abundance (Duan *et al*., 2018). This conclusion is congruent with the low quantity of isoflavone content in the TLBYT root. It is congruent with the higher leaf characteristics of these plants that we indicated above when compared with the three Vietnamese cultivars. Even though approximately half of the net yield of isoflavone in *P. mirifica* was recorded in tubers

(Cherdshewasart and Sriwatcharakul, 2007), leaves are more convenient in terms of harvesting. It takes at least three years for *P. mirifica* tuber to reach the mature stage; leaves only take about three months. Accordingly, we imply that selecting phytoestrogen sources from *P. mirifica* is critical depending on tissue types and environmental conditions.

The qRT-PCR analysis showed more insightful views of phytoestrogen synthesis routes in these plants. Even though the studied cultivars were conspecific and grown under the same conditions, there were elevated differences in the botanical phenotype characteristics and the (iso)flavone synthesis pathways. The gene expression and their secondary product concentration do not always correlate. The research of Dastmalchi and Dhaubhadel (2015) on soybeans indicated that its isoflavone accumulated mostly in leaves while its synthetic gene *IFS* expressed low there. Similar results were reported in previous studies on tobacco and *P. mirifica* (Yazaki *et al*., 2008; Suntichaikamolkul *et al*., 2019). In *Glycine max*, the dominant site for *CHS13* is also in the leaf while for *CHS11* is the root. As the same, these gene of the samples were mostly expressed in leaves. Their evidence proved although isoflavone synthesis genes *CHSs*, *CYPs*, *IFRs*, *IFSs*, and *PTs* expressed much more in leaves, the plant's isoflavone content accumulated mostly in the root. *UGT2* was known to be active with genistein and preferentially transcribed in the root (Wang *et al*. 2016) and the *CYP81E63* region, specifically exhibiting a key role in miroestrol biosynthesis (Suntichaikamolkul *et al*., 2023). Their high expression result recorded in the NA *P. mirifica* plant implies the good quality of phytoestrogen provided by this cultivar. The uniform expression of the tested genes in both NA leaves and stems may be related to the high accumulation of the two substances. In the research, *CHI2A* gene showed abundant expression in leaves. The *CHI2A* is known as *CHI* type I, which converts 6'-doxychalcone to 5-deoxyflavone in Fabaceae plants and was studied in detail in *G. max* (Yin *et al*., 2019). Previous research on *P. mirifica* transcriptome detected that *CHI2* expressed highest in tuberous cortices and cortex-excised tuber tissue (Suntichaikamolkul *et al*., 2019); however, it was not clear if it is *CHI2A* or not. The diversity of gene expression in different tissues may be due to their diverse promoter regions (Vadivel *et al*., 2018). In addition, *CHS* is a large gene family and is involved in various points in the phytoestrogen synthesis pathway.

The phytoestrogen content in these cultivars might accumulate mostly in larger tuber roots or after a longer culturing time. Another hypothesis is that perhaps this cultivar accumulates isoflavone mostly in leaves. Previous studies proved that three Thai *P. mirifica* cultivars contain a high net yield of total isoflavone level in leaves, reaching 61.6 ± 4 mg/100g of powder when cultivar PM-V were harvested in September (Jungsukcharoen *et al*., 2014). The authors also showed that the content and concentration of each substance in *P. mirifica* leaves depend on cultivars and cultivation conditions such as temperature and amount of rainfall. Genistin amount in leaf was the lowest in PM-V with 13.0 ± 0.1 mg/100g of powder but the isoflavone content was the highest in PM-III and PM-IV cultivars (Jungsukcharoen *et al*., 2014) and much even higher than the recorded content in the *P. mirifica* SL and TLCNX tubers in this research. Therefore, the leaves may be the potential source for isoflavone in applications. As described above, the expression of these genes is diverse not only among species but also among cultivars and organs in the same cultivar. This raises the hypothesis of a connection between the gene expression, gene activation, and secondary metabolite accumulation in medicinal plants. However, the regulation and expression of genes participating in the synthesis of secondary compounds include hundreds of genes, so the qRT-PCR results in the study showed their differential expression between varieties, which is similar to the differences in the content of analyzed substances in these varieties. Even though using multiple DNA barcoding markers together with other biochemicals, and molecular tools may take a lot of time and require a high budget, they provide us with advanced outcomes in human healthcare applications.

Conclusions: The research based on DNA markers - *mat*K, *rpo*C1, *rps*16, *trn*H-*psb*A showed a more efficient resolution at the subspecies level than ITS nuclear marker among *P. mirifica* cultivars. The sequences separated TLBYT into a distinct branch from *P. mirifica* and *P. lobata* plants and separated plants by tribes better than ITSs did. In addition, HPLC coupled with qRT-PCR provided detailed information about the phytoestrogen content and its biosynthesis route. The qRT-PCR was performed to evaluate eight phytoestrogen biosynthesis gene expressions in leaf and tuber, in which four novel genes *CHS11*, *CHS13*, *CHI2A*, and *UGT2* were examined for the first time in *P. mirifica*. Besides TLBYT's molecular characteristic, its botanical morphology is also different from the rest cultivars with the lowest content of the three substances tested.

Acknowledgment: This work was funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 106.02-2019.13.

Declaration of Interest's statement: The authors declare there are no competing interests.

Authors' contribution: Hue Thi Thu HUYNH, conceived, initiated the project, and designed the experiments; Thuy Thi Thu CAO, Canh Xuan NGUYEN, Duong Thuy NGUYEN analyzed the data; Ngoc Thi Bich NGUYEN, Dung Thi Thu CAO, Oanh Thi Kieu PHAM,

performed the experiments; Hue Thi Thu HUYNH, Thuy Thi Thu CAO, Giang Van NGUYEN wrote the article. All authors interpreted the data, critically revised the manuscript for important intellectual contents, and approved the final version.

REFERENCES

Adolfo, L.M., X. Rao and R.A. Dixon (2022). Identification of *Pueraria* spp. through DNA barcoding and comparative transcriptomics. BMC Plant Bio. 22(1): 10. doi: 10.1186/s12870-021-03383-x.

Baker, W.J., T.A. Hedderson and J. Dransfield (2000). Molecular phylogenetics of subfamily Calamoideae (Palmae) based on nrDNA ITS and cpDNA rps16 intron sequence data. Mol. Phylogenetics Evol. 14(2): 195-217. doi: 10.1006/mpev.1999.0696.

Chen, S., Y. Hui, H. Jianping, L. Chang, S. Jingyuan, S. Linchun, Z. Yingjie, M. Xinye, G. Ting, P. Xiaohui, L. Kun, L. Ying, L. Xiwen, J. Xiaocheng, L. Yuli and L. Christine (2010). Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. PLoS ONE. 5(1): e8613. doi: 10.1371/journal.pone.0008613.

Chen, S., Y. Xianmei, H. Jianping, S. Wei, Y. Hui, S. Jingyuan and L. Xiwen Lib (2023). DNA barcoding in herbal medicine: Retrospective and prospective. J Phar Anal. 13(5): 431- 441. doi: 10.1016/j.jpha.2023.03.008.

Cherdshewasart, W. and S. Sriwatcharakul (2007). Major Isoflavonoid contents of the 1-year-cultivated phytoestrogen-rich herb, *Pueraria mirifica*. Biosci. Biotechnol. Biochem. 71(10): 2527-2533. doi: 10.1271/bbb.70316. Epub 2007 Oct 7.

Coutinho, M.D., D. Still, M. Lum and A. Hirsch (2015). DNA-based authentication of botanicals and plant-derived dietary supplements: where have we been and where are we going? Planta Medica. 81(09): 687-695. doi: 10.1055/s-0035-1545843.

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

Deepa, K., T.E. Sheeja, R. Santhi, B. Sasikumar, C. Anu, P.V. Deepesh and D. Prasath (2014). A simple and efficient protocol for isolation of high-quality functional RNA from different tissues of turmeric (*Curcuma longa* L.). Physiol Mol Biol Plants. 20(2):263-271. doi: 10.1007/s12298-013-0218-y.

Doyle, J.J. and J.L. Doyle (1990). Isolation of plant DNA from fresh tissue. Focus. 12(1): 13-15.

Egan, A.N., M. Vatanparast and W. Cagle (2016). Parsing polyphyletic Pueraria: delimiting distinct evolutionary lineages through phylogeny. Mol.Phylogenetics Evol. 04: 44-59. doi:10.1016/j.ympev.2016.08.001.

El-Sherif, N. and M. Ibrahim (2020). Implications of rbcL and rpoC1 DNA barcoding in phylogenetic relationships of some Egyptian medicago sativa L. Cultivars. Egypt. J. Bot. 60(2): 451-460. doi: 10.21608/ejbo.2020.20028.1399.

Van der Maesen, L.J.G (1985). Revision of the genus Pueraria DC with some notes on Teyleria backer. Agricultural University Wageningen Papers (Netherlands). 5(26): 1-85.

Hebert, P.D.N., A. Cywinska, S.L. Ball and R. de W. Jeremy (2003). Biological identifications through DNA barcodes. Proc Biol Sci. 270(1512): 313-321. doi: 10.1098/rspb.2002.2218.

Hollingsworth, P.M., L.F. Laura, L.S. John, H. Mehrdad, R. Sujeevan, B. Michelle van W.C. Mark, S.C. Robyn, L.E. David, J.F. Aron, W.G. Sean, E.J. Karen, K. Ki-Joong, W.K. John, S. Harald, van A. Jonathan, C.H.B. Spencer, van den B. Cassio, B. Diego, S.B. Kevin, M.C. Kenneth, C. Mark, C. Juliana, C. Alexandra, J.C. James, C. Ferozah, S.D. Dion, S.F. Caroline, A.J.H. Terry, L.H. Michelle, C.H. Brian, J.K. Laura, R.K. Prasad, S.K. Jung, K. Young-Dong , L. Renaud, L. Hae-Lim, G.L. David, M. Santiago, M. Olivier, M. Isabelle, G.N. Steven, P. Chong-M.P. Diana, P. Gitte, E.R. James, A.S. Gerardo, S. Vincent, S. Ole, J.W. Michael, Y. Dong-Keun and P.L. Damon (2009). A DNA barcode for land plants. Proc Natl Acad Sci USA., 106(31): 12794-12797. doi: 10.1073/pnas.0905845106.

Hollingsworth, P.M., G.S. Sean and D.P. Little (2011). Choosing and using a plant DNA barcode. PLoS ONE. 6(5): e19254. doi: 10.1371/journal.pone.0019254.

Hunt, R.C., V.L. Simhadri, M. Iandoli, E.S. Zuben and K.S. Chava (2014). Exposing synonymous mutations. Trends Genet. 30 (7): 308-321. doi: 10.1016/j.tig.2014.04.006.

Intharuksa, A., K. Masashi, P. Nichakan, C. Wannaree, A. Hirokazu, S. Yohei and S.A. Panee (2020). Evaluation of white Kwao Krua (*Pueraria candollei* Grah. ex Benth.) products sold in Thailand by molecular, chemical, and microscopic analyses. J. Nat. Med. 74(1): 106-118. doi: 10.1007/s11418-019-01351-2.

Intharuksa, A., D. Jessada, C. Sunee, T. Kannika and S. Suchada (2023). HPLC and DNA barcoding

profiles for identification of the selected twelve Mucuna species and its application for detecting prohibited aphrodisiac Mucuna products. Heliyon. 9(3): e14130. doi: 10.1016/j.heliyon.2023.e14130.

Jungsukcharoen, J., D.A. Binar, C. Wichai, V. Nawaporn, S. Polkit and B. Chuenchit (2014). *Pueraria mirifica* leaves, an alternative potential isoflavonoid source. Biosci. Biotechnol. Biochem. 78(6): 917-926. doi: 10.1080/09168451.2014.910091.

Kress, W.J. (2017). Plant DNA barcodes: Applications today and in the future. J Syst Evol. 55(4): 291-307. doi: 10.1111/jse.12254.

Kress, W.J. and D.L. Erickson (2007). A Two-Locus global DNA barcode for land plants: the coding rbcl gene complements the non-coding trnh-psba spacer region. PLoS ONE. 2(6): e508. doi: 10.1371/journal.pone.0000508.

Lackey, J.A. (1977a). Neonotonia, a new generic name to include Glycine wightii (Arnott) Verdcourt (Leguminosae, Papilionoideae). Phytologia. 37(3): 209-212.

Lackey, J.A. (1977b). A revised classification of the tribe Phaseoleae (Leguminosae: Papilionoideae), and its relation to canavanine distribution. Bot. J. Linn. Soc. 74(2): 163-178. doi: 10.1111/j.1095-8339.1977.tb01173.x.

Lee, J. and T. Hymowitz (2001). A molecular phylogenetic study of the subtribe Glycininae (Leguminosae) derived from the chloroplast DNA rps16 intron sequences. Am. J. Bot. 88(11): 2064-2073. doi: 10.2307/3558432.

Lindgren, C.J., K.L. Castro and H.A. Coiner (2013). The biology of invasive alien plants in canada. 12. *Pueraria montana* var. lobata (Willd.) Sanjappa & Predeep. Can. J. Plant Sci. 93(1): 71-95. doi: 10.4141/cjps2012-128.

Livak, K.J. and T.D. Schmittgen (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 25: 402 - 408. doi: 10.1006/meth.2001.1262.

Mahima, K., N.S.K. Koppala, V.R. Kanakarajan, S.R. Parameswaran, S. Ashutosh and S. Ramalingam (2022). Advancements and future prospective of DNA barcodes in the herbal drug industry. Front Pharmacol. 21(13): 947512. doi: 10.3389/fphar.2022.947512.

Manzanilla, V., A. Kool, N.L. Nguyen, H.V. Nong, H.T.H. Le and H. J.D. Boer (2018). Phylogenomics and barcoding of Panax: toward the identification of ginseng species. BMC Evol. Biol. 18(1): 44. doi: 10.1186/s12862-018-1160-y.

Munger, G.T. (2002). *Pueraria montana* var. lobata. In: Fire Effects Information System. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fire Sciences Laboratory.

Ohashi, H. (2005). A new species of Pueraria (Leguminosae) from Guizhou, China. Jap. J. Bot. 80: 9-13.

Ohashi, H. and Y. Iokawa (2006). A new species, *Pueraria xyzhuii* (Leguminosae) from Yunnan, China, with pollen stainability and pollen morphology in comparison to related species. Jap. J. Bot. 81: 26-34.

Pan, B., B. Liu, Z.X. Yu and Y.Q. Yang (2015). *Pueraria grandiflora* (Fabaceae), a new species from Southwest China. Phytotaxa. 203(3): 287. doi: 10.11646/phytotaxa.203.3.8.

Pawar, R S., S M. Handy, R. Cheng, N. Shyong and E. Grundel (2017). Assessment of the authenticity of herbal dietary supplements: comparison of chemical and DNA barcoding. Methods Planta Med. 83: 921-936. doi: 10.1055/s-0043-107881.

Soltis, D.E. and R.K. Kuzoff (1995). Discordance between nuclear and chloroplast phylogenies in the Heuchera group (Saxifragaceae). Evolution. 49(4): 727-742. doi: 10.1111/j.1558-5646.1995.tb02309.x.

Suntichaikamolkul, N., A. Tomoyoshi, M. Panupong, S. Kamonpan, R. Thanyada, B. Jean-Etienne, S. Hubert Schaller, D.E. Wanchai, V. Sornkanok, Y. Mami and S. Supaart (2023). Daidzein hydroxylation by CYP81E63 is involved in the biosynthesis of miroestrol in *Pueraria mirifica*. Plant Cell Physiol. 64(1): 64-79. doi: 10.1093/pcp/pcac140.

Suntichaikamolkul, N., K. Tantisuwanichkul, P. Prombutara, K. Kobtrakul, J. Zumsteg, S. Wannachart, H. Schaller, M. Yamazaki, K. Saito, W. De-eknamkul, S. Vimolmangkang and S. Sirikantaramas (2019). Transcriptome analysis of *Pueraria candollei* var. mirifica for gene discovery in the biosyntheses of isoflavones and miroestrol. BMC Plant Biol. 19(1): 581. doi: 10.1186/s12870-019-2205-0.

Vadivel, A.K.A., K. Kevin, T. Gang and D. Sangeeta (2018). Genome-wide identification and localization of chalcone synthase family in soybean (*Glycine max* [L] Merr). BMC Plant Biol. 18(1): 1-13. doi: 10.1186/s12870-018-1569-x.

Wang, X., F. Rongyan, L. Jia, L. Changfu and Z. Yansheng (2016). Molecular cloning and functional characterization of a novel (iso)flavone 4′,7-O-diglucoside glucosyltransferase from pueraria lobata. Front. Plant Sci. 7: 387. doi: 10.3389/fpls.2016.00387

Duan, H., C. Ming'en, Y. Jian, L. Changjiangsheng, Z.

Liangping, H. Yu, P. Huasheng and H. Huang (2018). Qualitative analysis and the profiling of isoflavonoids in various tissues of *Pueraria lobata* roots by ultra performance liquid chromatography quadrupole/time-of-flight-mass spectrometry and high performance liquid chromatography separation and ultraviolent-visible detection. Pharmacogn. Mag. 14(56): 397-403. doi: 10.4103/pm.pm_139_17.

Yazaki, K., A. Sugiyama, M. Morita and N. Shitan (2008). Secondary transport as an efficient membrane transport mechanism for plant secondary metabolites. Phytochemistry Reviews. 7(3): 513-524. doi: 10.1007/s11101-007-9079-8.

Yin, Y.C., X.D. Zhang, Z.Q. Gao, T. Hu and Y. Liu (2019). The research progress of *Chalcone Isomerase* (*CHI*) in plants. Mol Biotechnol. 61(1):32-52. doi: 10.1007/s12033-018-0130-3.

Yusakul, G., P. Waraporn, U. Orapin, J. Thaweesak and C. Chaiyo (2011). Comparative analysis of the chemical constituents of two varieties of *Pueraria candollei*. Fitoterapia. 82(2): 203-207. doi: 10.1016/j.fitote.2010.09.009.

Zhang, G., L. Jinxin, G. Mei, K. Weijun, Z. Qing, S. Linchun and W. Qiuling (2020). Tracing the edible and medicinal plant *pueraria montana* and its products in the marketplace yields subspecies level distinction using dna barcoding and DNA metabarcoding. Front. Pharmacol. 11: 512963. doi: 10.3389/fphar.2020.00336.